

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

## The impact of targeted diabetic retinopathy training for graders in Vietnam and the implications for future diabetic retinopathy screening programmes

|                               |  |
|-------------------------------|--|
| Journal:                      | <i>BMJ Open</i>  |
| Manuscript ID                 | bmjopen-2021-059205  |
| Article Type:                 | Original research  |
| Date Submitted by the Author: | 11-Nov-2021  |
| Complete List of Authors:     | Curran, Katie; Queen's University Belfast School of Medicine Dentistry and Biomedical Sciences, Centre of Public Health<br>Congdon, Nathan; Queen's University of Belfast; ORBIS International<br>Hoang, Tung; Hanoi Medical University, Department of Ophthalmology;<br>The University of Sydney Save Sight Institute<br>Lohfeld, Lynne; Queen's University of Belfast; Eye Hospital of Wenzhou Medical University<br>Nguyen, Van; Orbis International in Vietnam<br>Nguyen, Hue; Orbis International in Vietnam<br>Nguyen, Quan; Ho Chi Minh Eye Hospital, Vitreo-Retina Department<br>Dardis, Catherine; Belfast Health and Social Care Trust, Department of Ophthalmology<br>Virgili, Gianni ; Queen's University of Belfast; Belfast Health and Social Care Trust, Department of Ophthalmology<br>Piyasena, Prabhath; Queen's University of Belfast<br>Tran, Huong; Orbis International in Vietnam<br>Salongcay, Recivall; Queen's University Belfast<br>Tung, Mai; Ho Chi Minh Eye Hospital, Vitreo-Retina Department<br>Peto, Tunde; Queen's University of Belfast; Belfast Health and Social Care Trust, Department of Ophthalmology |
| Keywords:                     | Diabetic retinopathy < DIABETES & ENDOCRINOLOGY, OPHTHALMOLOGY, PUBLIC HEALTH, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT  |
|                               |  |

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3 1 **Title: The impact of targeted diabetic retinopathy training for graders in Vietnam and**  
4  
5 2 **the implications for future diabetic retinopathy screening programmes**  
6  
7

8  
9 3 **Short title:** Training diabetic retinopathy graders in Vietnam  
10

11  
12 4 **Authors:**  
13

14  
15  
16 5 \*Katie Curran, MPH, PhD<sup>1</sup> ([k.curran@qub.ac.uk](mailto:k.curran@qub.ac.uk))  
17

18  
19 6 Nathan Congdon, MD, MPH<sup>1,2,3</sup> ([Nathan.Congdon@orbis.org](mailto:Nathan.Congdon@orbis.org))  
20

21  
22 7 Tung Thanh Hoang<sup>4,5</sup> ([tung.hoang@sydney.edu.au](mailto:tung.hoang@sydney.edu.au))  
23

24  
25 8 Lynne Lohfeld, MPH, PhD<sup>1,6</sup> ([l.lohfeld@qub.ac.uk](mailto:l.lohfeld@qub.ac.uk))  
26

27  
28 9 Van Thu Nguyen, BA<sup>7</sup> ([van.nguyen@orbis.org](mailto:van.nguyen@orbis.org))  
29

30  
31 10 Hue Thi Nguyen, MD, MPH<sup>7</sup> ([hue.nguyen@orbis.org](mailto:hue.nguyen@orbis.org))  
32

33  
34 11 Quan Nhu Nguyen, MD, PhD<sup>8</sup> ([bsquan2009@gmail.com](mailto:bsquan2009@gmail.com))  
35

36  
37 12 Catherine Dardis, MSc<sup>9</sup> ([cathy831oph@gmail.com](mailto:cathy831oph@gmail.com))  
38

39  
40 13 [Gianni Virgili MD<sup>1,9</sup>](mailto:G.Virgili@qub.ac.uk) ([G.Virgili@qub.ac.uk](mailto:G.Virgili@qub.ac.uk))  
41

42  
43 14 [Prabhath Piyasena, MB BS, PhD<sup>1</sup>](mailto:m.piyasena@qub.ac.uk), ([m.piyasena@qub.ac.uk](mailto:m.piyasena@qub.ac.uk))  
44

45  
46 15 Huong Tran, MPA<sup>7</sup> ([Huong.Tran@orbis.org](mailto:Huong.Tran@orbis.org))  
47

48  
49 16 Recivall Salongcay, MD, MPM<sup>1</sup> ([R.SALONGCAY@qub.ac.uk](mailto:R.SALONGCAY@qub.ac.uk))  
50

51  
52 17 Mai Quoc Tung, MD, PhD<sup>8</sup> ([maitungtn@gmail.com](mailto:maitungtn@gmail.com))  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 18 Tunde Peto, MD, PhD<sup>1,9</sup> ([T.Peto@qub.ac.uk](mailto:T.Peto@qub.ac.uk))  
4  
5

6  
7 **Affiliations:**  
8  
9

10 20 <sup>1</sup>Centre for Public Health, Queen's University of Belfast, Belfast, UK  
11  
12

13 21 <sup>2</sup>Orbis International, New York, USA  
14  
15

16 22 <sup>3</sup>Zhongshan Ophthalmic Centre, Guangzhou, China  
17  
18

19 23 <sup>4</sup>Ophthalmology Department, Hanoi Medical University, Hanoi, Vietnam  
20  
21  
22

23 24 <sup>5</sup>Save Sight Institute, The University of Sydney School of Medicine, Sydney, Australia  
24  
25  
26

27 25 <sup>6</sup>Eye Hospital of Wenzhou Medical University, Wenzhou, Zhejiang Province, China  
28  
29  
30

31 26 <sup>7</sup>Orbis International in Vietnam, Hanoi, Vietnam  
32  
33

34 27 <sup>8</sup>Vitreo-Retina Department, Ho Chi Minh Eye Hospital, Ho Chi Minh City, Vietnam  
35  
36  
37

38 28 <sup>9</sup>Department of Ophthalmology, Belfast Health and Social Care Trust, NI  
39  
40

41 29 **\*Corresponding author**  
42  
43

44 30 Katie Curran  
45  
46

47 31 Centre for Public Health  
48  
49

50 32 Institute of Clinical Sciences, A  
51  
52

53 33 Queen's University Belfast  
54  
55

56 34 BT12 6BA  
57  
58

59 35 Email: [k.curran@qub.ac.uk](mailto:k.curran@qub.ac.uk)  
60

1  
2  
3 36 **Keywords:** Diabetic retinopathy, diabetic retinopathy training, screening  
4  
5

6 37 **Word Count:** 2265  
7  
8

9 38 **Abstract**  
10  
11

12  
13 39 **Objectives:** To compare the accuracy of trained level 1 diabetic retinopathy (DR) graders  
14  
15 40 (nurses, endocrinologists, one general practitioner), level 2 graders (mid-level  
16  
17 41 ophthalmologists) and level 3 graders (senior ophthalmologists) in Vietnam against a  
18  
19 42 reference standard from the UK, and assess the impact of supplementary targeted grader  
20  
21 43 training.  
22  
23

24  
25 44 **Methods:** DR training was delivered to new Vietnamese graders in February 2018 by  
26  
27 45 National Health System (NHS) UK graders. Two-field retinal images were taken by trained  
28  
29 46 screeners and graded by 14 trained graders in Vietnam between August-October 2018 and  
30  
31 47 then re-graded retrospectively by an NHS-certified reference standard UK optometrist (Phase  
32  
33 48 I). Further directed DR training based on Phase I results was delivered to Vietnamese graders  
34  
35 49 between March-November 2019. After training was delivered, a randomised subset of images  
36  
37 50 from January-October 2020 was graded by 6 of the original cohort (Phase II). The reference  
38  
39 51 grader re-graded all images from Phase I and II retrospectively in masked fashion. Sensitivity  
40  
41 52 was calculated at the two different time points and Chi Squared was used to test significance.  
42  
43  
44  
45  
46  
47

48 53 **Results:** In Phase I, the sensitivity for detecting any DR for all grader groups in Vietnam  
49  
50 54 was low and improved in Phase II after additional training was delivered. The greatest  
51  
52 55 improvement was seen among level 1 graders ( $P < 0.001$ ) and the lowest improvement was  
53  
54 56 observed among level 3 graders ( $P = 0.326$ ). There was an improvement in sensitivity for  
55  
56 57 detecting DR and referable diabetic macular oedema between all grader levels and whilst the  
57  
58  
59  
60

1  
2  
3 58 differences were statistically significant, the post-training values were suboptimal (41.8% to  
4  
5 59 61.5%). The main disagreement was the detection of ungradable images.  
6  
7  
8

9 60 **Conclusions:** This is among the first studies to demonstrate that targeted training  
10  
11 61 interventions can improve accuracy of DR grading in a low-middle income country. These  
12  
13 62 findings have important implications for improving service delivery in DR screening  
14  
15  
16 63 programmes in low-resource settings.  
17  
18

## 19 64 **Article Summary**

### 23 65 **Strengths and limitations of this study**

- 26 66 • This is the first study describing the impact of a training intervention to improve the  
27  
28 67 quality of DR grading in an LMIC.
- 31 68 • Reinforcing training to identify ungradable images has been acknowledged.
- 33 69 • The sample size was smaller in Phase II compared to Phase I, however, there were no  
34  
35 70 statistically significant differences between the groups.  
36  
37  
38

39 71

42 72  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 74 **Introduction**

75 The prevalence of diabetes among adults in Vietnam is approximately 6% and has almost  
76 doubled over the past decade.[1] Early detection through diabetic eye screening programmes  
77 (DESPs) is important to reduce the risk of avoidable blindness due to diabetic retinopathy  
78 (DR). Since the introduction of systematic DESPs in the UK, a high-income country (HIC),  
79 diabetic retinopathy (DR) is no longer the leading cause of blindness among working age  
80 adults.[2] The key to such successful DESPs is implementing accurate, innovative and cost-  
81 effective models tailored to fit healthcare systems and contexts.

82 Investing in training personnel to increase human resources and procuring appropriate  
83 diagnostic and treatment equipment are essential to ensure that service providers can deliver  
84 optimum care for people with DR. In low-middle income countries (LMICs), there is often  
85 insufficient capacity to implement robust DESPs due to the lack of skilled human resources  
86 and infrastructure.[3,4] In Vietnam, there are only 14 ophthalmologists per million population  
87 compared to 49 per million in the UK.[5]

88 All screening programmes must provide evidence of their ability to detect the targeted  
89 condition and ensure that the service performs efficiently to improve screening accuracy  
90 when it falls short. To date, there is insufficient evidence on DR grading accuracy using non-  
91 mydriatic digital imaging by trained graders in LMICs, and even less about the capacity of  
92 DESPs in LMICs to improve where poor accuracy is detected. The current retrospective  
93 study is designed to assess accuracy of a range of graders in a non-governmental organisation  
94 (NGO)-supported DESP in Vietnam, and to study the efficacy of a quality-improvement  
95 intervention.

## 96 **Methods**



1  
2  
3 97 **Study participants:** The 14 participants from Vietnam in Phase I included: Level 1 DR  
4  
5 98 graders (6 nurses, 1 general practitioner and 2 endocrinologists, all with < 1 year grading  
6  
7 99 experience, 55.6% female), Level 2 DR graders (3 newly-qualified ophthalmologists with < 1  
8  
9 100 year formal DR grading experience, 100% female), and Level 2 DR graders (2 senior  
10  
11 101 ophthalmologists with >5 years' experience providing treatment for sight threatening DR, but  
12  
13 102 with <1 year formal DR grading experience, 100% male). In Phase II, 6/14 graders (3 Level  
14  
15 103 1, 2 Level 2, 1 Level 3) from Phase I were included. The reference standard from the UK  
16  
17 104 (KC) was a fully-qualified optometrist trained in DR grading and certified by the UK NHS  
18  
19 105 DESP.[6] Vietnamese Level 1, 2 and 3 graders are equivalent to primary, secondary and  
20  
21 106 arbitration graders, respectively, in UK DESPs.[7] In the current study, Vietnamese Level 1  
22  
23 107 and Level 2 graders graded all fundus images for DR. All images having disagreement  
24  
25 108 between graders, and an additional randomly-selected 40% of all images, were sent for  
26  
27 109 arbitration grading by Level 3 graders in Vietnam. All graders in Vietnam were masked to  
28  
29 110 any prior diagnoses or grades of the reference standard, while the reference standard was also  
30  
31 111 masked to results of grading in Vietnam.  
32  
33  
34  
35  
36  
37  
38

39 112 **DR training for graders in Vietnam:** As part of a DESP project supported by NGO Orbis  
40  
41 113 International, a team of five Vietnamese doctors and medical administrators visited a  
42  
43 114 Northern Ireland (NI) DESP in September 2017 to receive training on screening, programme  
44  
45 115 administration and quality control methods. In February 2018, a senior UK NHS grader from  
46  
47 116 the Belfast Trust (CD) and a fully-qualified optometrist, trained in DR grading and certified  
48  
49 117 by the NHS (KC), visited Vietnam to deliver DR training to graders involved in the DESPs.  
50  
51 118 Training focused on ocular anatomy, retinal diseases, DR signs and grading (based on the UK  
52  
53 119 National Screening Committee (NSC) classification system), and appropriate referral  
54  
55 120 pathways and management.[8]  
56  
57  
58  
59  
60

1  
2  
3 121 **Image acquisition and management:** Images were captured by trained nurses and  
4  
5 122 technicians in Vietnam. Two-field, 45° digital colour photographs (one disc-centred and one  
6  
7 123 macula-centred) were taken using a non-mydriatic camera (Canon CR2-AF, Canon Medical  
8  
9 124 Systems. Europe), in accordance with the UK's NHS DESP.[9] Nurses and technicians were  
10  
11 125 trained to repeat inadequate images as a quality control measure and take anterior segment  
12  
13 126 photographs where adequate fundus images were not possible. Images were anonymised and  
14  
15 127 uploaded to a cloud-based software system (Spectra)® for analysis by trained DR graders in  
16  
17 128 Vietnam. The images were transferred to a Queen's University Belfast (QUB) server for re-  
18  
19 129 grading by the reference standard.

20  
21  
22  
23  
24  
25 130 **Assessment of gradeability:** Image quality was defined as 'adequate' or 'inadequate' in  
26  
27 131 accordance with NHS DESP guidelines as outlined below;

- 28  
29  
30  
31 132 • Adequate disc-centred image: complete optic disc >2DD from edge of image and fine  
32  
33 133 vessels visible on surface of the disc.[9]  
34  
35 134 • Adequate macula-centred image: centre of fovea >2DD from edge of image and  
36  
37 135 vessels visible within 1DD of centre of fovea.[9]  
38  
39  
40

41 136 The disc-centred and macula-centred images for each eye were viewed as a pair and graded at  
42  
43 137 an individual eye level. The presence of DR and diabetic macular oedema (DMO) was also  
44  
45 138 determined at a patient level and based on the worst affected eye. Ungradable images were  
46  
47 139 referred for further slit-lamp examination. Where images were considered inadequate but  
48  
49 140 referable disease was detectable, the referable grade was recorded and the patients were moved  
50  
51 141 onto the appropriate referable grade pathway.[9]  
52  
53

54  
55  
56 142 Consecutive patients diagnosed with diabetes and undergoing evaluation for possible DR at  
57  
58 143 Ho Chi Minh City General Hospital and Ho Chi Minh Eye Hospital (tertiary hospitals), Tien  
59  
60

1  
2  
3 144 Giang General Hospital (provincial hospital) and Cai Ba General Hospital (district hospital)  
4  
5 145 in Vietnam were recruited. Fundus images from August to October 2018 (Phase I) were  
6  
7 146 graded by 14 graders in Vietnam and then re-graded retrospectively by a reference standard  
8  
9 147 from the UK in Phase I. Targeted remedial training, based on specific findings from the  
10  
11 148 Phase I analysis, was delivered in March 2019 and November 2019 by UK graders and Orbis.  
12  
13 149 Additionally, regular testing and training for quality assurance purposes was also introduced,  
14  
15 150 similar to UK DESP models. To evaluate the impact of this quality-improvement  
16  
17 151 intervention, a new subset of images was graded by six of the original cohort of graders  
18  
19 152 between January-October 2020 (Phase II) and re-graded by the reference standard from the  
20  
21 153 UK (KC) in September 2021.

22  
23  
24  
25  
26  
27 154 **Statistical analysis:** Data were entered into Microsoft Excel version 16.0 and then  
28  
29 155 transferred to Stata 16.0 (StataCorp LLC) for analysis. Intra and inter-grader agreement was  
30  
31 156 calculated using kappa and a stratified random sampling technique was utilised to ensure a  
32  
33 157 representative sample of images was re-graded (Supplementary Files S1 and S2). Diagnostic  
34  
35 158 test accuracy (DTA) comparing graders in Vietnam with the UK reference standard was  
36  
37 159 assessed by calculating sensitivity, specificity, positive predicative values (PPV) and negative  
38  
39 160 predictive values (NPV). Sensitivity was calculated at the two different time points (Phase I  
40  
41 161 and Phase II) and Chi Squared was used to test significance.

42  
43  
44  
45  
46  
47 162 Patients or the public were not involved in the design, or conduct, or reporting, or  
48  
49 163 dissemination plans of our research.

## 50 51 52 164 **Results**

53  
54  
55 165 **Patient characteristics:** In Phase I, 65.4% of patients were female with a mean age 59.4  
56  
57 166 years. In Phase II, 40.0% were female with a mean age of 59.8 years. Figures 1 and 2

1  
2  
3 167 describe enrolment of patients and capture and grading of images in Phase I and II of the  
4  
5 168 study respectively.  
6  
7

8 169 Figure 1  
9

10 170  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

171 Figure 2

172

For peer review only

173 **Initial grading performance analysis (Phase I):** The sensitivity for detecting any DR was  
 174 low against the reference standard in the UK for all grader groups in Vietnam. The sensitivity  
 175 for detecting referable DR and referable DMO was even lower for all grader groups (Table  
 176 1).

177 **Table 1. Diagnostic test accuracy of DR graders in Vietnam against a reference**  
 178 **standard from the UK, including ungradable images.**

|  | Level 1 graders<br>(n=410 patient<br>images)* | Level 2 graders<br>(n=410 patient<br>images)* | Level 3 graders<br>(n=260 patient<br>images)† |
|--|---|---|---|
| <b>Any DR</b>  |   |   |   |
| Sensitivity (%) (95% CI)   | 41.8 (33.9, 50.1)                             | 42.5 (34.5, 50.7)                             | 42.2 (33.1, 51.8)                             |
| Specificity (%) (95% CI)   | 87.9 (83.3, 91.7)                             | 98.8 (96.6, 99.8)                             | 100 (97.5, 100)                               |
| PPV (%) (95% CI)   | 67.4 (57.0, 76.6)                             | 95.6 (87.6, 99.1)                             | 100 (92.7, 100)                               |
| NPV (%) (95% CI)   | 71.7 (66.4, 76.7)                             | 74.3 (69.3, 78.8)                             | 68.2 (61.5, 74.5)                             |
| <b>Referable DR</b>  |   |   |   |
| Sensitivity (%) (95% CI)   | 19.2 (9.63, 32.5)                             | 13.5 (5.59, 25.8)                             | 10.5 (2.94, 24.8)                             |
| Specificity (%) (95% CI)   | 97.2 (94.9, 98.7)                             | 100 (99.0, 100)                               | 99.5 (97.5, 100)                              |
| PPV (%) (95% CI)   | 50.0 (27.2, 72.8)                             | 100 (59.0, 100)                               | 80.0 (28.4, 99.5)                             |
| NPV (%) (95% CI)   | 89.2 (85.7, 92.1)                             | 88.8 (85.3, 91.7)                             | 86.7 (81.9, 90.6)                             |
| <b>Referable DMO</b>   |   |   |   |
| Sensitivity (%) (95% CI)   | 5.8 (1.91, 13.0)                              | 20.9 (12.9, 31.0)                             | 16.9 (8.76, 28.3)                             |
| Specificity (%) (95% CI)   | 97.2 (94.8, 98.7)                             | 99.4 (97.8, 99.9)                             | 100 (98.1, 100)                               |
| PPV (%) (95% CI)   | 35.7 (12.8, 64.9)                             | 90.0 (68.3, 98.8)                             | 100 (71.5, 100)                               |
| NPV (%) (95% CI)   | 79.5 (75.2, 83.4)                             | 82.6 (78.4, 86.2)                             | 78.3 (72.7, 83.3)                             |
| <p><b>Abbreviations:</b> UK = United Kingdom, DR = Diabetic Retinopathy, DMO = Diabetic Macular Oedema, CI = Confidence Intervals,<br/>           Grading criteria: UK National Diabetic Eye Screening Programme (NDESP) classification system.<br/>           Criteria:<br/>           Any DR, is defined as grades R1, R2, R3s, R3a and U.<br/>           Referable DR is defined as grades R2, R3a and U<br/>           Referable DMO is defined as grades M1 and U<br/>           *Missing (n=2, 0.5%), †missing (n=2, 0.8%)</p> |   |   |   |

179

180 **Subsequent grading performance after retraining (Phase II):** Subsequently, a further  
 181 subset of images from 115 consecutive patients from January to October 2020 were graded  
 182 by 6 of the original cohort of 14 Vietnamese graders, and were regraded in the UK to  
 183 evaluate graders' performance after targeted training was delivered and quality control  
 184 measures were instituted. The greatest improvement in sensitivity for detecting any DR was  
 185 seen among level 1 graders (difference: +45.4%, 95%CI +33.1% to +57.8%; P<0.001). The  
 186 specificity increased from 87.9% in phase I to 95.6% in phase II which helps to avoid over  
 187 referrals (difference: +7.7%, 95%CI +1.4% to +13.9%; p=0.069). The lowest improvement in  
 188 sensitivity for detecting any DR was observed between level 3 graders in Vietnam  
 189 (difference; +9.0%, 95%CI: -9.0% to +27.1%; p=0.326), although their specificity remained  
 190 100% at phase I and phase II. There was an improvement in sensitivity for detecting DR and  
 191 referable DMO between all grader levels and whilst there were statistically significant  
 192 differences, sensitivities after training were still insufficient and comprised between about  
 193 40% and 61.5% (Table 2).

194 **Table 2: Diagnostic test accuracy of DR graders in Vietnam against a reference**  
 195 **standard from the UK after additional DR training was delivered.**

|   | <b>Level 1 graders<br/>(n=115 patient<br/>images)</b> | <b>Level 2<br/>graders (n=115<br/>patient images)</b> | <b>Level 3 graders<br/>(n=62 patient<br/>images)</b> |
|---|---|---|--|
| <b>Any DR</b>                               |   |   |  |
| Sensitivity (%) (95% CI)                    | 87.2 (74.3, 95.2)                                     | 68.1 (52.9, 80.9)                                     | 51.3 (34.8, 67.6)                                    |
| Specificity (%) (95% CI)                    | 95.6 (87.6, 99.1)                                     | 95.6 (87.6, 99.1)                                     | 100 (84.6, 100)                                      |
| PPV (%) (95% CI)                            | 93.2 (81.3, 98.6)                                     | 91.4 (76.9, 98.2)                                     | 100 (83.2, 100)                                      |
| NPV (%) (95% CI)                            | 91.5 (82.5, 96.8)                                     | 81.3 (71.0, 89.1)                                     | 53.7 (37.4, 69.3)                                    |
| P-value comparing sensitivity to<br>Phase I | P=0.000   | P=0.002   | P=0.326  |
| <b>Referable DR</b>                         |   |   |  |
| Sensitivity (%) (95% CI)                    | 53.3 (26.6, 78.7)                                     | 40.0 (16.3, 67.7)                                     | 58.3 (27.7, 84.8)                                    |
| Specificity (%) (95% CI)                    | 90.0 (82.4, 95.1)                                     | 93.0 (86.1, 97.1)                                     | 100 (92.7, 100)                                      |
| PPV (%) (95% CI)                            | 44.4 (21.5, 69.2)                                     | 46.2 (19.2, 74.9)                                     | 100 (59.0, 100)                                      |
| NPV (%) (95% CI)                            | 92.8 (85.7, 97.0)                                     | 91.2 (83, 95.9)                                       | 90.7 (79.7, 96.9)                                    |
| P-value comparing sensitivity to<br>Phase I | P=0.009   | P=0.022   | P=0.001  |
| <b>Referable DMO</b>                        |   |   |  |

|  |                   |                   |                   |
|--|-------------------|-------------------|-------------------|
| Sensitivity (%) (95% CI)                 | 56.3 (29.9, 80.2) | 43.8 (19.8, 70.1) | 61.5 (31.6, 86.1) |
| Specificity (%) (95% CI)                 | 97.0 (91.4, 99.4) | 93.9 (87.3, 97.7) | 100 (92.6, 100)   |
| PPV (%) (95% CI)                         | 75.0 (42.8, 94.5) | 53.8 (25.1, 80.8) | 100 (63.1, 100)   |
| NPV (%) (95% CI)                         | 93.2 (86.5, 97.2) | 91.2 (83.9, 95.9) | 90.6 (79.3, 96.9) |
| P-value comparing sensitivity to Phase I | P=0.000           | P=0.051           | P=0.002           |

**Abbreviations:** UK = United Kingdom, DR = Diabetic Retinopathy, DMO = Diabetic Macular Oedema, CI = Confidence Intervals,  
Grading criteria: UK National Diabetic Eye Screening Programme (NDESP) classification system.

Criteria:

Any DR is defined as grades R1, R2, R3s, R3a and U.

Referable DR is defined as grades R2, R3a and U

Referable DMO is defined as grades M1 and U

Chi-squared used to compare sensitivity between Phase I and II.

196

197 The main discordance is detecting ungradable images, which does not improve much with  
198 training; therefore, training should be reinforced to ungradable images for the future  
199 (Supplementary File S3).

## 200 Discussion

201 Results from our study demonstrate extremely poor sensitivity and specificity for detecting  
202 all levels of DR, especially referable DR, in the early stages of programme delivery. This  
203 translates into increased costs to the health care system due to missed opportunities for early  
204 treatment and un-necessary examinations for false-positive referrals. The quality of patient  
205 care also suffers. Didactic DR training was delivered to graders in Vietnam over a two-year  
206 period by trained DR graders from the UK and Vietnam. Training was specifically targeted to  
207 address problems identified in the Phase I testing, and quality control testing using  
208 international test and training (iTAT) were also undertaken. This study demonstrates that  
209 these steps led to improved grading accuracy for all classes of patients and graders; however,  
210 results remain suboptimal for a screening programme. The main discordance was detection of  
211 ungradable images, therefore, training should be reinforced to detect ungradable images.



1  
2  
3 212 Results can be poor in these settings for a variety of reasons, quality assessment is crucial,  
4  
5 213 and programmatic changes based on models such as the UK DESP can be successful in  
6  
7 214 enhancing grader accuracy in LMICs settings. However, it is fundamental for countries to  
8  
9 215 adapt their own DR classification system and referral pathways to meet their requirements.  
10  
11 216 Most importantly, the role of affiliated hospitals (and partnerships, coordination among  
12  
13 217 training institutions and practical hospitals) are crucial for DR grading quality improvement  
14  
15  
16  
17  
18 218 Studies in LMICs and HICs have assessed the accuracy of non-medical graders and medical  
19  
20 219 graders in the detection of DR and found that both grader types are capable of achieving  
21  
22 220 moderate-high sensitivity for detecting DR.[7, 10-13] Previous studies have briefly described  
23  
24 221 what training interventions were used to train their graders, although no study has outlined  
25  
26 222 whether additional training methods were employed to improve grading accuracy if needed.  
27  
28 223 In the UK, the DR grading course by the Gloucestershire Retinal Education Group is  
29  
30 224 compulsory for graders in addition to monthly iTAT.[6] This formal training qualification  
31  
32 225 and continuous monitoring and evaluation are crucial to achieve optimal sensitivity, which  
33  
34 226 may be more challenging in terms of costs and capacity for LMICs.  
35  
36  
37  
38  
39

40 227 **Strengths:** To our knowledge, this is the first study describing the impact of a training  
41  
42 228 intervention to improve the quality of DR grading in an LMIC. The inclusion of ungradable  
43  
44 229 images in this study was a logical decision, particularly when the prevalence of cataract  
45  
46 230 (which often renders DR images ungradable) is high in LMICs.[10] Dense cataracts normally  
47  
48 231 obstruct the view of the fundus, making it difficult to obtain clear fundus photographs and  
49  
50 232 assign a DR grade. In these instances, referring patients to an eye clinic for further  
51  
52 233 assessment and treatment as needed is required. Determining sensitivity and specificity at the  
53  
54 234 patient level is also important from a DESP implementation perspective. In the UK and  
55  
56 235 Vietnam, both eyes are typically examined for DR and a single outcome is assigned to the  
57  
58  
59  
60

1  
2  
3 236 patient, as was done here. For these reasons, we feel our analytic approach, and thus results,  
4  
5 237 are relevant to these settings.  
6  
7  
8

9 238 **Limitations:** Limitations for this study have also been acknowledged. Data from this study  
10  
11 239 represent routine clinical practice. In daily DR screening, not all patients undergoing primary  
12  
13 240 (Level 1) and secondary (Level 2) grading proceed to arbitration grading (Level 3). This  
14  
15 241 means a proportion of images were not graded by arbitration graders as outlined in figure 1  
16  
17 242 and figure 2. Second, only 6/14 graders from Phase I were included in Phase II grading;  
18  
19 243 however the distribution of grader levels was similar. Third, though the proportion of patients  
20  
21 244 excluded was small, we are unable to fully characterise the reasons for these exclusions, due  
22  
23 245 to the nature of the study as a programmatic evaluation. Some potential reasons for this are a  
24  
25 246 patient's unwillingness to participate in the study, graders having forgotten to ask for patient  
26  
27 247 consent to participate in the study, and patient inability to comply with image capture. Fourth,  
28  
29 248 pupil status was not recorded in this study and this can be important for LMICs. Finally, it  
30  
31 249 was not practical for the UK reference standard to examine patients clinically in Vietnam;  
32  
33 250 however, the method of grading by a certified DR grader or clinical specialist is widely used  
34  
35 251 as the reference standard in many screening programmes.  
36  
37  
38  
39  
40  
41

42 252 **Conclusions:** This paper shows how grading accuracy was particularly low among all grader  
43  
44 253 groups in Vietnam in the first six months of DESP implementation. Many factors may have  
45  
46 254 contributed to poor grader performance, including inadequate training and feedback,  
47  
48 255 insufficient time to participate in quality assurance testing and competing work  
49  
50 256 responsibilities. After additional training, testing and quality assurance systems were  
51  
52 257 implemented in Vietnam, DTA improved among all grader groups, however a significant  
53  
54 258 amount of work is still needed. In particular, training graders to detect ungradable cases is  
55  
56 259 crucial. A qualitative study to determine why the initial training intervention was less  
57  
58  
59  
60

1  
2  
3 260 successful should be explored. Since further improvements are required, understanding how  
4  
5 261 other countries implement such programmes would be beneficial. Future studies should  
6  
7  
8 262 outline what DR training interventions were used, state relevant training courses and explain  
9  
10 263 what quality assurance measures are in place. The findings from this study are important for  
11  
12 264 DESP programme planners in Vietnam and other LMICs, highlighting the importance of  
13  
14 265 quality monitoring and directed re-training as needed.  
15  
16  
17

18 266 **Acknowledgements:** The authors are very grateful to all participants in Vietnam and  
19  
20 267 Northern Ireland who generously gave their time for this study and to Orbis International.  
21  
22

23 268 **Author contributions:** All authors have made substantial contributions to the conception of  
24  
25 269 the design of the work. KC and GV analysed and interpreted the data. All authors read and  
26  
27 270 approved the final manuscript.  
28  
29  
30

31 271 **Funding:** The project was funded by the Department for the Economy (DfE) – Global  
32  
33 272 Challenges Research Fund (GCRF) Awards (Grant number N/A). None of the funders were  
34  
35 273 involved in the design or conduct of the study; preparation, review or approval of the  
36  
37 274 manuscript; or decision to submit the manuscript for publication. Nathan Congdon is  
38  
39 275 supported by the Ulverscroft Foundation (UK). Katie Curran is funded by the Wellcome  
40  
41 276 Trust.  
42  
43  
44  
45

46 277 **Competing Interests:** Nathan Congdon is employed as a Research Director by Orbis  
47  
48 278 International.  
49  
50

51  
52 279 **Patient consent for publication:** Not required.  
53  
54

55 280 **Ethics approval:** This research adhered to the tenets of the Declaration of Helsinki. Ethical  
56  
57 281 approval was granted by the Hanoi Medical University Institutional Review Board in Bio-  
58  
59  
60

1  
2  
3 282 Medical Research, Vietnam (No. 0518/HMU IRB). Written informed consent was obtained  
4  
5 283 from all participants prior to their being interviewed.  
6  
7  
8

9 284 **Data sharing statement:** All datasets relevant to the study are included in the article or  
10  
11 285 uploaded as supplementary information.  
12  
13

## 14 286 **References**

- 15  
16  
17 287 1. Ngoc NB, Lin ZL, Ahmed W. Diabetes: What challenges lie ahead for Vietnam? *Ann*  
18  
19 288 *Glob Heal.* 2020;86(1):1–9.  
20  
21  
22 289 2. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness  
23  
24 290 certifications in England and Wales in working age adults (16-64 years), 1999-2000  
25  
26 291 with 2009-2010. *BMJ Open.* 2014;4(2):1–6.  
27  
28  
29 292 3. Burgess PI, Msukwa G, Beare NAV. Diabetic retinopathy in sub-Saharan Africa:  
30  
31 293 Meeting the challenges of an emerging epidemic. *BMC Med.* 2013;11(1):1.  
32  
33  
34 294 4. Adriono G, Wang D, Octavianus C, Congdon N. Use of eye care services among  
35  
36 295 diabetic patients in urban Indonesia. *Arch Ophthalmol* (Chicago, Ill 1960).  
37  
38 296 2011;129(7):930–5.  
39  
40  
41 297 5. International Council of Ophthalmology. Number of Ophthalmologists in Practic and  
42  
43 298 Training Worldwide [Internet]. 2012 [cited 2020 Aug 13]. Available from:  
44  
45 299 <http://www.icoph.org/ophthalmologists-worldwide.html>  
46  
47  
48 300 6. Gloucestershire Retinal Education Group. Diabetic Retinopathy Grading Course  
49  
50 301 [Internet]. 2020. [cited 2021 April] Available from: [https://drscreening.org/diabetic-](https://drscreening.org/diabetic-retinopathy-grading-course)  
51  
52 302 [retinopathy-grading-course](https://drscreening.org/diabetic-retinopathy-grading-course)  
53  
54  
55  
56 303 7. Widdowson S, Taylor D. The management of grading quality: Good practice in the  
57  
58  
59  
60

- 1  
2  
3 304 quality assurance of grading. [Internet] 2015. [cited 2021 April] Available from:  
4  
5 305 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachmen](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/512832/The_Management_of_Grading.pdf)  
6  
7  
8 306 [t\\_data/file/512832/The\\_Management\\_of\\_Grading.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/512832/The_Management_of_Grading.pdf)  
9  
10  
11 307 8. Public Health England. NHS Diabetic Eye Screening Programme Grading definitions  
12  
13 308 for referable disease. [Internet] 2021 (cited 2021 September) Available from:  
14  
15 309 [https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-](https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-grading-criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-referable-disease)  
16  
17 310 [grading-criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-](https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-grading-criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-referable-disease)  
18  
19 311 [referable-disease](https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-grading-criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-referable-disease)  
20  
21  
22  
23 312 9. NHS DESP Core Team. Diabetic Eye Screening Programme Pathway for Adequate /  
24  
25 313 Inadequate Images and where images cannot be taken Outlines pathway and business  
26  
27 314 rules for image capture exceptions and ungradable images. [Internet] 2013.[cited 2021  
28  
29 315 April] Available from: [https://www.gov.uk/government/publications/diabetic-eye-](https://www.gov.uk/government/publications/diabetic-eye-screening-pathway-for-images-and-where-images-cannot-be-taken)  
30  
31 316 [screening-pathway-for-images-and-where-images-cannot-be-taken](https://www.gov.uk/government/publications/diabetic-eye-screening-pathway-for-images-and-where-images-cannot-be-taken)  
32  
33  
34  
35 317 10. Murthy GVS, John N, Shamanna BR, Pant HB. Elimination of avoidable blindness  
36  
37 318 due to cataract: Where do we prioritize and how should we monitor this decade. *Indian*  
38  
39 319 *J Ophthalmol.* 2012;60(5):438–45.  
40  
41  
42  
43 320 11. McKenna M, Chen T, McAneney H, Membrillo MAV, Jin L, Xiao W, et al. Accuracy  
44  
45 321 of trained rural ophthalmologists versus non-medical image graders in the diagnosis of  
46  
47 322 diabetic retinopathy in rural China. *Br J Ophthalmol.* 2018;102(11):1471–6.  
48  
49  
50 323 12. Amirul Islam FM. Accuracy and reliability of retinal photo grading for diabetic  
51  
52 324 retinopathy: Remote graders from a developing country and standard retinal photo  
53  
54 325 grader in Australia. *PLoS One.* 2017;12(6):1–12.  
55  
56  
57  
58 326 13. Bhargava M, Cheung CYL, Sabanayagam C, Kawasaki R, Harper CA, Lamoureux EL,  
59  
60 327 et al. Accuracy of diabetic retinopathy screening by trained non-physician graders

1  
2  
3 328 using non-mydriatic fundus camera. *Singapore Med J.* 2012;53(11):715–9.  
4  
5

6  
7 329 Figure 1: Flow diagram to illustrate enrolment of patients and management of images in Phase I from  
8  
9 330 August to October 2018 (Initial grading performance analysis). Level 1 and level 2 graders graded the  
10  
11 331 same set of photographs and level 3 graders graded a subset of these photographs: All disagreements  
12  
13 332 between Level 1 and 2 graders and a 40% random sample of all images.  
14  
15

16 333 Figure 2: Flow diagram illustrating the enrolment of patients and management of images included in  
17  
18 334 Phase II from January 2020 to October 2020 (Follow-up grading performance analysis after re-  
19  
20 335 training). Level 1 and level 2 graders graded the same set of photographs and level 3 graders graded a  
21  
22 336 subset of these images.  
23  
24  
25

26 337  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

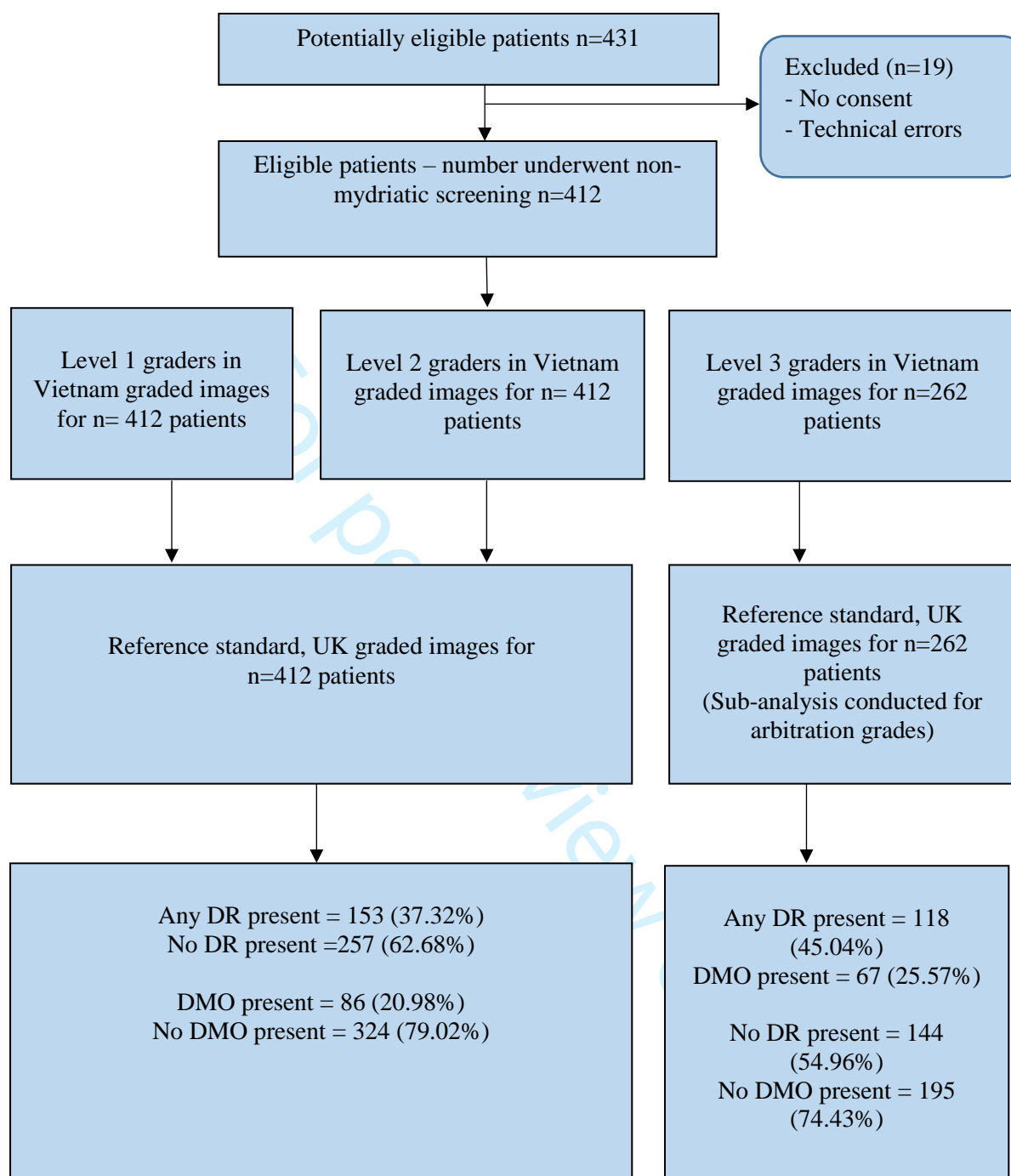


Figure 1: Flow diagram to illustrate enrolment of patients and management of images in Phase I from August to October 2018 (Initial grading performance analysis). Level 1 and level 2 graders graded the same set of photographs and level 3 graders graded a subset of these photographs: All disagreements between Level 1 and 2 graders and a 40% random sample of all images.

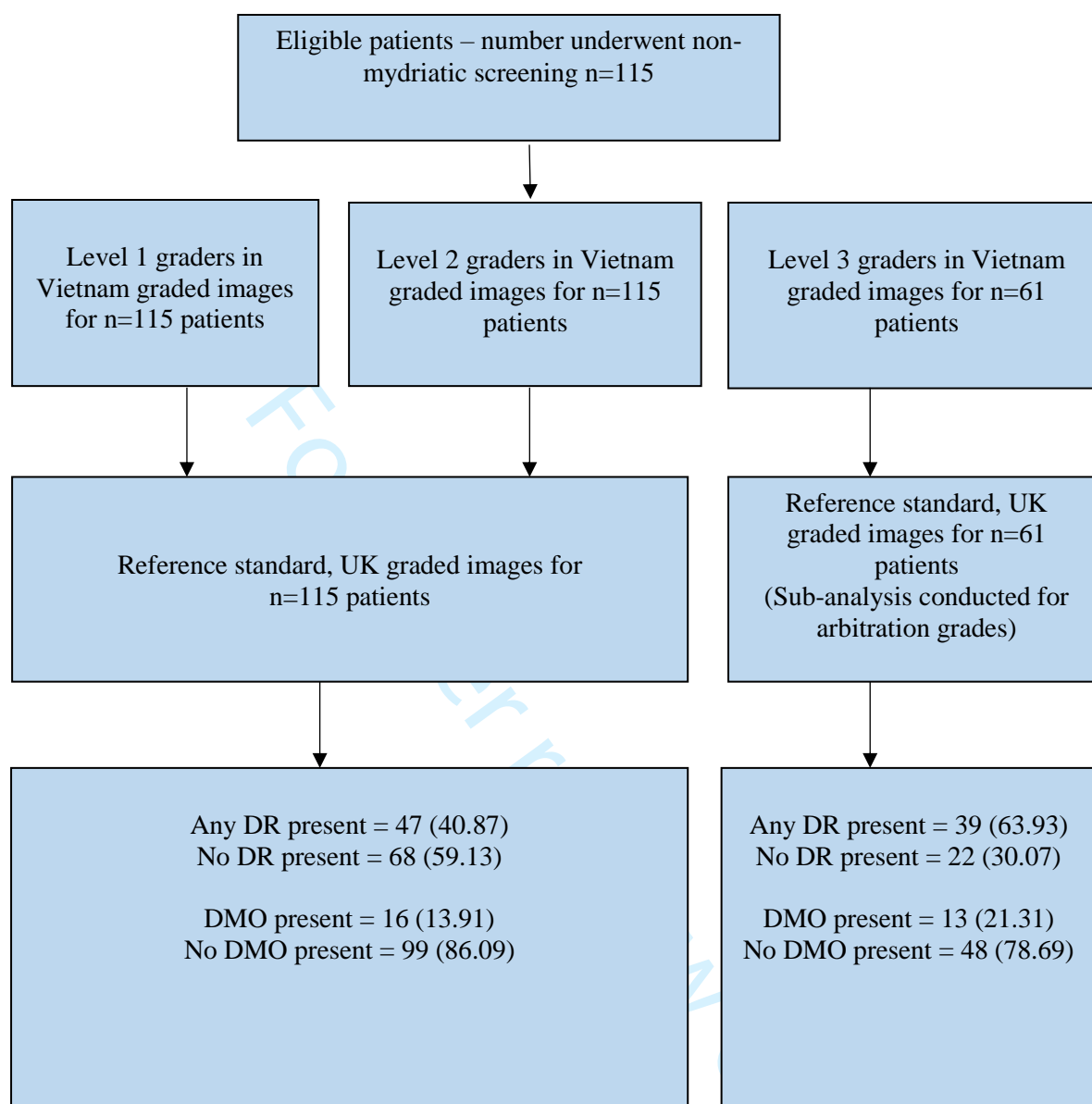


Figure 2: Flow diagram illustrating the enrolment of patients and management of images included in Phase II from January 2020 to October 2020 (Follow-up grading performance analysis after re-training). Level 1 and level 2 graders graded the same set of photographs and level 3 graders graded a subset of these images.



## Supplementary Material

**Table S1: Reference standards intra-rater agreement score using kappa statistic (first attempt versus second attempt)**

|  | <b>Intra-rater agreement (reference standard, UK), k (95% CI) (by eyes, n=106)</b> | <b>Intra-rater agreement (reference standard, UK), k (95% CI) (by worst eye, n=53)</b> |
|--|--|--|
| <b>Overall Diabetic Retinopathy Grading:</b>   |  |  |
| <b>Any DR</b>  | 0.96 (0.91, 1.00)  | 0.92 (0.82, 1.00)  |
| <b>Treatable DR</b>  | 0.81 (0.60, 1.00)  | 0.74 (0.47, 1.00)  |
| <b>Referable Maculopathy</b>   | 0.97 (0.92, 1.00)  | 1.00 (1.00, 1.00)  |
| Abbreviations: CI=confidence interval, k=kappa, DR=Diabetic retinopathy, DMO=diabetic macular oedema<br>Any DR defined as R1, R2, R3s, R3a and U<br>Treatable DR defined as R3a<br>Referable DMO defined as M1 and U |  |  |

**Table S2 Using kappa statistic to determine the inter-rater agreement between the reference standard and one senior grader from QUB grading centre**

|  | <b>Inter-rater agreement (reference standard vs a senior grader QUB), k (95% CI) (by eyes, n=106)</b> | <b>Inter-rater agreement (reference standard vs a senior grader QUB (by worst), k (95% CI) (by worst eye, n=53)</b> |
|--|---|---|
| <b>Overall Diabetic Retinopathy Grading:</b>   |   |   |
| <b>Any DR</b>  | 0.79 (0.67, 0.91)   | 0.74 (0.55, 0.92)   |
| <b>Treatable DR</b>  | 0.71 (0.48, 0.95)   | 0.68 (0.39, 0.97)   |
| <b>Referable Maculopathy</b>   | 0.75 (0.61, 0.90)   | 0.74 (0.55, 0.93)   |
| Abbreviations: CI=confidence interval, k=kappa, DR=Diabetic retinopathy, DMO=diabetic macular oedema<br>Any DR defined as R1, R2, R3s, R3a and U<br>Treatable DR defined as R3a<br>Referable DMO defined as M1 and U |   |   |

### Intra and inter-grader agreement

To ensure there was good intra-grader reliability as a reference standard, a stratified random sample of images were regraded. There was approximately one month between the first and second attempts to reduce the possibility of bias caused by memory. Additionally, inter-grader agreement was calculated using kappa to ensure there was good grading agreement between the reference standard and one senior grader from the Ophthalmic Reading Centre at QUB, Belfast. Any disagreements were discussed with

a retinal specialist until consensus was reached. Overall, the intra-grader agreement and inter-grader agreement ranged from substantial to almost perfect.

**Table S3: The prevalence of any diabetic retinopathy (DR), referable DR, any maculopathy and ungradable cases with the reference grader from Phase I and Phase II**

| Diabetic Retinopathy grades  | Phase I     | Phase II (post remedial training) | P-Value |
|--|-------------|-----------------------------------|---------|
| R0 (n,%)   | 257 (62.68) | 68 (59.13)                        | P=0.347 |
| R1 (n,%)   | 100 (24.39) | 32 (27.83)                        |         |
| R2 (n,%)   | 11 (2.68)   | 2 (1.74)                          |         |
| R3a (n,%)  | 10 (2.44)   | 7 (6.09)                          |         |
| R3s (n,%)  | 1 (0.24)    | 0 (0.00)                          |         |
| U (n,%)  | 31 (7.56)   | 6 (5.22)                          |         |
| <b>Any DR</b>  |             |                                   |         |
| - Yes (n,%)  | 153 (37.32) | 47 (40.87)                        | P=0.488 |
| - No (n,%)   | 257 (62.68) | 68(59.12)                         |         |
| <b>Referable DR</b>  |             |                                   |         |
| - Yes (n,%)  | 52 (12.68)  | 15 (13.04)                        | P=0.918 |
| - No (n,%)   | 358 (87.32) | 100 (86.96)                       |         |
| <b>Any DMO</b>   |             |                                   |         |
| - M0 (n,%)   | 324 (79.02) | 99 (86.09)                        | P=0.173 |
| - M1 (n,%)   | 43 (10.49)  | 10 (8.70)                         |         |
| - U (n,%)  | 43 (10.49)  | 6 (5.22)                          |         |
| <b>Abbreviations:</b> DR=diabetic retinopathy, DMO=Diabetic Macular Oedema, U=ungradable<br>Chi-Squares used to test significance. |             |                                   |         |

| Section & Topic          | No         | Item   | Reported on page #          |
|--------------------------|------------|--|-----------------------------|
| <b>TITLE OR ABSTRACT</b> |            |  |                             |
|                          | <b>1</b>   | Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)  | 1                           |
| <b>ABSTRACT</b>          |            |  |                             |
|                          | <b>2</b>   | Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)                                 | 3-4                         |
| <b>INTRODUCTION</b>      |            |  |                             |
|                          | <b>3</b>   | Scientific and clinical background, including the intended use and clinical role of the index test   | 5-6                         |
|                          | <b>4</b>   | Study objectives and hypotheses  | 5                           |
| <b>METHODS</b>           |            |  |                             |
| <i>Study design</i>      | <b>5</b>   | Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)     | 5                           |
| <i>Participants</i>      | <b>6</b>   | Eligibility criteria   | 5-6                         |
|                          | <b>7</b>   | On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)                 | 5-6                         |
|                          | <b>8</b>   | Where and when potentially eligible participants were identified (setting, location and dates)   | 5-6                         |
|                          | <b>9</b>   | Whether participants formed a consecutive, random or convenience series  | 6                           |
| <i>Test methods</i>      | <b>10a</b> | Index test, in sufficient detail to allow replication  | 5-6                         |
|                          | <b>10b</b> | Reference standard, in sufficient detail to allow replication  | 5-6                         |
|                          | <b>11</b>  | Rationale for choosing the reference standard (if alternatives exist)  | 5-6                         |
|                          | <b>12a</b> | Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory         | 8                           |
|                          | <b>12b</b> | Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory | 8                           |
|                          | <b>13a</b> | Whether clinical information and reference standard results were available to the performers/readers of the index test                                 | 6                           |
|                          | <b>13b</b> | Whether clinical information and index test results were available to the assessors of the reference standard  | 6                           |
| <i>Analysis</i>          | <b>14</b>  | Methods for estimating or comparing measures of diagnostic accuracy  | 8                           |
|                          | <b>15</b>  | How indeterminate index test or reference standard results were handled  | Supplementary file          |
|                          | <b>16</b>  | How missing data on the index test and reference standard were handled   | NA                          |
|                          | <b>17</b>  | Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory  | 8                           |
|                          | <b>18</b>  | Intended sample size and how it was determined   | NA                          |
| <b>RESULTS</b>           |            |  |                             |
| <i>Participants</i>      | <b>19</b>  | Flow of participants, using a diagram  | 9-10                        |
|                          | <b>20</b>  | Baseline demographic and clinical characteristics of participants  | 5-6,8                       |
|                          | <b>21a</b> | Distribution of severity of disease in those with the target condition   | 9-10,<br>Supplementary file |
|                          | <b>21b</b> | Distribution of alternative diagnoses in those without the target condition  | 9-10,<br>Supplementary file |
|                          | <b>22</b>  | Time interval and any clinical interventions between index test and reference standard   | 6-7                         |
| <i>Test results</i>      | <b>23</b>  | Cross tabulation of the index test results (or their distribution) by the results of the reference standard  | 10-11                       |
|                          | <b>24</b>  | Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)  | 10-11                       |
|                          | <b>25</b>  | Any adverse events from performing the index test or the reference standard  | NA                          |
| <b>DISCUSSION</b>        |            |  |                             |
|                          | <b>26</b>  | Study limitations, including sources of potential bias, statistical uncertainty, and generalisability  | 15                          |
|                          | <b>27</b>  | Implications for practice, including the intended use and clinical role of the index test  | 15-16                       |
| <b>OTHER INFORMATION</b> |            |  |                             |
|                          | <b>28</b>  | Registration number and name of registry   | NA                          |

|    |   |    |
|----|---|----|
| 29 | Where the full study protocol can be accessed         | NA |
| 30 | Sources of funding and other support; role of funders | 16 |

For peer review only



# BMJ Open

## The impact of targeted diabetic retinopathy training for graders in Vietnam and the implications for future diabetic retinopathy screening programmes: a diagnostic test accuracy study

|                                 |   |
|---------------------------------|---|
| Journal:                        | <i>BMJ Open</i>   |
| Manuscript ID                   | bmjopen-2021-059205.R1  |
| Article Type:                   | Original research   |
| Date Submitted by the Author:   | 13-May-2022   |
| Complete List of Authors:       | Curran, Katie; Queen's University Belfast School of Medicine Dentistry and Biomedical Sciences, Centre of Public Health<br>Congdon, Nathan; Queen's University of Belfast; ORBIS International<br>Hoang, Tung; Hanoi Medical University, Department of Ophthalmology; The University of Sydney Save Sight Institute<br>Lohfeld, Lynne; Queen's University of Belfast; Eye Hospital of Wenzhou Medical University<br>Nguyen, Van; Orbis International in Vietnam<br>Nguyen, Hue; Orbis International in Vietnam<br>Nguyen, Quan; Ho Chi Minh Eye Hospital, Vitreo-Retina Department<br>Dardis, Catherine; Belfast Health and Social Care Trust, Department of Ophthalmology<br>Virgili, Gianni ; Queen's University of Belfast; Belfast Health and Social Care Trust, Department of Ophthalmology<br>Piyasena, Prabhath; Queen's University of Belfast<br>Tran, Huong; Orbis International in Vietnam<br>Salongcay, Recivall; Queen's University Belfast<br>Tung, Mai; Ho Chi Minh Eye Hospital, Vitreo-Retina Department<br>Peto, Tunde; Queen's University of Belfast; Belfast Health and Social Care Trust, Department of Ophthalmology |
| <b>Primary Subject Heading</b>: | Ophthalmology   |
| Secondary Subject Heading:      | Diabetes and endocrinology, Medical education and training  |
| Keywords:                       | Diabetic retinopathy < DIABETES & ENDOCRINOLOGY, OPHTHALMOLOGY, PUBLIC HEALTH, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT   |
|                                 |   |

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3 1 **Title: The impact of targeted diabetic retinopathy training for graders in Vietnam and**  
4  
5 2 **the implications for future diabetic retinopathy screening programmes: a diagnostic test**  
6  
7 3 **accuracy study**  
8  
9

10  
11 4 **Short title:** Training diabetic retinopathy graders in Vietnam  
12  
13

14  
15 5 **Authors:**  
16

17  
18 6 \*Katie Curran, MPH, PhD<sup>1</sup> ([k.curran@qub.ac.uk](mailto:k.curran@qub.ac.uk))  
19

20  
21 7 Nathan Congdon, MD, MPH<sup>1,2,3</sup> ([Nathan.Congdon@orbis.org](mailto:Nathan.Congdon@orbis.org))  
22  
23

24  
25 8 Tung Thanh Hoang<sup>4,5</sup> ([tung.hoang@sydney.edu.au](mailto:tung.hoang@sydney.edu.au))  
26  
27

28  
29 9 Lynne Lohfeld, MPH, PhD<sup>1,6</sup> ([l.lohfeld@qub.ac.uk](mailto:l.lohfeld@qub.ac.uk))  
30  
31

32  
33 10 Van Thu Nguyen, BA<sup>7</sup> ([van.nguyen@orbis.org](mailto:van.nguyen@orbis.org))  
34  
35

36  
37 11 Hue Thi Nguyen, MD, MPH<sup>7</sup> ([hue.nguyen@orbis.org](mailto:hue.nguyen@orbis.org))  
38  
39

40  
41 12 Quan Nhu Nguyen, MD, PhD<sup>8</sup> ([bsquan2009@gmail.com](mailto:bsquan2009@gmail.com))  
42  
43

44  
45 13 Catherine Dardis, MSc<sup>9</sup> ([cathy831oph@gmail.com](mailto:cathy831oph@gmail.com))  
46  
47

48  
49 14 [Gianni Virgili MD<sup>1,9</sup>](mailto:G.Virgili@qub.ac.uk) ([G.Virgili@qub.ac.uk](mailto:G.Virgili@qub.ac.uk))  
50  
51

52  
53 15 [Prabhath Piyasena, MB BS, PhD<sup>1</sup>](mailto:m.piyasena@qub.ac.uk), ([m.piyasena@qub.ac.uk](mailto:m.piyasena@qub.ac.uk))  
54  
55

56  
57 16 Huong Tran, MPA<sup>7</sup> ([Huong.Tran@orbis.org](mailto:Huong.Tran@orbis.org))  
58  
59

60  
61 17 Recivall Salongcay, MD, MPM<sup>1</sup> ([R.SALONGCAY@qub.ac.uk](mailto:R.SALONGCAY@qub.ac.uk))  
62  
63

64  
65 18 Mai Quoc Tung, MD, PhD<sup>8</sup> ([maitungtn@gmail.com](mailto:maitungtn@gmail.com))  
66  
67

1  
2  
3 19 Tunde Peto, MD, PhD<sup>1,9</sup> ([T.Peto@qub.ac.uk](mailto:T.Peto@qub.ac.uk))  
4  
5

6  
7 **Affiliations:**  
8  
9

10 21 <sup>1</sup>Centre for Public Health, Queen's University of Belfast, Belfast, UK  
11  
12

13 22 <sup>2</sup>Orbis International, New York, USA  
14  
15

16 23 <sup>3</sup>Zhongshan Ophthalmic Centre, Guangzhou, China  
17  
18

19 24 <sup>4</sup>Ophthalmology Department, Hanoi Medical University, Hanoi, Vietnam  
20  
21  
22

23 25 <sup>5</sup>Save Sight Institute, The University of Sydney School of Medicine, Sydney, Australia  
24  
25  
26

27 26 <sup>6</sup>Eye Hospital of Wenzhou Medical University, Wenzhou, Zhejiang Province, China  
28  
29

30 27 <sup>7</sup>Orbis International in Vietnam, Hanoi, Vietnam  
31  
32

33 28 <sup>8</sup>Vitreo-Retina Department, Ho Chi Minh Eye Hospital, Ho Chi Minh City, Vietnam  
34  
35  
36

37 29 <sup>9</sup>Department of Ophthalmology, Belfast Health and Social Care Trust, NI  
38  
39  
40

41 30 **\*Corresponding author**  
42  
43

44 31 Katie Curran  
45  
46

47 32 Centre for Public Health  
48  
49

50 33 Institute of Clinical Sciences, A  
51  
52

53 34 Queen's University Belfast  
54  
55

56 35 BT12 6BA  
57  
58

59 36 Email: [k.curran@qub.ac.uk](mailto:k.curran@qub.ac.uk)  
60



1  
2  
3 37 **Keywords:** Diabetic retinopathy, diabetic retinopathy training, screening  
4  
5

6 38 **Word Count:** 2545  
7  
8

9 39 **Abstract**  
10  
11

12  
13 40 **Objectives:** To compare the accuracy of trained level 1 diabetic retinopathy (DR) graders  
14  
15 41 (nurses, endocrinologists, one general practitioner), level 2 graders (mid-level  
16  
17 42 ophthalmologists) and level 3 graders (senior ophthalmologists) in Vietnam against a  
18  
19 43 reference standard from the UK, and assess the impact of supplementary targeted grader  
20  
21 44 training.  
22  
23

24  
25 45 **Design:** Diagnostic test accuracy study.  
26  
27

28  
29 46 **Setting:** Secondary care hospitals in Southern Vietnam  
30  
31

32  
33 47 **Participants:** DR training was delivered to Vietnamese graders in February 2018 by National  
34  
35 48 Health System (NHS) UK graders. Two-field retinal images (412 patient images) were  
36  
37 49 graded by 14 trained graders in Vietnam between August-October 2018 and then re-graded  
38  
39 50 retrospectively by an NHS-certified reference standard UK optometrist (Phase I). Further  
40  
41 51 directed DR training based on Phase I results was delivered to graders in November 2019.  
42  
43 52 After training, a randomised subset of images from January-October 2020 (115 patient  
44  
45 53 images) was graded by 6 of the original cohort (Phase II). The reference grader re-graded all  
46  
47 54 images from Phase I and II retrospectively in masked fashion.  
48  
49  
50

51  
52 55 **Primary and secondary outcome measures:** Sensitivity was calculated at the two different  
53  
54 56 time points and Chi-Squared was used to test significance.  
55  
56  
57  
58  
59  
60

1  
2  
3 57 **Results:** In Phase I, the sensitivity for detecting any DR for all grader groups in Vietnam  
4  
5 58 was low and improved in Phase II after additional training was delivered. The greatest  
6  
7 59 improvement was seen among level 1 graders ( $P<0.001$ ) and the lowest improvement was  
8  
9 60 observed among level 3 graders ( $P=0.326$ ). There was an improvement in sensitivity for  
10  
11 61 detecting any DR and referable diabetic macular oedema between all grader levels and whilst  
12  
13 62 the differences were statistically significant, the post-training values were suboptimal (41.8%  
14  
15 63 to 61.5%). The main disagreement was the detection of ungradable images.

16  
17  
18  
19  
20 64 **Conclusions:** This study demonstrates that targeted training interventions can improve  
21  
22 65 accuracy of DR grading in a low-middle income country. These findings have important  
23  
24 66 implications for improving service delivery in DR screening programmes in low-resource  
25  
26 67 settings.

## 27 28 29 30 31 68 **Article Summary**

### 32 33 34 69 **Strengths and limitations of this study**

- 35  
36  
37 70 • Graders in Vietnam were trained to detect DR based on the UK's DR screening model
- 38  
39 71 • This study describes the impact of a training intervention to improve DR grading in  
40  
41 72 Vietnam
- 42  
43 73 • Reinforcing training to identify ungradable images has been acknowledged.
- 44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 77 **Introduction**

78 The prevalence of diabetes among adults in Vietnam is approximately 6% and has almost  
79 doubled over the past decade.[1] Early detection through diabetic eye screening programmes  
80 (DESPs) is important to reduce the risk of avoidable blindness due to diabetic retinopathy  
81 (DR). Since the introduction of systematic DESPs in the UK, a high-income country (HIC),  
82 diabetic retinopathy (DR) is no longer the leading cause of blindness among working age  
83 adults.[2] The key to such successful DESPs is implementing accurate, innovative and cost-  
84 effective models tailored to fit healthcare systems and contexts.

85 Investing in training personnel to increase human resources and procuring appropriate  
86 diagnostic and treatment equipment are essential to ensure that service providers can deliver  
87 optimum care for people with DR. In low-middle income countries (LMICs), there is often  
88 insufficient capacity to implement robust DESPs due to the lack of skilled human resources  
89 and infrastructure.[3,4] In Vietnam, there are only 14 ophthalmologists per million population  
90 compared to 49 per million in the UK.[5]

91 All screening programmes must provide evidence of their ability to detect the targeted  
92 condition and ensure that the service performs efficiently to improve screening accuracy  
93 when it falls short. To date, there is insufficient evidence on DR grading accuracy using non-  
94 mydriatic digital imaging by trained graders in LMICs, and even less about the capacity of  
95 DESPs in LMICs to improve where poor accuracy is detected. The current retrospective  
96 study is designed to assess accuracy of a range of graders in a non-governmental organisation  
97 (NGO)-supported DESP in Vietnam, and to study the efficacy of a quality-improvement  
98 intervention.

## 99 **Methods**

1  
2  
3 100 **Study participants:** The 14 participants from Vietnam in Phase I included: Level 1 DR  
4  
5 101 graders (6 nurses, 1 general practitioner and 2 endocrinologists, all with < 1 year grading  
6  
7 102 experience, 55.6% female), Level 2 DR graders (3 newly-qualified ophthalmologists with < 1  
8  
9 103 year formal DR grading experience, 100% female), and Level 3 DR graders (2 senior  
10  
11 104 ophthalmologists with >5 years' experience providing treatment for sight threatening DR, but  
12  
13 105 with <1 year formal DR grading experience, 100% male). In Phase II, 6/14 graders (3 Level  
14  
15 106 1, 2 Level 2, 1 Level 3) from Phase I were included. The reference standard from the UK  
16  
17 107 (KC) was a fully-qualified optometrist trained in DR grading and certified by the UK NHS  
18  
19 108 DESP.[6] Vietnamese Level 1, 2 and 3 graders are equivalent to primary, secondary and  
20  
21 109 arbitration graders, respectively, in UK DESPs.[7] In the current study, Vietnamese Level 1  
22  
23 110 and Level 2 graders graded all fundus images for DR. All images having disagreement  
24  
25 111 between graders, and an additional randomly-selected 40% of all images, were sent for  
26  
27 112 arbitration grading by Level 3 graders in Vietnam. All graders in Vietnam were masked to  
28  
29 113 any prior diagnoses or grades of the reference standard, while the reference standard was also  
30  
31 114 masked to results of grading in Vietnam. Fundus images were graded for 412 patients in  
32  
33 115 phase I and 115 patients in phase II (Figure 1 and figure 2).

34  
35  
36  
37  
38  
39  
40  
41 116 **DR training for graders in Vietnam:** As part of a DESP project supported by NGO Orbis  
42  
43 117 International, a team of five Vietnamese doctors and medical administrators visited a  
44  
45 118 Northern Ireland (NI) DESP in September 2017 to receive training on screening, programme  
46  
47 119 administration and quality control methods. In February 2018, a senior UK NHS grader from  
48  
49 120 the Belfast Trust (CD) and a fully-qualified optometrist, trained in DR grading and certified  
50  
51 121 by the NHS (KC), visited Vietnam to deliver DR training to graders involved in the DESPs.  
52  
53 122 (See supplementary material, figure S1 for training timeline). Training focused on ocular  
54  
55 123 anatomy, retinal diseases, DR signs and grading (based on the UK National Screening  
56  
57  
58  
59  
60

1  
2  
3 124 Committee (NSC) classification system), and appropriate referral pathways and management  
4  
5 125 (Supplementary material, Table S1).[8]  
6  
7  
8

9 126 **Image acquisition and management:** Images were captured by trained nurses and  
10  
11 127 technicians in Vietnam. Two-field, 45° digital colour photographs (one disc-centred and one  
12  
13 128 macula-centred) were taken using a tabletop non-mydratic fundus camera (Canon CR2-AF,  
14  
15 129 Canon Medical Systems. Europe), in accordance with the UK's NHS DESP.[9] Nurses and  
16  
17 130 technicians were trained to repeat inadequate images as a quality control measure and take  
18  
19 131 anterior segment photographs where adequate fundus images were not possible. Images were  
20  
21 132 anonymised and uploaded to a cloud-based software system (Spectra)® for analysis by  
22  
23 133 trained DR graders in Vietnam. The images were transferred to a Queen's University Belfast  
24  
25 134 (QUB) server for re-grading by the reference standard.  
26  
27  
28  
29

30  
31 135 **Assessment of gradeability:** Image quality was defined as 'adequate' or 'inadequate' in  
32  
33 136 accordance with NHS DESP guidelines as outlined below;  
34  
35

- 36  
37 137 • Adequate disc-centred image: complete optic disc >2DD from edge of image and fine  
38  
39 138 vessels visible on surface of the disc.[9]  
40  
41 139 • Adequate macula-centred image: centre of fovea >2DD from edge of image and  
42  
43 140 vessels visible within 1DD of centre of fovea.[9]  
44  
45  
46

47 141 The disc-centred and macula-centred images for each eye were viewed as a pair and graded at  
48  
49 142 an individual eye level. The presence of DR and diabetic macular oedema (DMO) was also  
50  
51 143 determined at a patient level and based on the worst affected eye. Participants with ungradable  
52  
53 144 images were referred for further slit-lamp examination. Where images were considered  
54  
55 145 inadequate but referable disease was detectable, the referable grade was recorded and the  
56  
57 146 patients were moved onto the appropriate referable grade pathway.[9]  
58  
59  
60

1  
2  
3 147 Consecutive patients diagnosed with diabetes and undergoing evaluation for possible DR at  
4  
5 148 Ho Chi Minh City General Hospital and Ho Chi Minh Eye Hospital (tertiary hospitals), Tien  
6  
7 149 Giang General Hospital (provincial hospital) and Cai Ba General Hospital (district hospital)  
8  
9  
10 150 in Vietnam were recruited. Fundus images from August to October 2018 (Phase I) were  
11  
12 151 graded by 14 graders in Vietnam and then re-graded retrospectively by a reference standard  
13  
14 152 from the UK in Phase I. Targeted remedial training, based on specific findings from the  
15  
16  
17 153 Phase I analysis, was delivered in March 2019 and November 2019 by UK graders and Orbis.  
18  
19 154 (Supplementary material, figure S1) Additionally, regular testing and training for quality  
20  
21 155 assurance purposes was also introduced, similar to UK DESP models. To evaluate the impact  
22  
23 156 of this quality-improvement intervention, a new subset of images was graded by six of the  
24  
25  
26 157 original cohort of graders between January-October 2020 (Phase II) and re-graded by the  
27  
28 158 reference standard from the UK (KC) in September 2021.

31  
32 159 **Statistical analysis:** Data were entered into Microsoft Excel version 16.0 and then  
33  
34 160 transferred to Stata 17.0 (StataCorp LLC) for analysis. Intra and inter-grader agreement was  
35  
36 161 calculated using kappa and a stratified random sampling technique was utilised to ensure a  
37  
38 162 representative sample of images was re-graded (Supplementary tables S2 and S3). Diagnostic  
39  
40 163 test accuracy (DTA) comparing graders in Vietnam with the UK reference standard was  
41  
42 164 assessed by calculating sensitivity, specificity, positive predictive values (PPV) and negative  
43  
44 165 predictive values (NPV). Sensitivity was calculated at the two different time points (Phase I  
45  
46 166 and Phase II) and Chi Squared was used to test significance.

47  
48  
49  
50  
51 167 Patients or the public were not involved in the design, or conduct, or reporting, or  
52  
53 168 dissemination plans of our research.

## 54 55 56 57 169 **Results**

1  
2  
3 170 **Patient characteristics:** In Phase I, 65.4% of patients were female with a mean age 59.4  
4  
5 171 years. In Phase II, 40.0% were female with a mean age of 59.8 years. Figures 1 and 2  
6  
7 172 describe enrolment of patients and capture and grading of images in Phase I and II of the  
8  
9 173 study respectively.  
10  
11  
12

13 174  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

175 Figure 1  
176

For peer review only



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

177 Figure 2

178

For peer review only

179 **Initial grading performance analysis (Phase I):** The sensitivity for detecting any DR was  
 180 low against the reference standard in the UK for all grader groups in Vietnam. The sensitivity  
 181 for detecting referable DR and referable DMO was even lower for all grader groups (Table  
 182 1). (See supplementary table S4 for prevalence of DR and table S5 for grading performance,  
 183 excluding ungradable images)

184 **Table 1. Diagnostic test accuracy of DR graders in Vietnam against a reference**  
 185 **standard from the UK, including ungradable images.**

|   | <b>Level 1 graders<br/>(n=410 patient<br/>images)*</b> | <b>Level 2 graders<br/>(n=410 patient<br/>images)*</b> | <b>Level 3 graders<br/>(n=260 patient<br/>images)†</b> |
|---|--|--|--|
| <b>Any DR</b>   |  |  |  |
| Sensitivity (%) (95% CI)  | 41.8 (33.9, 50.1)                                      | 42.5 (34.5, 50.7)                                      | 42.2 (33.1, 51.8)                                      |
| Specificity (%) (95% CI)  | 87.9 (83.3, 91.7)                                      | 98.8 (96.6, 99.8)                                      | 100 (97.5, 100)  |
| PPV (%) (95% CI)  | 67.4 (57.0, 76.6)                                      | 95.6 (87.6, 99.1)                                      | 100 (92.7, 100)  |
| NPV (%) (95% CI)  | 71.7 (66.4, 76.7)                                      | 74.3 (69.3, 78.8)                                      | 68.2 (61.5, 74.5)                                      |
| <b>Referable DR</b>   |  |  |  |
| Sensitivity (%) (95% CI)  | 19.2 (9.63, 32.5)                                      | 13.5 (5.59, 25.8)                                      | 10.5 (2.94, 24.8)                                      |
| Specificity (%) (95% CI)  | 97.2 (94.9, 98.7)                                      | 100 (99.0, 100)  | 99.5 (97.5, 100)                                       |
| PPV (%) (95% CI)  | 50.0 (27.2, 72.8)                                      | 100 (59.0, 100)  | 80.0 (28.4, 99.5)                                      |
| NPV (%) (95% CI)  | 89.2 (85.7, 92.1)                                      | 88.8 (85.3, 91.7)                                      | 86.7 (81.9, 90.6)                                      |
| <b>Referable DMO</b>  |  |  |  |
| Sensitivity (%) (95% CI)  | 5.8 (1.91, 13.0)                                       | 20.9 (12.9, 31.0)                                      | 16.9 (8.76, 28.3)                                      |
| Specificity (%) (95% CI)  | 97.2 (94.8, 98.7)                                      | 99.4 (97.8, 99.9)                                      | 100 (98.1, 100)  |
| PPV (%) (95% CI)  | 35.7 (12.8, 64.9)                                      | 90.0 (68.3, 98.8)                                      | 100 (71.5, 100)  |
| NPV (%) (95% CI)  | 79.5 (75.2, 83.4)                                      | 82.6 (78.4, 86.2)                                      | 78.3 (72.7, 83.3)                                      |
| <p><b>Abbreviations:</b> UK = United Kingdom, DR = Diabetic Retinopathy, DMO = Diabetic Macular Oedema, CI = Confidence Intervals,<br/>           Grading criteria: UK National Diabetic Eye Screening Programme (NDESP) classification system (See supplementary material, Table S1 for more details).<br/>           Any DR, is defined as grades R1, R2, R3s, R3a and U.<br/>           Referable DR is defined as grades R2, R3a and U<br/>           Referable DMO is defined as grades M1 and U<br/>           *Missing (n=2, 0.5%), †missing (n=2, 0.8%)<br/>           Sensitivity is the ability of a test to correctly identify patients with a disease and specificity is the ability of a test to correctly identify people without the disease</p> |  |  |  |

Positive predictive value (PPV) is the proportion of those who test positive who have the condition (true positives) and negative predictive value (NPV) is the proportion of those who test negative who do not have the condition (true negatives).

186

187 **Subsequent grading performance after retraining (Phase II):** Subsequently, a further  
188 subset of images from 115 consecutive patients from January to October 2020 were graded  
189 by 6 of the original cohort of 14 Vietnamese graders, and were regraded in the UK to  
190 evaluate graders' performance after targeted training was delivered and quality control  
191 measures were instituted. The greatest improvement in sensitivity for detecting any DR was  
192 seen among level 1 graders (difference: +45.4%, 95%CI +33.1% to +57.8%; P<0.001). The  
193 specificity increased from 87.9% in phase I to 95.6% in phase II which helps to avoid over  
194 referrals (difference: +7.7%, 95%CI +1.4% to +13.9%; p=0.069). The lowest improvement in  
195 sensitivity for detecting any DR was observed between level 3 graders in Vietnam  
196 (difference; +9.0%, 95%CI: -9.0% to +27.1%; p=0.326), although their specificity remained  
197 100% at phase I and phase II. There was an improvement in sensitivity for detecting DR and  
198 referable DMO between all grader levels and whilst there were statistically significant  
199 differences, sensitivities after training were still insufficient and comprised between about  
200 40% and 61.5% (Table 2). (See supplementary table S4 for prevalence of DR and table S6 for  
201 grading performance, excluding ungradable images).

202

203

204

205

206 **Table 2: Diagnostic test accuracy of DR graders in Vietnam against a reference**  
 207 **standard from the UK after additional DR training was delivered.**

|   | <b>Level 1 graders<br/>(n=115 patient<br/>images)</b> | <b>Level 2<br/>graders (n=115<br/>patient images)</b> | <b>Level 3 graders<br/>(n=62 patient<br/>images)</b> |
|---|---|---|--|
| <b>Any DR</b>   |   |   |  |
| Sensitivity (%) (95% CI)  | 87.2 (74.3, 95.2)                                     | 68.1 (52.9, 80.9)                                     | 51.3 (34.8, 67.6)                                    |
| Specificity (%) (95% CI)  | 95.6 (87.6, 99.1)                                     | 95.6 (87.6, 99.1)                                     | 100 (84.6, 100)                                      |
| PPV (%) (95% CI)  | 93.2 (81.3, 98.6)                                     | 91.4 (76.9, 98.2)                                     | 100 (83.2, 100)                                      |
| NPV (%) (95% CI)  | 91.5 (82.5, 96.8)                                     | 81.3 (71.0, 89.1)                                     | 53.7 (37.4, 69.3)                                    |
| P-value comparing sensitivity to Phase I  | P=0.000   | P=0.002   | P=0.326  |
| <b>Referable DR</b>   |   |   |  |
| Sensitivity (%) (95% CI)  | 53.3 (26.6, 78.7)                                     | 40.0 (16.3, 67.7)                                     | 58.3 (27.7, 84.8)                                    |
| Specificity (%) (95% CI)  | 90.0 (82.4, 95.1)                                     | 93.0 (86.1, 97.1)                                     | 100 (92.7, 100)                                      |
| PPV (%) (95% CI)  | 44.4 (21.5, 69.2)                                     | 46.2 (19.2, 74.9)                                     | 100 (59.0, 100)                                      |
| NPV (%) (95% CI)  | 92.8 (85.7, 97.0)                                     | 91.2 (83, 95.9)                                       | 90.7 (79.7, 96.9)                                    |
| P-value comparing sensitivity to Phase I  | P=0.009   | P=0.022   | P=0.001  |
| <b>Referable DMO</b>  |   |   |  |
| Sensitivity (%) (95% CI)  | 56.3 (29.9, 80.2)                                     | 43.8 (19.8, 70.1)                                     | 61.5 (31.6, 86.1)                                    |
| Specificity (%) (95% CI)  | 97.0 (91.4, 99.4)                                     | 93.9 (87.3, 97.7)                                     | 100 (92.6, 100)                                      |
| PPV (%) (95% CI)  | 75.0 (42.8, 94.5)                                     | 53.8 (25.1, 80.8)                                     | 100 (63.1, 100)                                      |
| NPV (%) (95% CI)  | 93.2 (86.5, 97.2)                                     | 91.2 (83.9, 95.9)                                     | 90.6 (79.3, 96.9)                                    |
| P-value comparing sensitivity to Phase I  | P=0.000   | P=0.051   | P=0.002  |
| <p><b>Abbreviations:</b> UK = United Kingdom, DR = Diabetic Retinopathy, DMO = Diabetic Macular Oedema, CI = Confidence Intervals,<br/>           Grading criteria: UK National Diabetic Eye Screening Programme (NDESP) classification system (See supplementary material, Table S1 for more details).<br/>           Criteria:<br/>           Any DR is defined as grades R1, R2, R3s, R3a and U.<br/>           Referable DR is defined as grades R2, R3a and U<br/>           Referable DMO is defined as grades M1 and U<br/>           Chi-squared used to compare sensitivity between Phase I and II.<br/>           Sensitivity is the ability of a test to correctly identify patients with a disease and specificity is the ability of a test to correctly identify people without the disease<br/>           Positive predictive value (PPV) is the proportion of those who test positive who have the condition (true positives) and negative predictive value (NPV) is the proportion of those who test negative who do not have the condition (true negatives).</p> |   |   |  |

208

## 209 Discussion

1  
2  
3 210 Results from our study demonstrate extremely poor sensitivity and specificity for detecting  
4  
5 211 all levels of DR, especially referable DR, in the early stages of programme delivery. This  
6  
7 212 translates into increased costs to the health care system due to missed opportunities for early  
8  
9 213 treatment and un-necessary examinations for false-positive referrals. The quality of patient  
10  
11 214 care also suffers. Didactic DR training was delivered to graders in Vietnam over a two-year  
12  
13 215 period by trained DR graders from the UK and Vietnam. Training was specifically targeted to  
14  
15 216 address problems identified in the Phase I testing [10], and quality control testing using  
16  
17 217 international test and training (iTAT) were also undertaken. The iTAT is an online platform  
18  
19 218 offering monthly quality assurance and training for graders who work in DR screening. It is a  
20  
21 219 useful platform for graders to improve their skills in the detection of DR from ophthalmic  
22  
23 220 images. In the UK, it is compulsory for graders to complete monthly test sets (each set  
24  
25 221 consisting of 20 retinal images with a range of DR severities). If graders fall below the agreed  
26  
27 222 threshold, additional training and support is provided.[7] This study demonstrates that these  
28  
29 223 steps led to improved grading accuracy for all classes of patients and graders; however,  
30  
31 224 results remain suboptimal for a screening programme. We found that the main discordance  
32  
33 225 was the graders' ability to detect ungradable images, therefore, targeted training must be  
34  
35 226 given to ensure patients are referred to the next level (slit-lamp examination). According to  
36  
37 227 the UK National Institute for Clinical Excellence (NICE) guidelines, DR screening tests must  
38  
39 228 have at least 80% sensitivity and 95% specificity with a technical failure of 5% or less. These  
40  
41 229 requirements may not be applicable to LMICs, especially at the start of the programme where  
42  
43 230 a relatively low number of patients are being screened.

44  
45 231 Results can be poor in these settings for a variety of reasons, quality assessment is crucial,  
46  
47 232 and programmatic changes based on models such as the UK DESP can be successful in  
48  
49 233 enhancing grader accuracy in LMICs settings. However, it is fundamental for countries to  
50  
51 234 adapt their own DR classification system and referral pathways to meet their requirements.

1  
2  
3 235 As an example, the UK system (England, Wales and Northern Ireland) uses the grade M0 for  
4  
5 236 no maculopathy and M1 for referable maculopathy. In Scotland, M0 denotes no  
6  
7  
8 237 maculopathy, M1 observable maculopathy and M2 referable maculopathy allowing some  
9  
10 238 monitoring of maculopathy to take place on screening level. This reduces the burden on the  
11  
12 239 hospital system. The implication for LMICs is that being aware of hospital capacity at the  
13  
14 240 planning stage might mean that they need to safely adapt an accepted grading system to their  
15  
16 241 needs. Most importantly, the role of affiliated hospitals (and partnerships, coordination  
17  
18 242 among training institutions and practical hospitals) are crucial for DR grading quality  
19  
20 243 improvement.  
21  
22  
23  
24

25 244 Studies in LMICs and HICs have assessed the accuracy of non-medical graders and medical  
26  
27 245 graders in the detection of DR and found that both grader types are capable of achieving  
28  
29 246 moderate-high sensitivity for detecting DR.[11-15] Some studies have described what  
30  
31 247 training interventions were used to train their graders and key elements may be incorporated  
32  
33 248 into our training programme in the future.[15-16]  
34  
35  
36  
37

38 249 In the UK, the DR grading course by the Gloucestershire Retinal Education Group is is  
39  
40 250 required for grader certification. The high costs of this course may be more challenging in  
41  
42 251 LMICs due to limited funding.[6]  
43  
44  
45

46 252 **Strengths:** This study describes the impact of a training intervention to improve the quality  
47  
48 253 of DR grading in an LMIC. The inclusion of ungradable images in this study was a logical  
49  
50 254 decision, particularly when the prevalence of cataract (which often renders DR images  
51  
52 255 ungradable) is high in LMICs.[11] Dense cataracts normally obstruct the view of the fundus,  
53  
54 256 making it difficult to obtain clear fundus photographs and assign a DR grade. In these  
55  
56 257 instances, referring patients to an eye clinic for further assessment and treatment as needed is  
57  
58 258 required. Determining sensitivity and specificity at the patient level is also important from a  
59  
60

1  
2  
3 259 DESP implementation perspective. In the UK and Vietnam, both eyes are typically examined  
4  
5 260 for DR and a single outcome is assigned to the patient, as was done here. For these reasons,  
6  
7  
8 261 we feel our analytic approach, and thus results, are relevant to these settings.  
9

10  
11 262 **Limitations:** Limitations for this study have also been acknowledged. Data from this study  
12  
13 263 represent routine clinical practice. In daily DR screening, not all patients undergoing primary  
14  
15 264 (Level 1) and secondary (Level 2) grading proceed to arbitration grading (Level 3). This  
16  
17 265 means a proportion of images were not graded by arbitration graders as outlined in figure 1  
18  
19 266 and figure 2. Second, only 6/14 graders from Phase I were included in Phase II grading;  
20  
21 267 however the distribution of grader levels was similar. Third, though the proportion of patients  
22  
23 268 excluded was small, we are unable to fully characterise the reasons for these exclusions, due  
24  
25 269 to the nature of the study as a programmatic evaluation. Some potential reasons for this are a  
26  
27 270 patient's unwillingness to participate in the study, graders having forgotten to ask for patient  
28  
29 271 consent to participate in the study, and patient inability to comply with image capture. Fourth,  
30  
31 272 pupil status (size and cataract status) was not recorded in this study and this can be important  
32  
33 273 for LMICs. Finally, it was not practical for the UK reference standard to examine patients  
34  
35 274 clinically in Vietnam; however, the method of grading by a certified DR grader or clinical  
36  
37 275 specialist is widely used as the reference standard in many screening programmes.  
38  
39  
40  
41  
42  
43

44 276 **Conclusions:** This paper shows how grading accuracy was particularly low among all grader  
45  
46 277 groups in Vietnam in the first six months of DESP implementation. Many factors may have  
47  
48 278 contributed to poor grader performance, including inadequate training and feedback,  
49  
50 279 insufficient time to participate in quality assurance testing and competing work  
51  
52 280 responsibilities. After additional training, testing and quality assurance systems were  
53  
54 281 implemented in Vietnam, DTA improved among all grader groups, however a significant  
55  
56 282 amount of work is still needed. In particular, training graders to detect ungradable cases is  
57  
58  
59  
60

1  
2  
3 283 crucial. With continuous quality improvement, monthly international test and training,  
4  
5 284 periodic DR workshops and reviewal of certification, we would expect the DR sensitivity and  
6  
7 285 specificity to improve. A qualitative study to determine why the initial training intervention  
8  
9 286 was less successful should be explored. Since further improvements are required,  
10  
11 287 understanding how other countries implement such programmes would be beneficial. Future  
12  
13 288 studies should outline what DR training interventions were used, state relevant training  
14  
15 289 courses and explain what quality assurance measures are in place. The findings from this  
16  
17 290 study are important for DESP programme planners in Vietnam and other LMICs,  
18  
19 291 highlighting the importance of quality monitoring and directed re-training as needed.  
20  
21  
22  
23  
24

25 292 **Acknowledgements:** The authors are very grateful to all participants in Vietnam and  
26  
27 293 Northern Ireland who generously gave their time for this study and to Orbis International.  
28  
29

30 294 **Author contributions:** All authors (KC, NC, TTH, LL, VTN, HTN, QHN, CD, GV, PP, HT, RS,  
31  
32 295 MQT, TP) have made substantial contributions to the conception of the design of the work. KC and  
33  
34 296 GV analysed and interpreted the data. All authors (KC, NC, TTH, LL, VTN, HTN, QHN, CD, GV,  
35  
36 297 PP, HT, RS, MQT, TP) read and approved the final manuscript.  
37  
38  
39

40 298 **Funding:** The project was funded by the Department for the Economy (DfE) – Global  
41  
42 299 Challenges Research Fund (GCRF) Awards (Grant number: DFEGCRF17-18/Peto). None of  
43  
44 300 the funders were involved in the design or conduct of the study; preparation, review or  
45  
46 301 approval of the manuscript; or decision to submit the manuscript for publication. Nathan  
47  
48 302 Congdon is supported by the Ulverscroft Foundation (UK) (No award/grant number). Katie  
49  
50 303 Curran is funded by the Wellcome Trust (No award/grant number).  
51  
52  
53

54 304 **Competing Interests:** Nathan Congdon is employed as a Research Director by Orbis  
55  
56 305 International.  
57  
58  
59  
60



1  
2  
3 306 **Patient consent for publication:** Not required.  
4  
5

6  
7 307 **Ethics approval:** This research adhered to the tenets of the Declaration of Helsinki. Ethical  
8  
9 308 approval was granted by the Hanoi Medical University Institutional Review Board in Bio-  
10  
11 309 Medical Research, Vietnam (No. 0518/HMU IRB). Written informed consent was obtained  
12  
13 310 from all participants prior to their being interviewed.  
14  
15

16  
17 311 **Data sharing statement:** All datasets relevant to the study are included in the article or  
18  
19 312 uploaded as supplementary information.  
20  
21

## 22 313 **References**

- 23  
24  
25  
26 314 1. Ngoc NB, Lin ZL, Ahmed W. Diabetes: What challenges lie ahead for Vietnam? *Ann*  
27  
28 315 *Glob Heal.* 2020;86(1):1–9.  
29  
30  
31 316 2. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness  
32  
33 317 certifications in England and Wales in working age adults (16-64 years), 1999-2000  
34  
35 318 with 2009-2010. *BMJ Open.* 2014;4(2):1–6.  
36  
37  
38 319 3. Burgess PI, Msukwa G, Beare NAV. Diabetic retinopathy in sub-Saharan Africa:  
39  
40 320 Meeting the challenges of an emerging epidemic. *BMC Med.* 2013;11(1):1.  
41  
42  
43  
44 321 4. Adriono G, Wang D, Octavianus C, Congdon N. Use of eye care services among  
45  
46 322 diabetic patients in urban Indonesia. *Arch Ophthalmol* (Chicago, Ill 1960).  
47  
48 323 2011;129(7):930–5.  
49  
50  
51 324 5. International Council of Ophthalmology. Number of Ophthalmologists in Practice and  
52  
53 325 Training Worldwide [Internet]. 2012 [cited 2020 Aug 13]. Available from:  
54  
55 326 <http://www.icoph.org/ophthalmologists-worldwide.html>  
56  
57  
58  
59 327 6. Gloucestershire Retinal Education Group. Diabetic Retinopathy Grading Course  
60

- 1  
2  
3 328 [Internet]. 2020. [cited 2021 April] Available from: <https://drscreening.org/diabetic->  
4  
5 329 [retinopathy-grading-course](https://drscreening.org/diabetic-retinopathy-grading-course)  
6  
7  
8 330 7. Widdowson S, Taylor D. The management of grading quality: Good practice in the  
9  
10 331 quality assurance of grading. [Internet] 2015. [cited 2021 April] Available from:  
11  
12 332 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachmen](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/512832/The_Management_of_Grading.pdf)  
13  
14 333 [t\\_data/file/512832/The\\_Management\\_of\\_Grading.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/512832/The_Management_of_Grading.pdf)  
15  
16  
17  
18 334 8. Public Health England. NHS Diabetic Eye Screening Programme Grading definitions  
19  
20 335 for referable disease. [Internet] 2021 (cited 2021 September) Available from:  
21  
22 336 [https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-](https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-grading-criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-referable-disease)  
23  
24 337 [grading-criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-](https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-grading-criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-referable-disease)  
25  
26 338 [referable-disease](https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-grading-criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-referable-disease)  
27  
28  
29  
30 339 9. NHS DESP Core Team. Diabetic Eye Screening Programme Pathway for Adequate /  
31  
32 340 Inadequate Images and where images cannot be taken Outlines pathway and business  
33  
34 341 rules for image capture exceptions and ungradable images. [Internet] 2013.[cited 2021  
35  
36 342 April] Available from: [https://www.gov.uk/government/publications/diabetic-eye-](https://www.gov.uk/government/publications/diabetic-eye-screening-pathway-for-images-and-where-images-cannot-be-taken)  
37  
38 343 [screening-pathway-for-images-and-where-images-cannot-be-taken](https://www.gov.uk/government/publications/diabetic-eye-screening-pathway-for-images-and-where-images-cannot-be-taken)  
39  
40  
41  
42 344 10. Curran K, Congdon N, Peto T, Dardis C, Nguyen NN, Hoange TT et al. Capturing the  
43  
44 345 clinical decision-making processes of expert and novice diabetic retinal graders using a  
45  
46 346 ‘think-aloud’ approach. *Eye*. 2022;36:1019-1026  
47  
48  
49  
50 347 11. Murthy GVS, John N, Shamanna BR, Pant HB. Elimination of avoidable blindness  
51  
52 348 due to cataract: Where do we prioritize and how should we monitor this decade. *Indian*  
53  
54 349 *J Ophthalmol*. 2012;60(5):438–45.  
55  
56  
57  
58 350 12. McKenna M, Chen T, McAneney H, Membrillo MAV, Jin L, Xiao W, et al. Accuracy  
59  
351 of trained rural ophthalmologists versus non-medical image graders in the diagnosis of

- 1  
2  
3 352 diabetic retinopathy in rural China. *Br J Ophthalmol*. 2018;102(11):1471–6.  
4  
5  
6 353 13. Amirul Islam FM. Accuracy and reliability of retinal photo grading for diabetic  
7  
8 354 retinopathy: Remote graders from a developing country and standard retinal photo  
9  
10 355 grader in Australia. *PLoS One*. 2017;12(6):1–12.  
11  
12  
13 356 14. Bhargava M, Cheung CYL, Sabanayagam C, Kawasaki R, Harper CA, Lamoureux EL,  
14  
15 357 et al. Accuracy of diabetic retinopathy screening by trained non-physician graders  
16  
17 358 using non-mydriatic fundus camera. *Singapore Med J*. 2012;53(11):715–9.  
18  
19  
20 359 15. Rani PK, Peguda HK, Chandrashekher M, Swarna S, Jonnadula GB, James J et al.  
21  
22 360 Capacity building for diabetic retinopathy screening by optometrists in India: Model  
23  
24 361 description and pilot results. *Indian J Ophthalmol* 2021.69:655-659.  
25  
26  
27 362 16. Srinivasan S, Shetty S, Natarajan V, Sharma T, Raman R. Development and  
28  
29 363 Validation of a Diabetic Retinopathy Referral Algorithm Based on Single-Field  
30  
31 364 Fundus Photography. *PLoS One* 2016.11:e0163108.  
32  
33  
34  
35  
36

37 365 Figure 1: Flow diagram to illustrate enrolment of patients and management of images in Phase I from  
38  
39 366 August to October 2018 (Initial grading performance analysis). Level 1 and level 2 graders graded the  
40  
41 367 same set of photographs and level 3 graders graded a subset of these photographs: All disagreements  
42  
43 368 between Level 1 and 2 graders and a 40% random sample of all images.  
44  
45

46 369 Figure 2: Flow diagram illustrating the enrolment of patients and management of images included in  
47  
48 370 Phase II from January 2020 to October 2020 (Follow-up grading performance analysis after re-  
49  
50 371 training). Level 1 and level 2 graders graded the same set of photographs and level 3 graders graded a  
51  
52 372 subset of these images.  
53  
54  
55

56 373  
57  
58  
59  
60

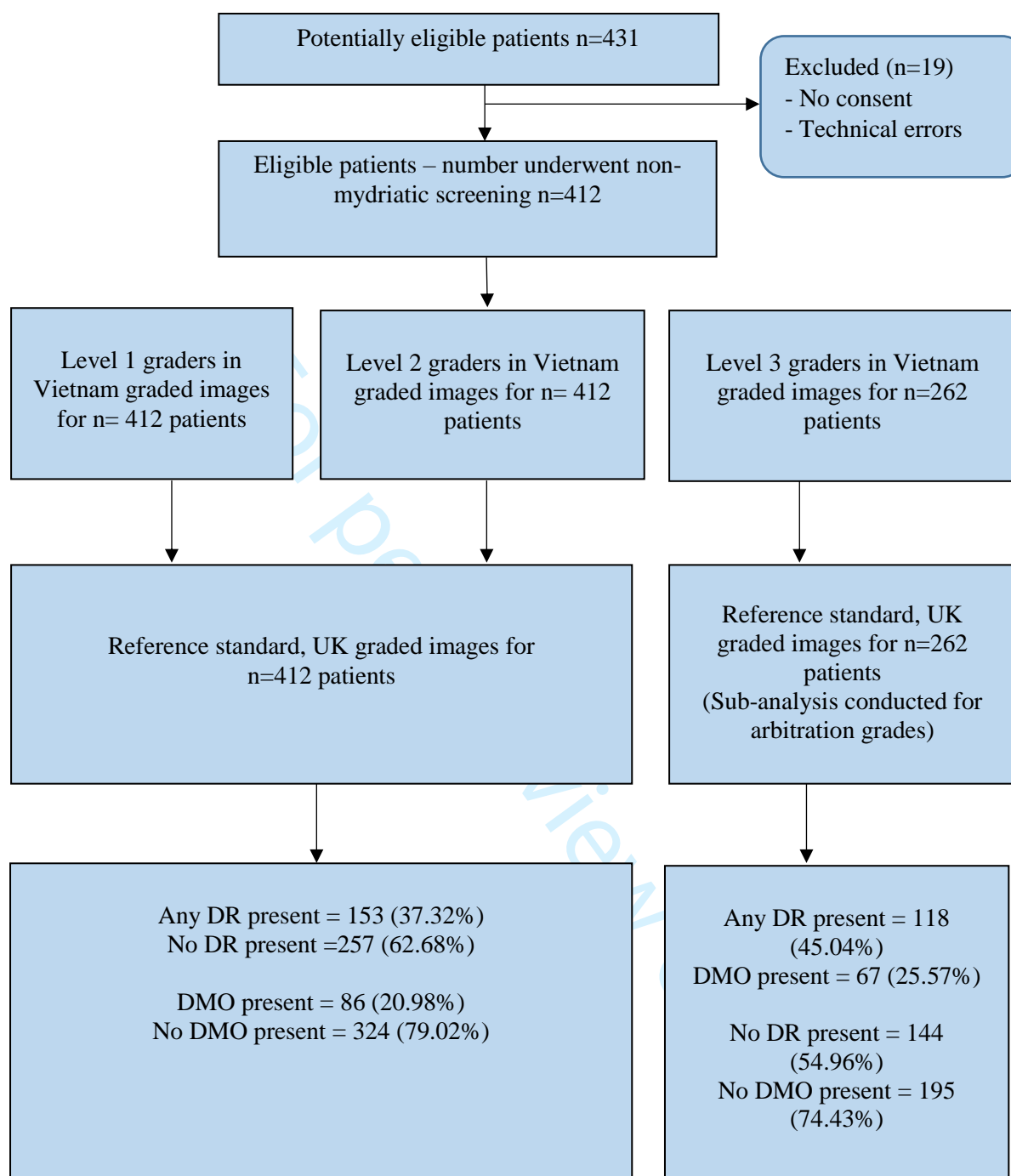
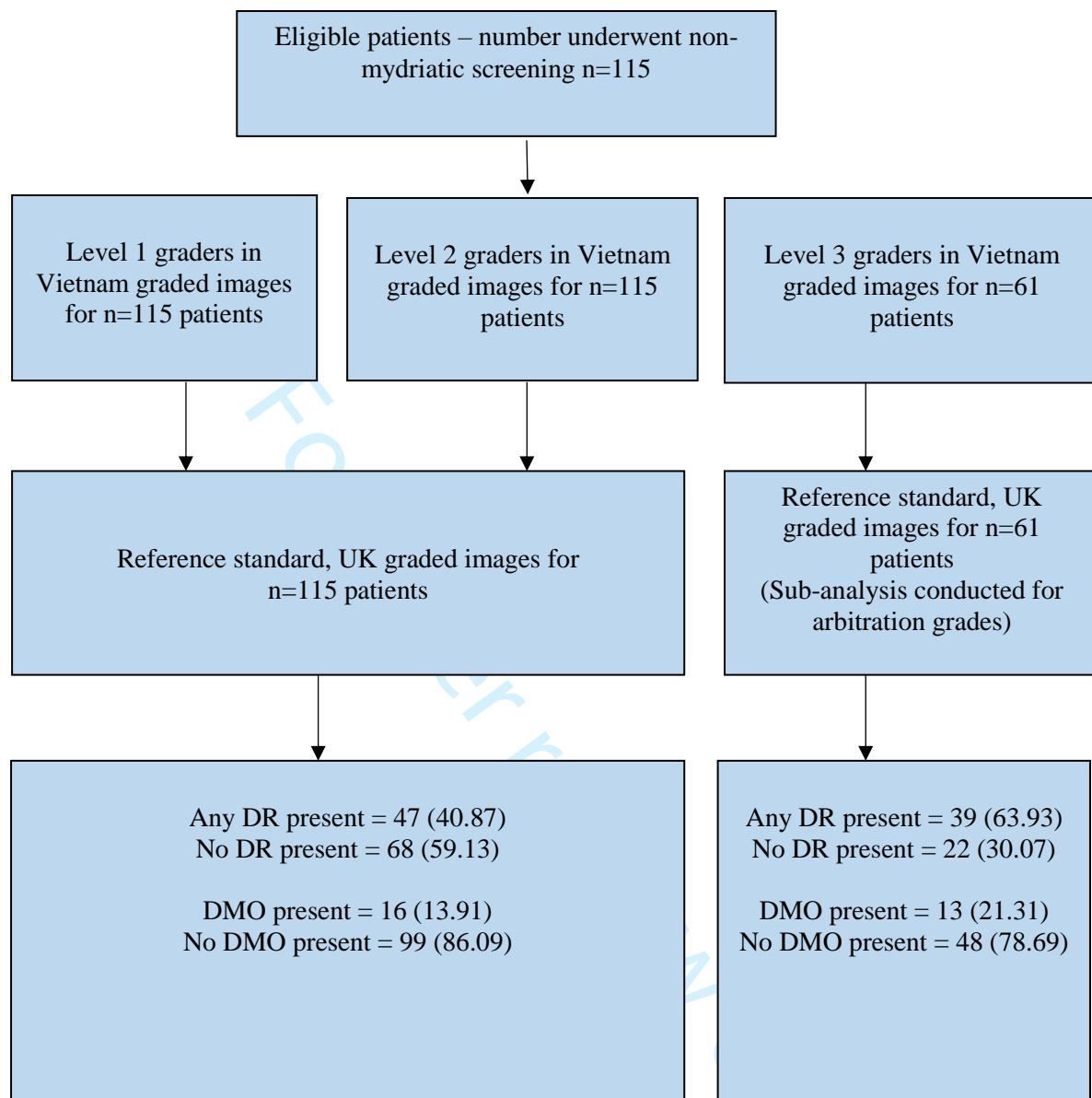


Figure 1: Flow diagram to illustrate enrolment of patients and management of images in Phase I from August to October 2018 (Initial grading performance analysis). Level 1 and level 2 graders graded the same set of photographs and level 3 graders graded a subset of these photographs: All disagreements between Level 1 and 2 graders and a 40% random sample of all images.



43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Figure 2: Flow diagram illustrating the enrolment of patients and management of images included in Phase II from January 2020 to October 2020 (Follow-up grading performance analysis after re-training). Level 1 and level 2 graders graded the same set of photographs and level 3 graders graded a subset of these images.

**SUPPLEMENTARY MATERIAL****September 2017 (5 days)**

- Five Vietnamese doctors and medical administrators visited Northern Ireland to receive screener/graders training in administration and failsafe methods.
- These doctors delivered training to new graders in Vietnam following their visit to Northern Ireland.

**January 2018 (1 week) – Delivered in person in Vietnam by two UK graders (senior ophthalmic nurse and optometrist, certified in DR grading)**

- Observation in retina clinics.
- Hands on training with tabletop CR-2 Canon Fundus Cameras.
- Topics covered: ocular anatomy, retinal disease, DR signs, DR grading (based on the UK DESP grading classification system) and appropriate referral pathways and management (PowerPoint presentation and interactive sessions).
- All graders received a module workbook.
- Certification provided by Orbis.

**March 2018 - Grading began in Vietnam**

- Graders began grading as part of pilot DESP.
- Ongoing training was delivered by the Orbis team and the lead ophthalmologist for DR screening in Vietnam over the course of the following months.

**June 2018 (1-2 days) – Delivered by Orbis partners**

- UK graders developed a PowerPoint presentation based on DR case examples and this was delivered by Orbis.

**March 2019 (2 days) - Delivered by UK grader in Vietnam**

- More training on DR case examples.

**November 2019 (3 days) - Delivered in person in Vietnam by two UK graders (senior ophthalmic nurse and optometrist, certified in DR grading)**

- Refresher DR training, incorporating think-aloud techniques into practical teaching sessions.
- Pre and post training assessments.
- Encouraged use of international test and training (iTAT) for quality assurance purposes. Practical sessions on iTAT.

*Figure S1: A flow chart to highlight the training timeframe for graders in Vietnam*

**Diabetic retinopathy workshop for graders in Vietnam**

|  |                                      |                           |
|--|--------------------------------------|---------------------------|
| <b>AGENDA</b>  |                                      |                           |
| <b>DAY 1-3 Monday 29<sup>th</sup> January – Wednesday 31<sup>st</sup> January</b>                |                                      |                           |
| Visit diabetic retinopathy screening sites to observe and provide hands on training and support. |                                      | UK graders and Orbis team |
| <b>DAY 4- Thursday 1<sup>st</sup> February</b>   |                                      |                           |
| <b>MORNING</b>   |                                      |                           |
| <b>9:00 - 9:30</b>   | Check in                             | UK graders and Orbis team |
| <b>9:30 – 10.30</b>  | Introduction on Diabetic Retinopathy |                           |
| <b>10.30-10.45</b>   | Tea Break                            |                           |
| <b>10.45-11.30</b>   | Basic Screening Component            |                           |
|  | Lunch                                |                           |
| <b>AFTERNOON</b>   |                                      |                           |
| <b>13.30-14:00</b>   | NHS Grading System                   | UK graders and Orbis team |
| <b>14:00-14:45</b>   | Image Quality                        |                           |
| <b>14:45-15:00</b>   | Tea Break                            |                           |
| <b>15:00-16:00</b>   | Image Grading                        |                           |
| <b>16:0-16:30</b>  | Hospital pathway                     |                           |
| <b>DAY 5- Friday 2<sup>nd</sup> February</b>   |                                      |                           |
| <b>MORNING</b>   |                                      |                           |
| <b>9:00-9:30</b>   | Other ocular findings                | UK graders and Orbis team |
| <b>9:30-11:30</b>  | Practice on grading                  |                           |
| <b>11:30-12:00</b>   | Wrap up                              |                           |

1  
2  
3 **Refresher training programme for DR graders in Vietnam (delivered by UK graders and Orbis)**  
4  
5 **in November 2019**  
6  
7

| Time               | Topics  | Method        | Who   | Preparation |
|--------------------|---|---------------|---|-------------|
| <b>DAY 1</b>       |   |               |   |             |
| <b>8:45-9:15</b>   | Introduction/ Pre course quiz   |               | Orbis VN/ UK graders                                      | UK graders  |
| <b>9:15-9:45</b>   | <b>Diabetic Retinopathy (DR) - New Challenges of Blindness Prevention</b><br><i>Objective: Understand the problem of DR and current efforts to manage vision loss. Aim to motivate graders to be involved in DESP</i> | Presentation  | Orbis VN  | Orbis       |
| <b>9:45-10:15</b>  | <b>Retina Anatomy</b><br><i>Objective : Understand the pathobiology of diabetic complications and pathogenesis of retinal damage</i>  | Presentation  | Ho Chi Minh Eye Hospital                                  | Orbis       |
| <b>10:30-11:15</b> | <b>Diabetic Retinopathy (DR) Pathophysiology</b><br><i>Objective : Understand Diabetic Retinopathy</i>  | Presentation  | UK graders  | UK graders  |
| <b>11:15-12:00</b> | <b>Grading system and DR Grading pathway ( UK system)/ How to systematically grade a retinal image</b><br><i>Objective : Understanding the grading system and referral pathway ( UK standard)</i>                     | Presentation  | UK graders  | UK graders  |
| <b>13:30-14:15</b> | <b>Image quality</b><br><i>Objective : Understand the requirements/criteria of image quality for accurate grading</i>   | Presentation  | UK graders  | UK graders  |
| <b>14:15-14:45</b> | <b>Spectra Software</b><br><i>Objective : How to use the current Spectra software for uploading, grading, and managing DR cases</i>   | Demonstration | Senior graders of Tien Giang and Ho Chi Minh Eye Hospital |             |



|              |   |                    |  |   |
|--------------|---|--------------------|--|---|
| 15:00-17:00  | <p><b>Practical Training</b><br/> <b>Parallel session:</b><br/> <b>1- Taking retina images of the patients following the procedure ,and provide counseling to the patients</b></p> <p><i>Objective: Practice experience of taking fundus images</i></p> | Practical training | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and UK graders | 2 fundus cameras:<br>4 Groups: make sure every participant is able to practice at least once        |
| 15:00-17:00  | <p><b>Parallel session :</b><br/> <b>1- Grading DR in the Spectra</b></p> <p><i>Objective: Practical experience of how to do DR grading</i></p>   | Practical training | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and UK graders | 3-4 accounts of Spectra<br>4 Groups : make sure every participant is able to practice at least once |
| <b>DAY 2</b> |   |                    |  |   |
| 08:45-09:00  | <i>Recap of day 1/ introduction to day 2</i>  |                    | Orbis  |   |
| 9:00-9:30    | <p><b>Analysis of retinal fundus images for grading of diabetic retinopathy severity</b><br/> Objective : How to read the image and protocol for retinal image analysis</p>   |                    | Ophthalmologist Ho Chi Minh Eye Hospital                                   |   |
| 9:30-10:15   | <p><b>DR Screening Procedure: Best Practice</b><br/> <i>Objective: Discuss how to build the “best screening procedures” into DESPs.</i></p>   | Presentation       | UK graders   | UK graders  |
| 10:30-11:15  | <p><b>Other Ocular Findings</b><br/> <i>Objective: Awareness of other ocular pathology during DR screening</i></p>  | Presentation       | UK graders   | UK graders  |
| 11:15-12:15  | <b>Image grading case studies competition</b>   | Practical          | UK graders   | We need to organise people into groups of 3 with one experienced grader in each group               |

|  |  |  |                                     |   |  |
|--|--|--|-------------------------------------|---|--|
| 1<br>2<br>3<br>4<br>5<br>6<br>7<br>8<br>9<br>10                                  | <b>13:30-14:30</b>                                       | <b>Counselling and delivering messages to patients during the DR screening</b><br><i>Objective: Important to provide counselling for the patients and deliver messages effectively.</i>  | Presentation/<br>practical training | Orbis Vietnam   |  |
| 11<br>12<br>13<br>14<br>15<br>16<br>17<br>18<br>19<br>20<br>21<br>22<br>23<br>24 | <b>14:30-16:30</b>                                       | <b>Practical Training</b><br><b>Parallel session:</b><br><b>1- Taking retina images of the patients following the procedure ,and provide counseling to the patients</b><br><br><i>Objective: Experience on how to take good fundus photographs</i> | Practical training                  | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and UK graders              | 2 fundus camera<br>4 Group: make sure every participants are able to practice at least one time        |
| 25<br>26<br>27<br>28<br>29<br>30<br>31<br>32<br>33                               |  | <b>Parallel session :</b><br><b>1- Grading DR in the Spectra</b><br><br><i>Objective: Practical experience on DR grading</i>   | Practical training                  | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and UK graders              | -4 accounts of Spectra<br>4 Groups: make sure every participant is able to practice at least one time. |
| 34<br>35   | <b>DAY 3</b><br><b>Part 1 (Final practical training)</b> |  |                                     |   |  |
| 36<br>37   | <b>08:45-9:00</b>  | Recap of day 2/ introduction to day 3  |                                     |   |  |
| 38<br>39<br>40<br>41<br>42<br>43<br>44<br>45<br>46<br>47                         | <b>09:00-09:30</b>                                       | <b>Quality Assurance in Diabetic Screening</b><br><i>Objective : Understand why quality assurance is important and the correct steps required to ensure good quality assurance procedures are in place</i>   | Presentation                        | UK graders  | UK graders   |
| 48<br>49<br>50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59<br>60       | <b>09:30-10:30</b>                                       | <b>Practice : Grading in iTAT</b><br><i>Objective: Know the Online training for DR grading and the importance of lifelong learning for DR grading</i>  | Practice                            | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and support from UK graders | ITAT accounts for practicing   |

|  |   |   |                       |  |   |
|--|---|---|-----------------------|--|---|
| 1<br>2<br>3<br>4<br>5<br>6<br>7<br>8<br>9<br>10<br>11<br>12<br>13<br>14  | <b>10:45-12:30</b>                              | <b><u>Assessment</u></b><br><br><b>Parallel session:</b><br><b>1- Taking retina images of the patients following the procedure ,and provide counseling to the patients</b><br><b>2- Grading DR in the Spectra</b>     | Practical Training    | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and UK graders | 2 fundus cameras<br>4 Group and 4 accounts of Spectra<br>4 Groups |
| 15<br>16   | <b>DAY 3</b><br><b>Part 2 (Future planning)</b> |   |                       |  |   |
| 17<br>18<br>19   | <b>13:30-14:15</b>                              | <b>Post course Quiz and results</b>   |                       | UK graders   | UK graders  |
| 20<br>21<br>22<br>23<br>24   | <b>14:15-15:00</b>                              | <b>Teaching Methodology for adults</b><br><i>Objective : How to train new graders effectively</i>   | Think aloud work shop | Orbis VN Supported by UK graders   | UK graders  |
| 25<br>26<br>27<br>28<br>29<br>30<br>31<br>32<br>33   | <b>15:15-16:00</b>                              | <b>Supportive Supervision Methodology: Developing quality improvement</b><br><i>Objective: How to plan, implement the supervision trips to correct / improve other graders' performances. Provide checklist tools</i> | Think aloud work shop | Orbis VN Supported by UK graders   | UK graders  |
| 34<br>35<br>36   | <b>16:00-16:45</b>                              | <b>Feedbacks and Plan for next steps</b>  | Discussion            |  |   |
| 37<br>38<br>39<br>40<br>41<br>42<br>43<br>44<br>45<br>46<br>47<br>48<br>49<br>50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59<br>60 | <b>16:45-17:00</b>                              | <b>Certificates for Vietnamese graders in attendance</b>  |                       |  |   |

Table S1: UK DR Grading Classification Scale

| NSC        | International Term                                     | Symptoms                         | Features  | Action                 |
|------------|--|----------------------------------|---|------------------------|
| <b>R0</b>  | No DR  | None                             | No signs of diabetic retinopathy  | Annual rescreen        |
| <b>R1</b>  | Mild non-proliferative (mild pre-proliferative)        | None                             | Haemorrhages & microaneurysms, only   | Annual rescreen        |
| <b>R2</b>  | Moderate non-proliferative, moderate pre-proliferative | None                             | Extensive Microaneurysms, intraretinal haemorrhages, hard exudates, venous abnormalities, large blot haemorrhages, cotton wool spots (small infarcts), venous beading, venous loop, venous reduplication.   | Refer routinely to HES |
| <b>R3s</b> | Stable proliferative diabetic retinopathy              |                                  | No haemorrhages or exudates or new vessels, laser scars   | Annual rescreen        |
| <b>R3a</b> | Active proliferative diabetic retinopathy              | Floaters, central loss of vision | New vessel formation either at the disc (NVD) or elsewhere (NVE). Extensive fibrovascular proliferation, retinal detachment, pre-retinal or vitreous haemorrhage.   | Urgent referral to HES |
| <b>M0</b>  |  |                                  | No maculopathy  | Annual rescreen        |
| <b>M1</b>  | Diabetic maculopathy                                   | Blurred central vision           | <p>The macula is defined as a circle centred on the fovea, with a radius of the distance to the disc margin.</p> <p>If the leakage involves or is near the fovea the condition is termed clinically significant macular oedema (CSME).</p> <p>Exudative maculopathy presents with leakage, retinal thickening, microaneurysms, hard exudates at the macula. Ischaemic form can have a featureless macular with NVE and poor vision.</p> <p>Milder forms:</p> <ul style="list-style-type: none"> <li>• exudate &lt; or = 1DD of centre of fovea</li> <li>• circinate or group of exudates within macula</li> <li>• any microaneurysm or haemorrhage &lt; or = 1DD of centre of fovea only is associated with a best VA of &lt; or = 6/12 retinal thickening &lt; or =</li> </ul> | Refer to HES           |

|   |                  |                                |   |                                       |
|---|------------------|--------------------------------|---|---------------------------------------|
|   |                  |                                | 1DD of centre of fovea<br>(if stereos available)  |                                       |
| <b>P</b>  | Photocoagulation | Reduced night<br>vision, glare | Small retinal scars throughout<br>the peripheral retina.  |                                       |
| <b>U</b>  | Ungradable       |                                | Ungradable is usually due to<br>cataract, small pupils, other<br>lesions usually referred for<br>assessment | Refer for slit<br>lamp<br>examination |
| <b>Abbreviations:</b> DR = diabetic retinopathy, NPDR = none-proliferative retinopathy, NVE = new vessels elsewhere, IRMAs = intraretinal microvascular abnormalities (part of severe pre-proliferative retinopathy, vessels will not leak with angiogram, otherwise they would be 'new vessels' making the condition 'proliferative'), MO=macular oedema, MA= microaneurysm, DD=disc diameter, HES= hospital eye service |                  |                                |   |                                       |

**Table S2: Reference standards intra-rater agreement score using kappa statistic (first attempt versus second attempt)**

|  | <b>Intra-rater agreement<br/>(reference standard, UK),<br/>k (95% CI)<br/>(by eyes, n=106)</b> | <b>Intra-rater agreement<br/>(reference standard, UK), k<br/>(95% CI)<br/>(by worst eye, n=53)</b> |
|--|--|--|
| <b>Overall Diabetic Retinopathy Grading:</b>   |  |  |
| <b>Any DR</b>  | 0.96 (0.91, 1.00)  | 0.92 (0.82, 1.00)  |
| <b>Treatable DR</b>  | 0.81 (0.60, 1.00)  | 0.74 (0.47, 1.00)  |
| <b>Referable Maculopathy</b>   | 0.97 (0.92, 1.00)  | 1.00 (1.00, 1.00)  |
| Abbreviations: CI=confidence interval, k=kappa, DR=Diabetic retinopathy, DMO=diabetic macular oedema<br>Any DR defined as R1, R2, R3s, R3a and U<br>Treatable DR defined as R3a<br>Referable DMO defined as M1 and U |  |  |

**Table S3: Using kappa statistic to determine the inter-rater agreement between the reference standard and one senior grader from QUB grading centre**

|  | <b>Inter-rater agreement<br/>(reference standard vs a<br/>senior grader QUB), k<br/>(95% CI)<br/>(by eyes, n=106)</b> | <b>Inter-rater agreement<br/>(reference standard vs a<br/>senior grader QUB (by<br/>worst), k (95% CI)<br/>(by worst eye, n=53)</b> |
|--|---|---|
| <b>Overall Diabetic Retinopathy Grading:</b>   |   |   |
| <b>Any DR</b>  | 0.79 (0.67, 0.91)   | 0.74 (0.55, 0.92)   |
| <b>Treatable DR</b>  | 0.71 (0.48, 0.95)   | 0.68 (0.39, 0.97)   |
| <b>Referable Maculopathy</b>   | 0.75 (0.61, 0.90)   | 0.74 (0.55, 0.93)   |
| Abbreviations: CI=confidence interval, k=kappa, DR=Diabetic retinopathy, DMO=diabetic macular oedema<br>Any DR defined as R1, R2, R3s, R3a and U<br>Treatable DR defined as R3a<br>Referable DMO defined as M1 and U |   |   |

### **Intra and inter-grader agreement**

To ensure there was good intra-grader reliability as a reference standard, a stratified random sample of images were regraded. There was approximately one month between the first and second attempts to reduce the possibility of bias caused by memory. Additionally, inter-grader agreement was calculated using kappa to ensure there was good grading agreement between the reference standard and one senior grader from the Ophthalmic Reading Centre at QUB, Belfast. Any disagreements were discussed with

a retinal specialist until consensus was reached. Overall, the intra-grader agreement and inter-grader agreement ranged from substantial to almost perfect.

**Table S4: The prevalence of any diabetic retinopathy (DR), referable DR, any maculopathy and ungradable cases with the reference grader from Phase I and Phase II**

| Diabetic Retinopathy grades  | Phase I     | Phase II (post remedial training) | P-Value |
|--|-------------|-----------------------------------|---------|
| R0 (n,%)   | 257 (62.68) | 68 (59.13)                        | P=0.347 |
| R1 (n,%)   | 100 (24.39) | 32 (27.83)                        |         |
| R2 (n,%)   | 11 (2.68)   | 2 (1.74)                          |         |
| R3a (n,%)  | 10 (2.44)   | 7 (6.09)                          |         |
| R3s (n,%)  | 1 (0.24)    | 0 (0.00)                          |         |
| U (n,%)  | 31 (7.56)   | 6 (5.22)                          |         |
| <b>Any DR</b>  |             |                                   |         |
| - Yes (n,%)  | 153 (37.32) | 47 (40.87)                        | P=0.488 |
| - No (n,%)   | 257 (62.68) | 68(59.12)                         |         |
| <b>Referable DR</b>  |             |                                   |         |
| - Yes (n,%)  | 52 (12.68)  | 15 (13.04)                        | P=0.918 |
| - No (n,%)   | 358 (87.32) | 100 (86.96)                       |         |
| <b>Any DMO</b>   |             |                                   |         |
| - M0 (n,%)   | 324 (79.02) | 99 (86.09)                        | P=0.173 |
| - M1 (n,%)   | 43 (10.49)  | 10 (8.70)                         |         |
| - U (n,%)  | 43 (10.49)  | 6 (5.22)                          |         |
| <b>Abbreviations:</b> DR=diabetic retinopathy, DMO=Diabetic Macular Oedema, U=ungradable<br>Chi-Squares used to test significance. |             |                                   |         |

**Table S5. Diagnostic test accuracy of DR graders in Vietnam against a reference standard from the UK, excluding ungradable images.**

|   | Level 1 graders<br>(n=373 patient<br>images) | Level 2 graders<br>(n=373 patient<br>images)* | Level 3 graders<br>(n=235 patient<br>images) † |
|---|--|---|--|
| <b>Any DR</b>   |  |   |  |
| Sensitivity (%) (95% CI)  | 47.9 (38.8, 57.2)                            | 50.8 (41.6, 60.0)                             | 49.0 (38.7, 59.3)                              |
| Specificity (%) (95% CI)  | 89.7 (85.1, 93.0)                            | 98.8 (96.3, 99.7)                             | 100 (96.8, 100)                                |
| PPV (%) (95% CI)  | 69.0 (57.9, 78.4)                            | 95.3 (86.2, 98.8)                             | 100 (90.6, 100)                                |
| NPV (%) (95% CI)  | 78.2 (72.9, 82.7)                            | 80.9 (76.0, 85.0)                             | 74.6 (67.8, 80.5)                              |
| <b>Referable DR</b>   |  |   |  |
| Sensitivity (%) (95% CI)  | 38.1 (19.0, 61.3)                            | 28.6 (12.2, 52.3)                             | 22.2 (7.4, 48.1)                               |
| Specificity (%) (95% CI)  | 98.9 (96.9, 99.6)                            | 100 (98.7, 100)                               | 99.5 (97.1, 99.9)                              |
| PPV (%) (95% CI)  | 66.7 (35.4, 88.7)                            | 100 (51.7, 100)                               | 80.0 (29.9, 98.9)                              |
| NPV (%) (95% CI)  | 96.4 (93.8, 97.9)                            | 96.0 (93.3, 97.6)                             | 94.0 (89.9, 96.6)                              |
| <b>Referable DMO</b>  |  |   |  |
| Sensitivity (%) (95% CI)  | 9.3 (3.0, 23.1)                              | 37.2 (23.4, 53.3)                             | 26.5 (13.5, 44.7)                              |
| Specificity (%) (95% CI)  | 99.1 (97.0, 99.8)                            | 99.4 (97.5, 99.9)                             | 100 (97.6, 100)                                |
| PPV (%) (95% CI)  | 57.1 (20.2, 88.2)                            | 88.9 (63.9, 98.1)                             | 100 (62.9, 100)                                |
| NPV (%) (95% CI)  | 88.9 (85.1, 92.0)                            | 92.3 (88.9, 94.7)                             | 88.6 (83.5, 92.4)                              |
| <p><b>Abbreviations:</b> UK = United Kingdom, DR = Diabetic Retinopathy, DMO = Diabetic Macular Oedema, CI = Confidence Intervals,<br/>           Grading criteria: UK National Diabetic Eye Screening Programme (NDESP) classification system (See supplementary material, Table S1 for more details).<br/>           Any DR, is defined as grades R1, R2, R3s and R3a.<br/>           Referable DR is defined as grades R2 and R3a.<br/>           Referable DMO is defined as grades M1<br/>           Sensitivity is the ability of a test to correctly identify patients with a disease and specificity is the ability of a test to correctly identify people without the disease<br/>           Positive predictive value (PPV) is the proportion of those who test positive who have the condition (true positives) and negative predictive value (NPV) is the proportion of those who test negative who do not have the condition (true negatives).</p> |  |   |  |



**Table S6: Diagnostic test accuracy of DR graders in Vietnam against a reference standard from the UK after additional DR training was delivered, excluding ungradable images**

|   | <b>Level 1 graders<br/>(n=115 patient<br/>images)</b> | <b>Level 2<br/>graders (n=115<br/>patient images)</b> | <b>Level 3 graders<br/>(n=62 patient<br/>images)</b> |
|---|---|---|--|
| <b>Any DR</b>   |   |   |  |
| Sensitivity (%) (95% CI)  | 97.6 (85.6, 99.9)                                     | 72.5 (55.9, 84.9)                                     | 55.6 (38.1, 72.1)                                    |
| Specificity (%) (95% CI)  | 95.6 (86.8, 99.8)                                     | 100 (93.5, 100)                                       | 100 (84.6, 100)                                      |
| PPV (%) (95% CI)  | 93.0 (79.9, 98.2)                                     | 100 (85.4, 100)                                       | 100 (80.0, 100)                                      |
| NPV (%) (95% CI)  | 98.5 (90.7, 99.9)                                     | 85.5 (75.2, 92.2)                                     | 57.9 (10.9, 73.2)                                    |
| <b>Referable DR</b>   |   |   |  |
| Sensitivity (%) (95% CI)  | 88.9 (50.7, 99.4)                                     | 55.6 (22.7, 84.7)                                     | 77.8 (40.0, 97.2)                                    |
| Specificity (%) (95% CI)  | 90.0 (81.9, 94.8)                                     | 96.9 (90.5, 99.2)                                     | 100 (92.8, 100)                                      |
| PPV (%) (95% CI)  | 44.4 (22.4, 68.7)                                     | 62.5 (25.9, 89.8)                                     | 100 (56.1, 100)                                      |
| NPV (%) (95% CI)  | 98.9 (93.4, 99.9)                                     | 95.9 (89.2, 98.7)                                     | 96.1 (87.8, 98.8)                                    |
| <b>Referable DMO</b>  |   |   |  |
| Sensitivity (%) (95% CI)  | 90.0 (54.1, 99.5)                                     | 60.0 (26.4, 86.3)                                     | 80.0 (44.4, 97.5)                                    |
| Specificity (%) (95% CI)  | 97.0 (91.8, 99.2)                                     | 97.9 (91.9, 99.6)                                     | 100 (92.6, 100)                                      |
| PPV (%) (95% CI)  | 75.0 (42.8, 93.3)                                     | 75.0 (35.6, 95.5)                                     | 100 (59.8, 100)                                      |
| NPV (%) (95% CI)  | 99.0 (93.6, 99.9)                                     | 95.9 (89.2, 98.7)                                     | 96.0 (87.4, 99.6)                                    |
| <p><b>Abbreviations:</b> UK = United Kingdom, DR = Diabetic Retinopathy, DMO = Diabetic Macular Oedema, CI = Confidence Intervals,<br/>           Grading criteria: UK National Diabetic Eye Screening Programme (NDESP) classification system (See supplementary material, Table S1 for more details).<br/>           Criteria:<br/>           Any DR is defined as grades R1, R2, R3s, and R3a.<br/>           Referable DR is defined as grades R2 and R3a.<br/>           Referable DMO is defined as grades M1.<br/>           Sensitivity is the ability of a test to correctly identify patients with a disease and specificity is the ability of a test to correctly identify people without the disease<br/>           Positive predictive value (PPV) is the proportion of those who test positive who have the condition (true positives) and negative predictive value (NPV) is the proportion of those who test negative who do not have the condition (true negatives).</p> |   |   |  |

| Section & Topic          | No  | Item   | Reported on page #          |
|--------------------------|-----|--|-----------------------------|
| <b>TITLE OR ABSTRACT</b> |     |  |                             |
|                          | 1   | Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)  | 1                           |
| <b>ABSTRACT</b>          |     |  |                             |
|                          | 2   | Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)                                 | 3-4                         |
| <b>INTRODUCTION</b>      |     |  |                             |
|                          | 3   | Scientific and clinical background, including the intended use and clinical role of the index test   | 5-6                         |
|                          | 4   | Study objectives and hypotheses  | 5                           |
| <b>METHODS</b>           |     |  |                             |
| <i>Study design</i>      | 5   | Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)     | 5                           |
| <i>Participants</i>      | 6   | Eligibility criteria   | 6-7                         |
|                          | 7   | On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)                 | 6-7                         |
|                          | 8   | Where and when potentially eligible participants were identified (setting, location and dates)   | 6-7                         |
|                          | 9   | Whether participants formed a consecutive, random or convenience series  | 7                           |
| <i>Test methods</i>      | 10a | Index test, in sufficient detail to allow replication  | 6-7                         |
|                          | 10b | Reference standard, in sufficient detail to allow replication  | 6                           |
|                          | 11  | Rationale for choosing the reference standard (if alternatives exist)  | 6                           |
|                          | 12a | Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory         |                             |
|                          | 12b | Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory |                             |
|                          | 13a | Whether clinical information and reference standard results were available to the performers/readers of the index test                                 | 6-7                         |
|                          | 13b | Whether clinical information and index test results were available to the assessors of the reference standard  | 6-7                         |
| <i>Analysis</i>          | 14  | Methods for estimating or comparing measures of diagnostic accuracy  | 12-14                       |
|                          | 15  | How indeterminate index test or reference standard results were handled  | 12-14<br>Supplementary file |
|                          | 16  | How missing data on the index test and reference standard were handled   | 12-14                       |
|                          | 17  | Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory  | 12-14                       |
|                          | 18  | Intended sample size and how it was determined   | NA                          |
| <b>RESULTS</b>           |     |  |                             |
| <i>Participants</i>      | 19  | Flow of participants, using a diagram  | 10-11                       |
|                          | 20  | Baseline demographic and clinical characteristics of participants  | 9                           |
|                          | 21a | Distribution of severity of disease in those with the target condition   | Supplementary file          |
|                          | 21b | Distribution of alternative diagnoses in those without the target condition  | Supplementary file          |
|                          | 22  | Time interval and any clinical interventions between index test and reference standard   | Supplementary file          |
| <i>Test results</i>      | 23  | Cross tabulation of the index test results (or their distribution) by the results of the reference standard  | 12-14 and supplementary     |
|                          | 24  | Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)  | 12-14                       |
|                          | 25  | Any adverse events from performing the index test or the reference standard  | NA                          |
| <b>DISCUSSION</b>        |     |  |                             |
|                          | 26  | Study limitations, including sources of potential bias, statistical uncertainty, and generalisability  | 17                          |
|                          | 27  | Implications for practice, including the intended use and clinical role of the index test  | 17-18                       |
| <b>OTHER INFORMATION</b> |     |  |                             |
|                          | 28  | Registration number and name of registry   | NA                          |
|                          | 29  | Where the full study protocol can be accessed  | NA                          |

For peer review only

# BMJ Open

## The impact of targeted diabetic retinopathy training for graders in Vietnam and the implications for future diabetic retinopathy screening programmes: a diagnostic test accuracy study

|                                 |   |
|---------------------------------|---|
| Journal:                        | <i>BMJ Open</i>   |
| Manuscript ID                   | bmjopen-2021-059205.R2  |
| Article Type:                   | Original research   |
| Date Submitted by the Author:   | 25-Jul-2022   |
| Complete List of Authors:       | Curran, Katie; Queen's University Belfast School of Medicine Dentistry and Biomedical Sciences, Centre of Public Health<br>Congdon, Nathan; Queen's University of Belfast; ORBIS International<br>Hoang, Tung; Hanoi Medical University, Department of Ophthalmology; The University of Sydney Save Sight Institute<br>Lohfeld, Lynne; Queen's University of Belfast; Eye Hospital of Wenzhou Medical University<br>Nguyen, Van; Orbis International in Vietnam<br>Nguyen, Hue; Orbis International in Vietnam<br>Nguyen, Quan; Ho Chi Minh Eye Hospital, Vitreo-Retina Department<br>Dardis, Catherine; Belfast Health and Social Care Trust, Department of Ophthalmology<br>Virgili, Gianni ; Queen's University of Belfast; Belfast Health and Social Care Trust, Department of Ophthalmology<br>Piyasena, Prabhath; Queen's University of Belfast<br>Tran, Huong; Orbis International in Vietnam<br>Salongcay, Recivall; Queen's University Belfast<br>Tung, Mai; Ho Chi Minh Eye Hospital, Vitreo-Retina Department<br>Peto, Tunde; Queen's University of Belfast; Belfast Health and Social Care Trust, Department of Ophthalmology |
| <b>Primary Subject Heading</b>: | Ophthalmology   |
| Secondary Subject Heading:      | Diabetes and endocrinology, Medical education and training  |
| Keywords:                       | Diabetic retinopathy < DIABETES & ENDOCRINOLOGY, OPHTHALMOLOGY, PUBLIC HEALTH, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT   |
|                                 |   |

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3 1 **Title: The impact of targeted diabetic retinopathy training for graders in Vietnam and**  
4  
5 2 **the implications for future diabetic retinopathy screening programmes: a diagnostic test**  
6  
7 3 **accuracy study**  
8  
9

10  
11 4 **Short title:** Training diabetic retinopathy graders in Vietnam  
12  
13

14  
15 5 **Authors:**  
16

17  
18 6 \*Katie Curran, MPH, PhD<sup>1</sup> ([k.curran@qub.ac.uk](mailto:k.curran@qub.ac.uk))  
19

20  
21  
22 7 Nathan Congdon, MD, MPH<sup>1,2,3</sup> ([Nathan.Congdon@orbis.org](mailto:Nathan.Congdon@orbis.org))  
23

24  
25 8 Tung Thanh Hoang<sup>4,5</sup> ([tung.hoang@sydney.edu.au](mailto:tung.hoang@sydney.edu.au))  
26

27  
28 9 Lynne Lohfeld, MPH, PhD<sup>1,6</sup> ([l.lohfeld@qub.ac.uk](mailto:l.lohfeld@qub.ac.uk))  
29

30  
31  
32 10 Van Thu Nguyen, BA<sup>7</sup> ([van.nguyen@orbis.org](mailto:van.nguyen@orbis.org))  
33

34  
35 11 Hue Thi Nguyen, MD, MPH<sup>7</sup> ([hue.nguyen@orbis.org](mailto:hue.nguyen@orbis.org))  
36

37  
38 12 Quan Nhu Nguyen, MD, PhD<sup>8</sup> ([bsquan2009@gmail.com](mailto:bsquan2009@gmail.com))  
39

40  
41  
42 13 Catherine Dardis, MSc<sup>9</sup> ([cathy831oph@gmail.com](mailto:cathy831oph@gmail.com))  
43

44  
45 14 [Gianni Virgili MD<sup>1,9</sup>](mailto:G.Virgili@qub.ac.uk) ([G.Virgili@qub.ac.uk](mailto:G.Virgili@qub.ac.uk))  
46

47  
48 15 [Prabhath Piyasena, MB BS, PhD<sup>1</sup>](mailto:m.piyasena@qub.ac.uk), ([m.piyasena@qub.ac.uk](mailto:m.piyasena@qub.ac.uk))  
49

50  
51  
52 16 Huong Tran, MPA<sup>7</sup> ([Huong.Tran@orbis.org](mailto:Huong.Tran@orbis.org))  
53

54  
55 17 Recivall Salongcay, MD, MPM<sup>1</sup> ([R.SALONGCAY@qub.ac.uk](mailto:R.SALONGCAY@qub.ac.uk))  
56

57  
58 18 Mai Quoc Tung, MD, PhD<sup>8</sup> ([maitungtn@gmail.com](mailto:maitungtn@gmail.com))  
59  
60

1  
2  
3 19 Tunde Peto, MD, PhD<sup>1,9</sup> ([T.Peto@qub.ac.uk](mailto:T.Peto@qub.ac.uk))  
4  
5

6  
7 **Affiliations:**  
8  
9

10 21 <sup>1</sup>Centre for Public Health, Queen's University of Belfast, Belfast, UK  
11  
12

13 22 <sup>2</sup>Orbis International, New York, USA  
14  
15

16 23 <sup>3</sup>Zhongshan Ophthalmic Centre, Guangzhou, China  
17  
18

19 24 <sup>4</sup>Ophthalmology Department, Hanoi Medical University, Hanoi, Vietnam  
20  
21  
22

23 25 <sup>5</sup>Save Sight Institute, The University of Sydney School of Medicine, Sydney, Australia  
24  
25  
26

27 26 <sup>6</sup>Eye Hospital of Wenzhou Medical University, Wenzhou, Zhejiang Province, China  
28  
29

30 27 <sup>7</sup>Orbis International in Vietnam, Hanoi, Vietnam  
31  
32

33 28 <sup>8</sup>Vitreo-Retina Department, Ho Chi Minh Eye Hospital, Ho Chi Minh City, Vietnam  
34  
35  
36

37 29 <sup>9</sup>Department of Ophthalmology, Belfast Health and Social Care Trust, NI  
38  
39  
40

41 30 **\*Corresponding author**  
42  
43

44 31 Katie Curran  
45  
46

47 32 Centre for Public Health  
48  
49

50 33 Institute of Clinical Sciences, A  
51  
52

53 34 Queen's University Belfast  
54  
55

56 35 BT12 6BA  
57  
58

59 36 Email: [k.curran@qub.ac.uk](mailto:k.curran@qub.ac.uk)  
60

1  
2  
3 37 **Keywords:** Diabetic retinopathy, diabetic retinopathy training, screening  
4  
5

6 38 **Word Count:** 2785  
7  
8

9 39 **Abstract**  
10  
11

12  
13 40 **Objectives:** To compare the accuracy of trained level 1 diabetic retinopathy (DR) graders  
14  
15 41 (nurses, endocrinologists, one general practitioner), level 2 graders (mid-level  
16  
17 42 ophthalmologists) and level 3 graders (senior ophthalmologists) in Vietnam against a  
18  
19 43 reference standard from the UK, and assess the impact of supplementary targeted grader  
20  
21 44 training.  
22  
23

24  
25 45 **Design:** Diagnostic test accuracy study.  
26  
27

28  
29 46 **Setting:** Secondary care hospitals in Southern Vietnam.  
30  
31

32  
33 47 **Participants:** DR training was delivered to Vietnamese graders in February 2018 by National  
34  
35 48 Health System (NHS) UK graders. Two-field retinal images (412 patient images) were  
36  
37 49 graded by 14 trained graders in Vietnam between August-October 2018 and then re-graded  
38  
39 50 retrospectively by an NHS-certified reference standard UK optometrist (Phase I). Further DR  
40  
41 51 training based on Phase I results was delivered to graders in November 2019. After training, a  
42  
43 52 randomised subset of images from January-October 2020 (115 patient images) was graded by  
44  
45 53 six of the original cohort (Phase II). The reference grader re-graded all images from Phase I  
46  
47 54 and II retrospectively in masked fashion.  
48  
49  
50

51  
52 55 **Primary and secondary outcome measures:** Sensitivity was calculated at the two different  
53  
54 56 time points and Chi-Squared was used to test significance.  
55  
56  
57  
58  
59  
60



1  
2  
3 57 **Results:** In Phase I, the sensitivity for detecting any DR for all grader groups in Vietnam  
4  
5 58 was low (41.8-42.2%) and improved in Phase II after additional training was delivered (51.2-  
6  
7 59 87.3%). The greatest improvement was seen among level 1 graders ( $P<0.001$ ) and the lowest  
8  
9 60 improvement was observed among level 3 graders ( $P=0.326$ ). There was a statistically-  
10  
11 61 significant improvement in sensitivity for detecting referable DR and referable diabetic  
12  
13 62 macular oedema between all grader levels. The post-training values ranged from 40.0-61.5%  
14  
15 63 (including ungradable images) and 55.6%-90.0% (excluding ungradable images).  
16  
17  
18  
19

20 64 **Conclusions:** This study demonstrates that targeted training interventions can improve  
21  
22 65 accuracy of DR grading. These findings have important implications for improving service  
23  
24 66 delivery in DR screening programmes in low-resource settings.  
25  
26  
27

## 28 67 **Article Summary**

### 29 30 31 68 **Strengths and limitations of this study**

- 32  
33 69 • Graders in Vietnam were trained to detect DR based on the UK's DR screening  
34  
35 70 model.
- 36  
37 71 • This study describes the impact of a training intervention to improve DR grading in  
38  
39 72 Vietnam
- 40  
41 73 • Gradable and ungradable fundus image grading were included in the analysis.
- 42  
43 74 • The sample size was smaller in Phase II compared to Phase I.  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 78 **Introduction**

79 The prevalence of diabetes among adults in Vietnam is approximately 6% and has almost  
80 doubled over the past decade.[1] Early detection through diabetic eye screening programmes  
81 (DESPs) is important to reduce the risk of avoidable blindness due to diabetic retinopathy  
82 (DR). Since the introduction of systematic DESPs in the UK, a high-income country (HIC),  
83 diabetic retinopathy (DR) is no longer the leading cause of blindness among working age  
84 adults.[2] The key to such successful DESPs is implementing accurate, innovative and cost-  
85 effective models tailored to fit healthcare systems and contexts.

86 Investing in training personnel to increase human resources and procuring appropriate  
87 diagnostic and treatment equipment are essential to ensure that service providers can deliver  
88 optimum care for people with DR. In low-middle income countries (LMICs), there is often  
89 insufficient capacity to implement robust DESPs due to the lack of skilled human resources  
90 and infrastructure.[3,4] In Vietnam, there are only 14 ophthalmologists per million population  
91 compared to 49 per million in the UK.[5]

92 All screening programmes must provide evidence of their ability to detect the targeted  
93 condition and ensure that the service performs efficiently to improve screening accuracy  
94 when it falls short. To date, there is insufficient evidence on DR grading accuracy using non-  
95 mydriatic digital imaging by trained graders in LMICs, and even less about the capacity of  
96 DESPs in LMICs to improve where poor accuracy is detected. The current retrospective  
97 study is designed to assess accuracy of a range of graders in a non-governmental organisation  
98 (NGO)-supported DESP in Vietnam, and to study the efficacy of a quality-improvement  
99 intervention.

## 100 **Methods**

1  
2  
3 101 **Study participants:** The 14 participants from Vietnam in Phase I included: Level 1 DR  
4  
5 102 graders (6 nurses, 1 general practitioner and 2 endocrinologists, all with < 1 year grading  
6  
7 103 experience, 55.6% female), Level 2 DR graders (3 newly-qualified ophthalmologists with < 1  
8  
9 104 year formal DR grading experience, 100% female), and Level 3 DR graders (2 senior  
10  
11 105 ophthalmologists with >5 years' experience providing treatment for sight threatening DR, but  
12  
13 106 with <1 year formal DR grading experience, 100% male). In Phase II, 6/14 graders (3 Level  
14  
15 107 1, 2 Level 2, 1 Level 3) from Phase I were included. The reference standard from the UK  
16  
17 108 (KC) was a fully-qualified optometrist trained in DR grading and certified by the UK NHS  
18  
19 109 DESP.[6] Vietnamese Level 1, 2 and 3 graders are equivalent to primary, secondary and  
20  
21 110 arbitration graders, respectively, in UK DESPs.[7] In the current study, Vietnamese Level 1  
22  
23 111 and Level 2 graders graded all fundus images for DR. All images having disagreement  
24  
25 112 between graders, and an additional randomly-selected 40% of all images, were sent for  
26  
27 113 arbitration grading by Level 3 graders in Vietnam. All graders in Vietnam were masked to  
28  
29 114 any prior diagnoses or grades of the reference standard, while the reference standard was also  
30  
31 115 masked to results of grading in Vietnam. Fundus images were graded for 412 patients in  
32  
33 116 phase I and 115 patients in phase II (Figure 1 and figure 2).

34  
35  
36  
37  
38  
39  
40  
41 117 **DR training for graders in Vietnam:** As part of a DESP project supported by NGO Orbis  
42  
43 118 International, a team of five Vietnamese doctors and medical administrators visited a  
44  
45 119 Northern Ireland (NI) DESP in September 2017 to receive training on screening, programme  
46  
47 120 administration and quality control methods. In February 2018, a senior UK NHS grader from  
48  
49 121 the Belfast Trust (CD) and a fully-qualified optometrist, trained in DR grading and certified  
50  
51 122 by the NHS (KC), visited Vietnam to deliver DR training to graders involved in the DESPs.  
52  
53 123 (Supplementary, Figure S1 for training timeline). Training focused on ocular anatomy, retinal  
54  
55 124 diseases, DR signs and grading (based on the UK National Screening Committee (NSC)  
56  
57  
58  
59  
60

1  
2  
3 125 classification system), and appropriate referral pathways and management (Supplementary,  
4  
5 126 Tables S1-S3).[8]  
6  
7  
8

9 127 **Image acquisition and management:** Images were captured by trained nurses and  
10  
11 128 technicians in Vietnam. Two-field, 45° digital colour photographs (one disc-centred and one  
12  
13 129 macula-centred) were taken using a tabletop non-mydratic fundus camera (Canon CR2-AF,  
14  
15 130 Canon Medical Systems. Europe), in accordance with the UK's NHS DESP.[9] Nurses and  
16  
17 131 technicians were trained to repeat inadequate images as a quality control measure and take  
18  
19 132 anterior segment photographs where adequate fundus images were not possible. Images were  
20  
21 133 anonymised and uploaded to a cloud-based software system (Spectra)® for analysis by  
22  
23 134 trained DR graders in Vietnam. The images were transferred to a Queen's University Belfast  
24  
25 135 (QUB) server for re-grading by the reference standard.  
26  
27  
28  
29

30  
31 136 **Assessment of gradeability:** Image quality was defined as 'adequate' or 'inadequate' in  
32  
33 137 accordance with NHS DESP guidelines as outlined below;  
34  
35

- 36  
37 138 • Adequate disc-centred image: complete optic disc  $>2DD$  from edge of image and fine  
38  
39 139 vessels visible on surface of the disc.[9]  
40  
41 140 • Adequate macula-centred image: centre of fovea  $>2DD$  from edge of image and  
42  
43 141 vessels visible within 1DD of centre of fovea.[9]  
44  
45  
46

47 142 The disc-centred and macula-centred images for each eye were viewed as a pair and graded at  
48  
49 143 an individual eye level. The presence of DR and diabetic macular oedema (DMO) was also  
50  
51 144 determined at a patient level and based on the worst affected eye. Participants with ungradable  
52  
53 145 images were referred for further slit-lamp examination. Where images were considered  
54  
55 146 inadequate but referable disease was detectable, the referable grade was recorded and the  
56  
57 147 patients were moved onto the appropriate referable grade pathway.[9]  
58  
59  
60

1  
2  
3 148 Consecutive patients diagnosed with diabetes and undergoing evaluation for possible DR at  
4  
5 149 Ho Chi Minh City General Hospital and Ho Chi Minh Eye Hospital (tertiary hospitals), Tien  
6  
7  
8 150 Giang General Hospital (provincial hospital) and Cai Ba General Hospital (district hospital)  
9  
10 151 in Vietnam were recruited. Fundus images from August to October 2018 (Phase I) were  
11  
12 152 graded by 14 graders in Vietnam and then re-graded retrospectively by a reference standard  
13  
14 153 from the UK in Phase I. Targeted remedial training, based on specific findings from the  
15  
16  
17 154 Phase I analysis, was delivered in March 2019 and November 2019 by UK graders and Orbis.  
18  
19 155 (Supplementary material, figure S1) Additionally, regular testing and training for quality  
20  
21 156 assurance purposes was also introduced, similar to UK DESP models. To evaluate the impact  
22  
23  
24 157 of this quality-improvement intervention, a new subset of images was graded by six of the  
25  
26 158 original cohort of graders between January-October 2020 (Phase II) and re-graded by the  
27  
28 159 reference standard from the UK (KC) in September 2021.

30  
31  
32 160 **Statistical analysis:** Data were entered into Microsoft Excel version 16.0 and then  
33  
34 161 transferred to Stata 17.0 (StataCorp LLC) for analysis. Intra and inter-grader agreement was  
35  
36 162 calculated using kappa and a stratified random sampling technique was utilised to ensure a  
37  
38 163 representative sample of images was re-graded (Supplementary, Tables S4 and S5).  
39  
40  
41 164 Diagnostic test accuracy (DTA) comparing graders in Vietnam with the UK reference  
42  
43 165 standard was assessed by calculating sensitivity, specificity, positive predicative values  
44  
45 166 (PPV) and negative predictive values (NPV). Sensitivity was calculated at the two different  
46  
47 167 time points (Phase I and Phase II) and Chi Squared was used to test significance.

48  
49  
50  
51 168 **Patient and Public Involvement:** Patients or the public were not involved in the design, or  
52  
53 169 conduct, or reporting, or dissemination plans of our research  
54  
55  
56

57 170  
58  
59  
60

1  
2  
3 171 **Results**  
4

5 172 **Patient characteristics:** In Phase I, 65.4% of patients were female with a mean age 59.4  
6  
7 173 years. In Phase II, 40.0% were female with a mean age of 59.8 years. Figures 1 and 2  
8  
9 174 describe enrolment of patients and capture and grading of images in Phase I and II of the  
10  
11 175 study respectively.  
12  
13  
14

15 176  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

177 Figure 1

178

For peer review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

179 Figure 2

180

For peer review only



181 **Initial grading performance analysis (Phase I):** The sensitivity for detecting any DR was  
 182 low against the reference standard in the UK for all grader groups in Vietnam. The sensitivity  
 183 for detecting referable DR and referable DMO was even lower for all grader groups (Table  
 184 1). The sensitivity increased when ungradable images were excluded from the analysis,  
 185 though it still remained low (47.9-50.8% for any DR; 22.2%-38.1% for referable DR and 9.3-  
 186 26.5% for referable DMO) (Supplementary, Table S6).

187 **Table 1. Diagnostic test accuracy of DR graders in Vietnam against a reference**  
 188 **standard from the UK, including ungradable images.**

|  | <b>Level 1 graders<br/>(n=410 patient<br/>images)*</b> | <b>Level 2 graders<br/>(n=410 patient<br/>images)*</b> | <b>Level 3 graders<br/>(n=260 patient<br/>images) †</b> |
|--|--|--|---|
| <b>Any DR</b>  |  |  |   |
| Sensitivity (%) (95% CI)   | 41.8 (33.9, 50.1)                                      | 42.5 (34.5, 50.7)                                      | 42.2 (33.1, 51.8)                                       |
| Specificity (%) (95% CI)   | 87.9 (83.3, 91.7)                                      | 98.8 (96.6, 99.8)                                      | 100 (97.5, 100)   |
| PPV (%) (95% CI)   | 67.4 (57.0, 76.6)                                      | 95.6 (87.6, 99.1)                                      | 100 (92.7, 100)   |
| NPV (%) (95% CI)   | 71.7 (66.4, 76.7)                                      | 74.3 (69.3, 78.8)                                      | 68.2 (61.5, 74.5)                                       |
| <b>Referable DR</b>  |  |  |   |
| Sensitivity (%) (95% CI)   | 19.2 (9.63, 32.5)                                      | 13.5 (5.59, 25.8)                                      | 10.5 (2.94, 24.8)                                       |
| Specificity (%) (95% CI)   | 97.2 (94.9, 98.7)                                      | 100 (99.0, 100)  | 99.5 (97.5, 100)  |
| PPV (%) (95% CI)   | 50.0 (27.2, 72.8)                                      | 100 (59.0, 100)  | 80.0 (28.4, 99.5)                                       |
| NPV (%) (95% CI)   | 89.2 (85.7, 92.1)                                      | 88.8 (85.3, 91.7)                                      | 86.7 (81.9, 90.6)                                       |
| <b>Referable DMO</b>   |  |  |   |
| Sensitivity (%) (95% CI)   | 5.8 (1.91, 13.0)                                       | 20.9 (12.9, 31.0)                                      | 16.9 (8.76, 28.3)                                       |
| Specificity (%) (95% CI)   | 97.2 (94.8, 98.7)                                      | 99.4 (97.8, 99.9)                                      | 100 (98.1, 100)   |
| PPV (%) (95% CI)   | 35.7 (12.8, 64.9)                                      | 90.0 (68.3, 98.8)                                      | 100 (71.5, 100)   |
| NPV (%) (95% CI)   | 79.5 (75.2, 83.4)                                      | 82.6 (78.4, 86.2)                                      | 78.3 (72.7, 83.3)                                       |
| <b>Abbreviations:</b> UK = United Kingdom, DR = Diabetic Retinopathy, DMO = Diabetic Macular Oedema, CI = Confidence Intervals,<br>Grading criteria: UK National Diabetic Eye Screening Programme (NDESP) classification system (See supplementary material, Table S1 for more details).<br>Any DR, is defined as grades R1, R2, R3s, R3a and U.<br>Referable DR is defined as grades R2, R3a and U<br>Referable DMO is defined as grades M1 and U<br>*Missing (n=2, 0.5%), †missing (n=2, 0.8%) |  |  |   |

Sensitivity is the ability of a test to correctly identify patients with a disease and specificity is the ability of a test to correctly identify people without the disease  
Positive predictive value (PPV) is the proportion of those who test positive who have the condition (true positives) and negative predictive value (NPV) is the proportion of those who test negative who do not have the condition (true negatives).

189

**Subsequent grading performance after retraining (Phase II):** Subsequently, a further subset of images from 115 consecutive patients from January to October 2020 were graded by six of the original cohort of 14 Vietnamese graders, and were regraded in the UK to evaluate graders' performance after targeted training was delivered and quality control measures were instituted. The greatest improvement in sensitivity for detecting any DR was seen among level 1 graders (difference: +45.4%, 95%CI +33.1% to +57.8%;  $P < 0.001$ ). The specificity increased from 87.9% in phase I to 95.6% in phase II which helps to avoid over referrals (difference: +7.7%, 95%CI +1.4% to +13.9%;  $p = 0.069$ ). The lowest improvement in sensitivity for detecting any DR was observed between level 3 graders in Vietnam (difference; +9.0%, 95%CI: -9.0% to +27.1%;  $p = 0.326$ ), although their specificity remained 100% at phase I and phase II. There was a significant improvement in sensitivity for detecting DR and referable DMO at all grader levels: sensitivities after training ranged between 40% and 61.5% (Table 2). Further improvement in sensitivity was observed when ungradable images were excluded from the analysis in Phase II: sensitivities ranged from 55.6 to 97.6% for any DR, 55.6%-88.9% for referable DR and 60.0-90.0% for referable DMO (Supplementary, Table S7). The overall prevalence of DR in this study can be found in Supplementary, Table S8.

207

208

209

211 **Table 2: Diagnostic test accuracy of DR graders in Vietnam against a reference**  
 212 **standard from the UK after additional DR training was delivered.**

|   | <b>Level 1 graders<br/>(n=115 patient<br/>images)</b> | <b>Level 2<br/>graders (n=115<br/>patient images)</b> | <b>Level 3 graders<br/>(n=62 patient<br/>images)</b> |
|---|---|---|--|
| <b>Any DR</b>   |   |   |  |
| Sensitivity (%) (95% CI)  | 87.2 (74.3, 95.2)                                     | 68.1 (52.9, 80.9)                                     | 51.3 (34.8, 67.6)                                    |
| Specificity (%) (95% CI)  | 95.6 (87.6, 99.1)                                     | 95.6 (87.6, 99.1)                                     | 100 (84.6, 100)                                      |
| PPV (%) (95% CI)  | 93.2 (81.3, 98.6)                                     | 91.4 (76.9, 98.2)                                     | 100 (83.2, 100)                                      |
| NPV (%) (95% CI)  | 91.5 (82.5, 96.8)                                     | 81.3 (71.0, 89.1)                                     | 53.7 (37.4, 69.3)                                    |
| P-value comparing sensitivity to<br>Phase I   | P=0.000   | P=0.002   | P=0.326  |
| <b>Referable DR</b>   |   |   |  |
| Sensitivity (%) (95% CI)  | 53.3 (26.6, 78.7)                                     | 40.0 (16.3, 67.7)                                     | 58.3 (27.7, 84.8)                                    |
| Specificity (%) (95% CI)  | 90.0 (82.4, 95.1)                                     | 93.0 (86.1, 97.1)                                     | 100 (92.7, 100)                                      |
| PPV (%) (95% CI)  | 44.4 (21.5, 69.2)                                     | 46.2 (19.2, 74.9)                                     | 100 (59.0, 100)                                      |
| NPV (%) (95% CI)  | 92.8 (85.7, 97.0)                                     | 91.2 (83, 95.9)                                       | 90.7 (79.7, 96.9)                                    |
| P-value comparing sensitivity to<br>Phase I   | P=0.009   | P=0.022   | P=0.001  |
| <b>Referable DMO</b>  |   |   |  |
| Sensitivity (%) (95% CI)  | 56.3 (29.9, 80.2)                                     | 43.8 (19.8, 70.1)                                     | 61.5 (31.6, 86.1)                                    |
| Specificity (%) (95% CI)  | 97.0 (91.4, 99.4)                                     | 93.9 (87.3, 97.7)                                     | 100 (92.6, 100)                                      |
| PPV (%) (95% CI)  | 75.0 (42.8, 94.5)                                     | 53.8 (25.1, 80.8)                                     | 100 (63.1, 100)                                      |
| NPV (%) (95% CI)  | 93.2 (86.5, 97.2)                                     | 91.2 (83.9, 95.9)                                     | 90.6 (79.3, 96.9)                                    |
| P-value comparing sensitivity to<br>Phase I   | P=0.000   | P=0.051   | P=0.002  |
| <p><b>Abbreviations:</b> UK = United Kingdom, DR = Diabetic Retinopathy, DMO = Diabetic Macular Oedema, CI = Confidence Intervals,<br/>           Grading criteria: UK National Diabetic Eye Screening Programme (NDESP) classification system (See supplementary material, Table S1 for more details).<br/>           Criteria:<br/>           Any DR is defined as grades R1, R2, R3s, R3a and U.<br/>           Referable DR is defined as grades R2, R3a and U<br/>           Referable DMO is defined as grades M1 and U<br/>           Chi-squared used to compare sensitivity between Phase I and II.<br/>           Sensitivity is the ability of a test to correctly identify patients with a disease and specificity is the ability of a test to correctly identify people without the disease<br/>           Positive predictive value (PPV) is the proportion of those who test positive who have the condition (true positives) and negative predictive value (NPV) is the proportion of those who test negative who do not have the condition (true negatives).</p> |   |   |  |

213

214 **Discussion**

1  
2  
3 215 Results from our study demonstrated poor sensitivity and specificity for detecting all levels of  
4  
5 216 DR, especially referable DR, in the early stages of programme delivery. This translates into  
6  
7  
8 217 increased costs to the health care system due to missed opportunities for early treatment and  
9  
10 218 un-necessary examinations for false-positive referrals. The quality of patient care also suffers.  
11  
12 219 Didactic DR training was delivered to graders in Vietnam over a two-year period by trained  
13  
14 220 DR graders from the UK and Vietnam. Training was specifically targeted to address  
15  
16  
17 221 problems identified in the Phase I testing [10], and quality control testing using international  
18  
19 222 test and training (iTAT) were also undertaken. The iTAT is an online platform offering  
20  
21 223 monthly quality assurance and training for graders who work in DR screening. It is a useful  
22  
23 224 platform for graders to improve their skills in the detection of DR from ophthalmic images. In  
24  
25  
26 225 the UK, it is compulsory for graders to complete monthly test sets (each set consisting of 20  
27  
28 226 retinal images with a range of DR severities). If graders fall below the agreed threshold,  
29  
30 227 additional training and support is provided.[7] This study demonstrates that these steps led to  
31  
32 228 improved grading accuracy for all classes of patients and graders. We found that the main  
33  
34 229 discordance between graders lay in their ability to detect ungradable images; therefore,  
35  
36 230 targeted training must be given to ensure such patients are referred to the next level (slit-lamp  
37  
38 231 examination).

39  
40  
41  
42  
43 232 According to the UK National Institute for Clinical Excellence (NICE) guidelines, DR  
44  
45 233 screening tests must have at least 80% sensitivity and 95% specificity with a technical failure  
46  
47 234 of 5% or less.[11] These requirements were not met here for sensitivity, but may not be  
48  
49 235 applicable to LMICs. Results can be poor in these settings for a variety of reasons, including  
50  
51 236 higher prevalence of un-operated lens opacity impacting clarity of photographs, use of nurses  
52  
53 237 rather than professional photographers for image capture and poor compliance with  
54  
55 238 photography among patients who have not previously undergone such examinations.  
56  
57  
58  
59  
60

1  
2  
3 239 Quality assessment in such settings is crucial, and programmatic changes based on models  
4  
5 240 such as the UK DESP can be successful in enhancing grader accuracy in LMICs settings.  
6  
7  
8 241 However, it is important for countries to adapt their own DR classification system and  
9  
10 242 referral pathways to meet their requirements. As an example, the UK system (England, Wales  
11  
12 243 and Northern Ireland) uses the grade M0 for no maculopathy and M1 for referable  
13  
14 244 maculopathy. In Scotland, M0 denotes no maculopathy, M1 observable maculopathy and M2  
15  
16 245 referable maculopathy allowing some monitoring of maculopathy to take place on screening  
17  
18 246 level. This reduces the burden on the hospital system. The implication for LMICs is that  
19  
20 247 being aware of hospital capacity at the planning stage might mean that they need to safely  
21  
22 248 adapt an accepted grading system to their needs. Most importantly, the role of affiliated  
23  
24 249 hospitals (and partnerships, coordination among training institutions and practical hospitals)  
25  
26 250 are crucial for DR grading quality improvement.  
27  
28  
29  
30  
31  
32 251 Studies in LMICs have assessed the accuracy of non-medical graders and medical graders in  
33  
34 252 the detection of DR and found that both grader types are capable of achieving moderate-high  
35  
36 253 sensitivity for detecting DR.[12-15] Comparable with our findings, a study in China found  
37  
38 254 that non-medical DR graders achieved higher sensitivity (0.82-0.94%) and specificity (0.91-  
39  
40 255 0.98%) compared to rural ophthalmologists (sensitivity=0.65-0.95%, specificity=0.59-0.95%)  
41  
42 256 [16]. In DR screening, it is vital to detect referable and STDR to prevent blindness, but it is  
43  
44 257 equally important to detect normal cases to prevent unnecessary referrals to already  
45  
46 258 overburdened hospital clinics. Screening provides an opportunity for graders to discuss with  
47  
48 259 patients the importance of managing diabetes to reduce the risk of visual impairment from  
49  
50 260 DR.  
51  
52  
53  
54  
55  
56 261 Some studies have described what training interventions were used to train their graders and  
57  
58 262 key elements may be incorporated into our training programme in the future.[15, 17-18] In  
59  
60

1  
2  
3 263 the UK, the DR grading course by the Gloucestershire Retinal Education Group is required  
4  
5 264 for grader certification. The high costs of this course may be more challenging in LMICs due  
6  
7  
8 265 to limited funding.[6]  
9

10  
11 266 **Strengths:** This study describes the impact of a training intervention to improve the quality  
12  
13 267 of DR grading in an LMIC. The inclusion of ungradable images in this study was a logical  
14  
15 268 decision, particularly when the prevalence of cataract (which often renders DR images  
16  
17 ungradable) is high in LMICs.[19] Dense cataracts normally obstruct the view of the fundus,  
18 269 making it difficult to obtain clear fundus photographs and assign a DR grade. In these  
19  
20 270 instances, referring patients to an eye clinic for further assessment and treatment as needed is  
21  
22 271 required. Determining sensitivity and specificity at the patient level is also important from a  
23  
24 272 DESP implementation perspective. In the UK and Vietnam, both eyes are typically examined  
25  
26 273 for DR and a single outcome is assigned to the patient, as was done here. For these reasons,  
27  
28 274 we feel our analytic approach, and thus results, are relevant to these settings.  
29  
30  
31  
32  
33

34  
35 276 **Limitations:** Limitations for this study have also been acknowledged. Data from this study  
36  
37 277 represent routine clinical practice. In daily DR screening, not all patients undergoing primary  
38  
39 278 (Level 1) and secondary (Level 2) grading proceed to arbitration grading (Level 3). This  
40  
41 279 means a proportion of images were not graded by arbitration graders as outlined in figure 1  
42  
43 280 and figure 2. Second, only 6/14 graders from Phase I were included in Phase II grading;  
44  
45 281 however the distribution of grader levels was similar. Third, though the proportion of patients  
46  
47 282 excluded was small, we are unable to fully characterise the reasons for these exclusions, due  
48  
49 283 to the nature of the study as a programmatic evaluation. Some potential reasons for this are a  
50  
51 284 patient's unwillingness to participate in the study, graders having forgotten to ask for patient  
52  
53 285 consent to participate in the study, and patient inability to comply with image capture. Fourth,  
54  
55 286 pupil status (size and cataract status) was not recorded in this study and this can be important  
56  
57  
58  
59  
60

1  
2  
3 287 for LMICs. Finally, it was not practical for the UK reference standard to examine patients  
4  
5 288 clinically in Vietnam; however, the method of grading by a certified DR grader or clinical  
6  
7  
8 289 specialist is widely used as the reference standard in many screening programmes.  
9

10  
11 290 **Conclusions:** This paper shows how grading accuracy was particularly low among all grader  
12  
13 291 groups in Vietnam in the first six months of DESP implementation. Many factors may have  
14  
15 292 contributed to poor grader performance, including inadequate training and feedback,  
16  
17 293 insufficient time to participate in quality assurance testing and competing work  
18  
19 294 responsibilities. After additional training, testing and quality assurance systems were  
20  
21 295 implemented in Vietnam, DTA improved among all grader groups, however more work is  
22  
23 296 still needed. In particular, training graders to detect ungradable cases is crucial. With  
24  
25 297 continuous quality improvement, monthly international test and training, periodic DR  
26  
27 298 workshops and reviewal of certification, we would expect the DR sensitivity and specificity  
28  
29 299 to improve further. A qualitative study to determine why the initial training intervention was  
30  
31 300 less successful should be explored. Since further improvements are required, understanding  
32  
33 301 how other countries implement such programmes would be beneficial. Future studies should  
34  
35 302 outline what DR training interventions were used, state relevant training courses and explain  
36  
37 303 what quality assurance measures are in place. The findings from this study are important for  
38  
39 304 DESP programme planners in Vietnam and other LMICs, highlighting the importance of  
40  
41 305 quality monitoring and directed re-training as needed.  
42  
43  
44  
45  
46  
47  
48

49 306 Artificial intelligence (AI) is likely to significantly change future approaches to DR grading.  
50  
51 307 Continued attention to maximising accuracy of human graders is still highly relevant today,  
52  
53 308 especially for low-resource settings, as AI systems must be validated locally against a gold  
54  
55 309 standard of proven expert human graders. Differences between the high-quality images used  
56  
57 310 to train most existing AI systems and the types of images encountered in low-resource  
58  
59  
60

1  
2  
3 311 settings, with high rates of prevalent lens opacity, less-well-trained photographers, and lower-  
4  
5 312 cost cameras, mean that such validation must almost certainly occur at the local level. The  
6  
7  
8 313 continued importance of reliable human graders in low-resource settings is further  
9  
10 314 underscored by the fact that few systems are able to function fully autonomously without  
11  
12 315 input from existing graders.

13  
14  
15  
16 316 **Acknowledgements:** The authors are very grateful to all participants in Vietnam and  
17  
18 317 Northern Ireland who generously gave their time for this study and to Orbis International.

19  
20  
21 318 **Author contributions:** All authors (KC, NC, TTH, LL, VTN, HTN, QHN, CD, GV, PP, HT,  
22  
23 319 RS, MQT, TP) have made substantial contributions to the conception of the design of the  
24  
25 320 work. KC and GV analysed and interpreted the data. All authors (KC, NC, TTH, LL, VTN,  
26  
27 321 HTN, QHN, CD, GV, PP, HT, RS, MQT, TP) read and approved the final manuscript.

28  
29  
30  
31 322 **Funding:** The project was funded by the Department for the Economy (DfE) – Global  
32  
33 323 Challenges Research Fund (GCRF) Awards (Grant number: DFEGCRF17-18/Peto). None of  
34  
35 324 the funders were involved in the design or conduct of the study; preparation, review or  
36  
37 325 approval of the manuscript; or decision to submit the manuscript for publication. Nathan  
38  
39 326 Congdon is supported by the Ulverscroft Foundation (UK) (Grant number: N/A). Katie  
40  
41 327 Curran is funded by the Wellcome Trust (Grant number: 204835/Z/16/A).

42  
43  
44  
45  
46 328 **Competing Interests:** Nathan Congdon is employed as a Research Director by Orbis  
47  
48 329 International.

49  
50  
51  
52 330 **Patient consent for publication:** Not required.

53  
54  
55  
56 331 **Ethics approval:** This study involves human participants and was approved by an Ethics  
57  
58 332 Committee or Institutional Board. This research adhered to the tenets of the Declaration of



1  
2  
3 333 Helsinki. Ethical approval was granted by the Hanoi Medical University Institutional Review  
4  
5 334 Board in Bio-Medical Research, Vietnam (No. 0518/HMU IRB). Written informed consent  
6  
7  
8 335 was obtained from all participants prior to their being interviewed.  
9

10  
11 336 **Data sharing statement:** All data relevant to the study are included in the article or uploaded  
12  
13 337 as supplementary information.  
14  
15

## 16 17 338 **References**

- 18  
19  
20 339 1. Ngoc NB, Lin ZL, Ahmed W. Diabetes: What challenges lie ahead for Vietnam? *Ann*  
21  
22 340 *Glob Heal.* 2020;86(1):1–9.  
23  
24  
25 341 2. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness  
26  
27 342 certifications in England and Wales in working age adults (16-64 years), 1999-2000 with  
28  
29 343 2009-2010. *BMJ Open.* 2014;4(2):1–6.  
30  
31  
32 344 3. Burgess PI, Msukwa G, Beare NAV. Diabetic retinopathy in sub-Saharan Africa:  
33  
34 345 Meeting the challenges of an emerging epidemic. *BMC Med.* 2013;11(1):1.  
35  
36  
37 346 4. Adriono G, Wang D, Octavianus C, Congdon N. Use of eye care services among  
38  
39 347 diabetic patients in urban Indonesia. *Arch Ophthalmol* (Chicago, Ill 1960). 2011;129(7):930.  
40  
41  
42 348 5. International Council of Ophthalmology. Number of Ophthalmologists in Practice and  
43  
44 349 Training Worldwide [Internet]. 2012 [cited 2020 Aug 13]. Available from:  
45  
46 350 <http://www.icoph.org/ophthalmologists-worldwide.html>  
47  
48  
49 351 6. Gloucestershire Retinal Education Group. Diabetic Retinopathy Grading Course  
50  
51 352 [Internet]. 2020. [cited 2021 April] Available from: [https://drscreening.org/diabetic-](https://drscreening.org/diabetic-retinopathy-grading-course)  
52  
53 353 [retinopathy-grading-course](https://drscreening.org/diabetic-retinopathy-grading-course)  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 354 7. Widdowson S, Taylor D. The management of grading quality: Good practice in the  
4  
5 355 quality assurance of grading. [Internet] 2015. [cited 2021 April] Available from:  
6  
7 356 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data)  
8  
9 357 [/file/512832/The\\_Management\\_of\\_Grading.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/512832/The_Management_of_Grading.pdf)  
10  
11  
12  
13 358 8. Public Health England. NHS Diabetic Eye Screening Programme Grading definitions  
14  
15 359 for referable disease. [Internet] 2021 (cited 2021 September) Available from:  
16  
17 360 [https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-grading-](https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-grading-criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-referable-disease)  
18  
19 361 [criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-referable-disease](https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-grading-criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-referable-disease)  
20  
21  
22  
23 362 9. NHS DESP Core Team. Diabetic Eye Screening Programme Pathway for Adequate /  
24  
25 363 Inadequate Images and where images cannot be taken Outlines pathway and business rules  
26  
27 364 for image capture exceptions and ungradable images. [Internet] 2013.[cited 2021 April]  
28  
29 365 Available from: [https://www.gov.uk/government/publications/diabetic-eye-screening-](https://www.gov.uk/government/publications/diabetic-eye-screening-pathway-for-images-and-where-images-cannot-be-taken)  
30  
31 366 [pathway-for-images-and-where-images-cannot-be-taken](https://www.gov.uk/government/publications/diabetic-eye-screening-pathway-for-images-and-where-images-cannot-be-taken)  
32  
33  
34  
35 367 10. Curran K, Congdon N, Peto T, Dardis C, Nguyen NN, Hoange TT et al. Capturing the  
36  
37 368 clinical decision-making processes of expert and novice diabetic retinal graders using a  
38  
39 369 ‘think-aloud’ approach. *Eye*. 2022;36:1019-1026  
40  
41  
42  
43 370 11. Scanlon PH. The English national screening programme for diabetic retinopathy  
44  
45 371 2003–2016. *Acta Diabetol*. 2017;54(6):515–525.  
46  
47  
48 372 12. Amirul Islam FM. Accuracy and reliability of retinal photo grading for diabetic  
49  
50 373 retinopathy: Remote graders from a developing country and standard retinal photo grader in  
51  
52 374 Australia. *PLoS One*. 2017;12(6):1–12.  
53  
54  
55 375 13. Bhargava M, Cheung CYL, Sabanayagam C, Kawasaki R, Harper CA, Lamoureux  
56  
57 376 EL, et al. Accuracy of diabetic retinopathy screening by trained non-physician graders using  
58  
59 377 non-mydratric fundus camera. *Singapore Med J*. 2012;53(11):715–9.

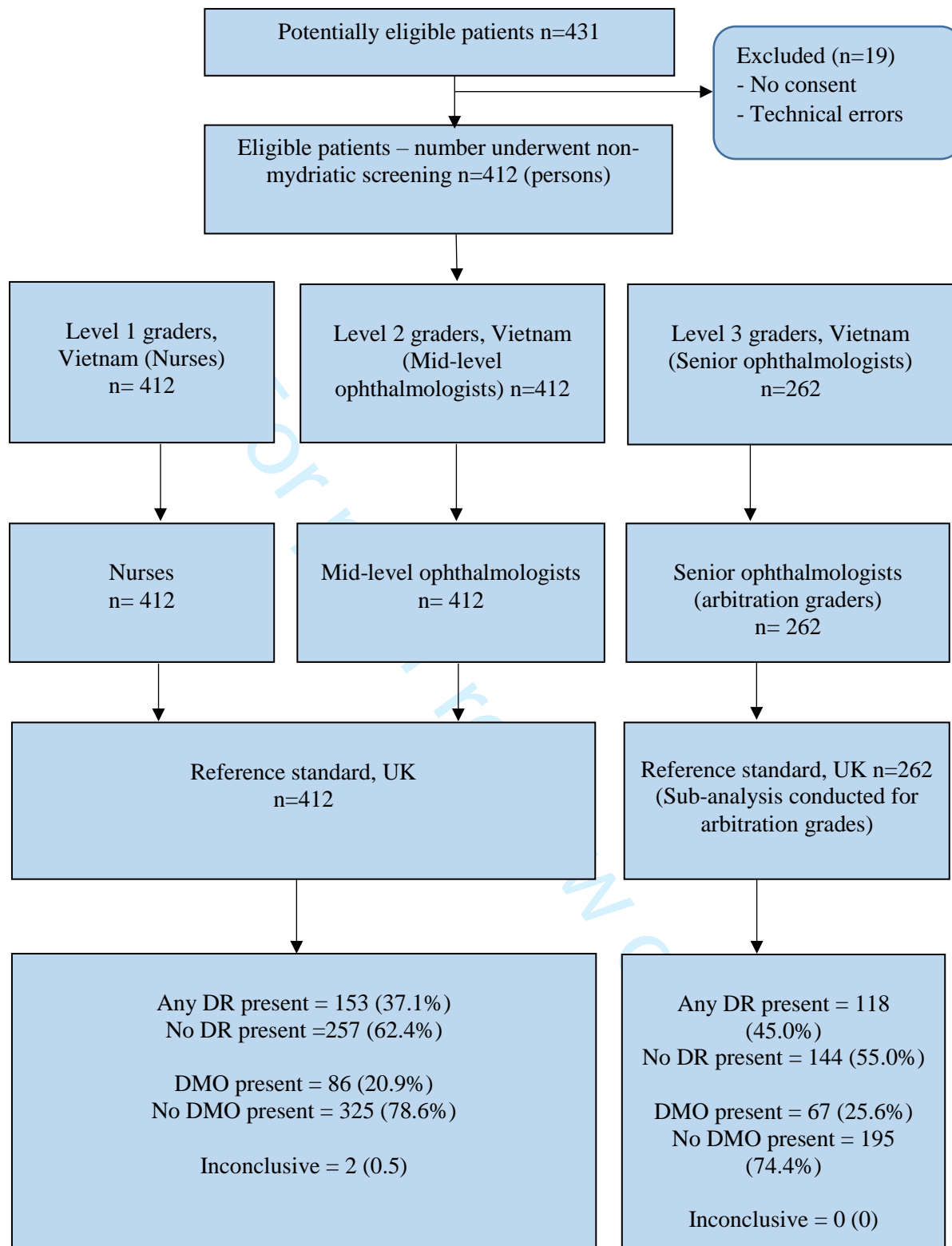
- 1  
2  
3 378 14, Piyasena MMPN, Yip JLY, MacLeod D, Kim M, Gudlavalleti VSM. Diagnostic test  
4  
5 379 accuracy of diabetic retinopathy screening by physician graders using a hand-held non-  
6  
7 380 mydriatic retinal camera at a tertiary level medical clinic. *BMC Ophthalmol.* 2019;19(1):89.  
8  
9  
10 381 15. Thapa R, Bajimaya S, Pradhan E, Sharma S, Kshetri B, Paudyal G. Agreement on  
11  
12 382 Grading Retinal Findings of Patients with Diabetes Using Fundus Photographs by Allied  
13  
14 383 Medical Personnel when Compared to an Ophthalmologist at a Diabetic Retinopathy  
15  
16 384 Screening Program in Nepal. *Clin Ophthalmol.* 2020;14:2731-2737.  
17  
18  
19 385 16. McKenna M, Chen T, McAneney H, Membrillo MAV, Jin L, Xiao W, et al. Accuracy  
20  
21 386 of trained rural ophthalmologists versus non-medical image graders in the diagnosis of  
22  
23 387 diabetic retinopathy in rural China. *Br J Ophthalmol.* 2018;102(11):1471–6.  
24  
25  
26 388 17. Rani PK, Peguda HK, Chandrashekher M, Swarna S, Jonnadula GB, James J et al.  
27  
28 389 Capacity building for diabetic retinopathy screening by optometrists in India: Model  
29  
30 390 description and pilot results. *Indian J Ophthalmol.* 2021.69:655-659.  
31  
32  
33 391 18. Srinivasan S, Shetty S, Natarajan V, Sharma T, Raman R. Development and  
34  
35 392 Validation of a Diabetic Retinopathy Referral Algorithm Based on Single-Field Fundus  
36  
37 393 Photography. *PLoS One* 2016.11:e0163108.  
38  
39  
40 394 19. Murthy GVS, John N, Shamanna BR, Pant HB. Elimination of avoidable blindness  
41  
42 395 due to cataract: Where do we prioritize and how should we monitor this decade. *Indian*  
43  
44 396 *J Ophthalmol.* 2012;60(5):438–45. (19)

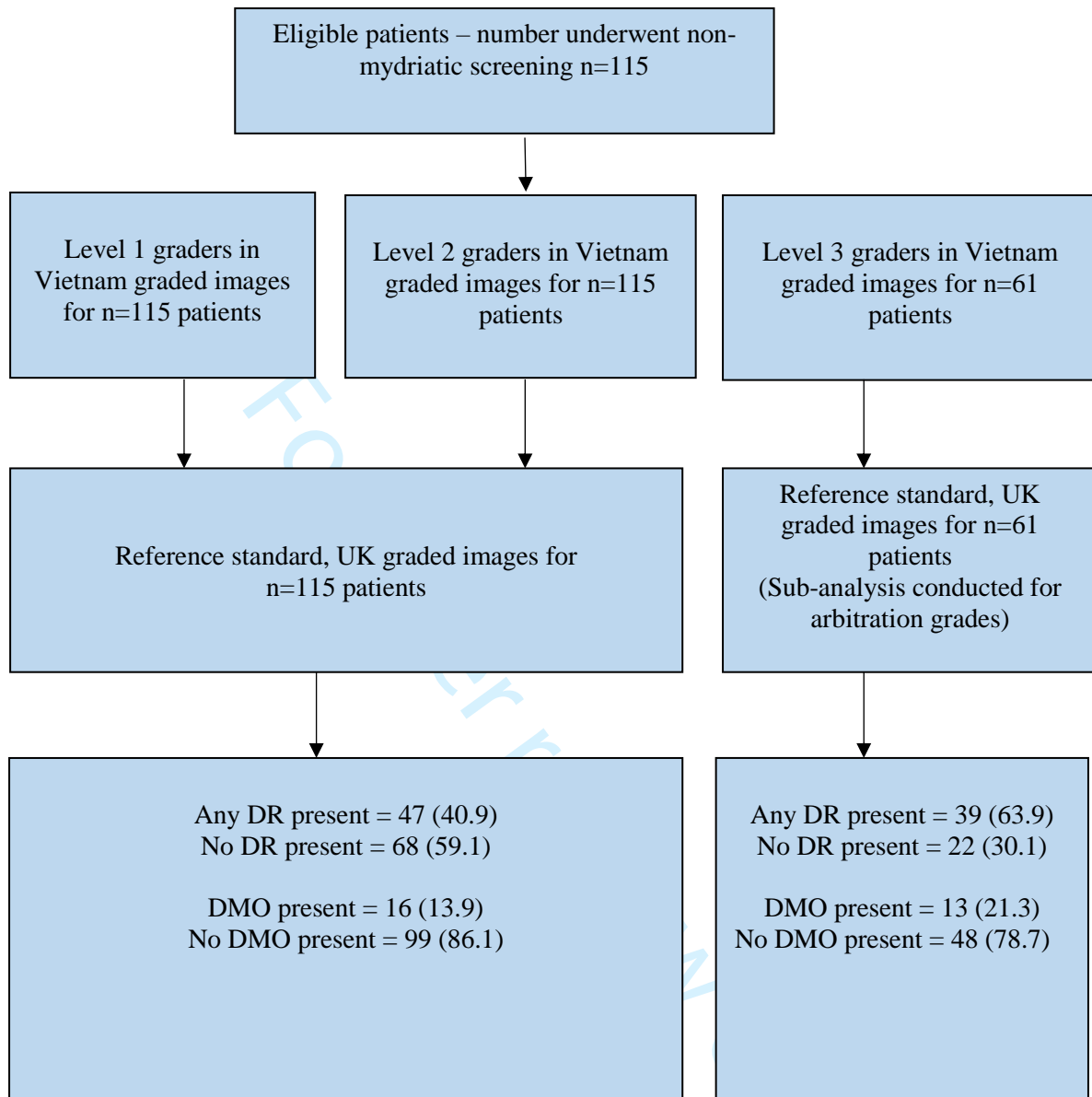
45  
46  
47  
48  
49  
50  
51 397 Figure 1: Flow diagram to illustrate enrolment of patients and management of images in Phase I from  
52  
53 398 August to October 2018 (Initial grading performance analysis). Level 1 and level 2 graders graded the  
54  
55 399 same set of photographs and level 3 graders graded a subset of these photographs: All disagreements  
56  
57 400 between Level 1 and 2 graders and a 40% random sample of all images.  
58  
59  
60

1  
2  
3 401 Figure 2: Flow diagram illustrating the enrolment of patients and management of images included in  
4  
5 402 Phase II from January 2020 to October 2020 (Follow-up grading performance analysis after re-  
6  
7 403 training). Level 1 and level 2 graders graded the same set of photographs and level 3 graders graded a  
8  
9 404 subset of these images.  
10

11  
12  
13 405  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only





## SUPPLEMENTARY MATERIAL

### Figure S1: Training timeframe for graders in Vietnam

#### September 2017 (5 days)

- Five Vietnamese doctors and medical administrators visited Northern Ireland to receive screener/graders training in administration and failsafe methods.
- These doctors delivered training to new graders in Vietnam following their visit to Northern Ireland.

#### January 2018 (1 week) – Delivered in person in Vietnam by two UK graders (senior ophthalmic nurse and optometrist, certified in DR grading)

- Observation in retina clinics.
- Hands on training with tabletop CR-2 Canon Fundus Cameras.
- Topics covered: ocular anatomy, retinal disease, DR signs, DR grading (based on the UK DESP grading classification system) and appropriate referral pathways and management (PowerPoint presentation and interactive sessions).
- All graders received a module workbook.
- Certification provided by Orbis.

#### March 2018 - Grading began in Vietnam

- Graders began grading as part of pilot DESP.
- Ongoing training was delivered by the Orbis team and the lead ophthalmologist for DR screening in Vietnam over the course of the following months.

#### June 2018 (1-2 days) – Delivered by Orbis partners

- UK graders developed a PowerPoint presentation based on DR case examples and this was delivered by Orbis.

#### March 2019 (2 days) - Delivered by UK grader in Vietnam

- More training on DR case examples.

#### November 2019 (3 days) - Delivered in person in Vietnam by two UK graders (senior ophthalmic nurse and optometrist, certified in DR grading)

- Refresher DR training, incorporating think-aloud techniques into practical teaching sessions.
- Pre and post training assessments.
- Encouraged use of international test and training (iTAT) for quality assurance purposes. Practical sessions on iTAT.

**Table S1: Diabetic retinopathy workshop for graders in Vietnam**

| <b>AGENDA</b>  |                                      |                           |
|--|--------------------------------------|---------------------------|
| <b>DAY 1-3 Monday 29<sup>th</sup> January – Wednesday 31<sup>st</sup> January</b>                |                                      |                           |
| Visit diabetic retinopathy screening sites to observe and provide hands on training and support. |                                      | UK graders and Orbis team |
| <b>DAY 4- Thursday 1<sup>st</sup> February</b>   |                                      |                           |
| <b>MORNING</b>   |                                      |                           |
| <b>9:00 - 9:30</b>   | Check in                             | UK graders and Orbis team |
| <b>9:30 – 10.30</b>  | Introduction on Diabetic Retinopathy |                           |
| <b>10.30-10.45</b>   | Tea Break                            |                           |
| <b>10.45-11.30</b>   | Basic Screening Component            |                           |
|  | Lunch                                |                           |
| <b>AFTERNOON</b>   |                                      |                           |
| <b>13.30-14:00</b>   | NHS Grading System                   | UK graders and Orbis team |
| <b>14:00-14:45</b>   | Image Quality                        |                           |
| <b>14:45-15:00</b>   | Tea Break                            |                           |
| <b>15:00-16:00</b>   | Image Grading                        |                           |
| <b>16:0-16:30</b>  | Hospital pathway                     |                           |
| <b>DAY 5- Friday 2<sup>nd</sup> February</b>   |                                      |                           |
| <b>MORNING</b>   |                                      |                           |
| <b>9:00-9:30</b>   | Other ocular findings                | UK graders and Orbis team |
| <b>9:30-11:30</b>  | Practice on grading                  |                           |
| <b>11:30-12:00</b>   | Wrap up                              |                           |



**Table S2: Refresher training programme for DR graders in Vietnam (delivered by UK graders and Orbis) in November 2019**

| Time               | Topics  | Method        | Who   | Preparation |
|--------------------|---|---------------|---|-------------|
| <b>DAY 1</b>       |   |               |   |             |
| <b>8:45-9:15</b>   | Introduction/ Pre course quiz   |               | Orbis VN/ UK graders                                      | UK graders  |
| <b>9:15-9:45</b>   | <b>Diabetic Retinopathy (DR) - New Challenges of Blindness Prevention</b><br><i>Objective: Understand the problem of DR and current efforts to manage vision loss. Aim to motivate graders to be involved in DESP</i> | Presentation  | Orbis VN  | Orbis       |
| <b>9:45-10:15</b>  | <b>Retina Anatomy</b><br><i>Objective : Understand the pathobiology of diabetic complications and pathogenesis of retinal damage</i>  | Presentation  | Ho Chi Minh Eye Hospital                                  | Orbis       |
| <b>10:30-11:15</b> | <b>Diabetic Retinopathy (DR) Pathophysiology</b><br><i>Objective : Understand Diabetic Retinopathy</i>  | Presentation  | UK graders  | UK graders  |
| <b>11:15-12:00</b> | <b>Grading system and DR Grading pathway ( UK system)/ How to systematically grade a retinal image</b><br><i>Objective : Understanding the grading system and referral pathway ( UK standard)</i>                     | Presentation  | UK graders  | UK graders  |
| <b>13:30-14:15</b> | <b>Image quality</b><br><i>Objective : Understand the requirements/criteria of image quality for accurate grading</i>   | Presentation  | UK graders  | UK graders  |
| <b>14:15-14:45</b> | <b>Spectra Software</b><br><i>Objective : How to use the current Spectra software for uploading, grading, and managing DR cases</i>   | Demonstration | Senior graders of Tien Giang and Ho Chi Minh Eye Hospital |             |

|              |   |                    |  |   |
|--------------|---|--------------------|--|---|
| 15:00-17:00  | <p><b>Practical Training</b><br/> <b>Parallel session:</b><br/> <b>1- Taking retina images of the patients following the procedure ,and provide counseling to the patients</b></p> <p><i>Objective: Practice experience of taking fundus images</i></p> | Practical training | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and UK graders | 2 fundus cameras:<br>4 Groups: make sure every participant is able to practice at least once        |
| 15:00-17:00  | <p><b>Parallel session :</b><br/> <b>1- Grading DR in the Spectra</b></p> <p><i>Objective: Practical experience of how to do DR grading</i></p>   | Practical training | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and UK graders | 3-4 accounts of Spectra<br>4 Groups : make sure every participant is able to practice at least once |
| <b>DAY 2</b> |   |                    |  |   |
| 08:45-09:00  | <i>Recap of day 1/ introduction to day 2</i>  |                    | Orbis  |   |
| 9:00-9:30    | <p><b>Analysis of retinal fundus images for grading of diabetic retinopathy severity</b><br/> Objective : How to read the image and protocol for retinal image analysis</p>   |                    | Ophthalmologist Ho Chi Minh Eye Hospital                                   |   |
| 9:30-10:15   | <p><b>DR Screening Procedure: Best Practice</b><br/> <i>Objective: Discuss how to build the “best screening procedures” into DESPs.</i></p>   | Presentation       | UK graders   | UK graders  |
| 10:30-11:15  | <p><b>Other Ocular Findings</b><br/> <i>Objective: Awareness of other ocular pathology during DR screening</i></p>  | Presentation       | UK graders   | UK graders  |
| 11:15-12:15  | <b>Image grading case studies competition</b>   | Practical          | UK graders   | We need to organise people into groups of 3 with one experienced grader in each group               |

|  |  |   |                                     |   |  |
|--|--|---|-------------------------------------|---|--|
| 1<br>2<br>3<br>4<br>5<br>6<br>7<br>8<br>9<br>10                                  | <b>13:30-14:30</b>                                       | <b>Counselling and delivering messages to patients during the DR screening</b><br><i>Objective: Important to provide counselling for the patients and deliver messages effectively.</i>   | Presentation/<br>practical training | Orbis Vietnam   |  |
| 11<br>12<br>13<br>14<br>15<br>16<br>17<br>18<br>19<br>20<br>21<br>22<br>23<br>24 | <b>14:30-16:30</b>                                       | <b><u>Practical Training</u></b><br><b>Parallel session:</b><br><b>1- Taking retina images of the patients following the procedure ,and provide counseling to the patients</b><br><br><i>Objective: Experience on how to take good fundus photographs</i> | Practical training                  | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and UK graders              | 2 fundus camera<br>4 Group: make sure every participants are able to practice at least one time        |
| 25<br>26<br>27<br>28<br>29<br>30<br>31<br>32<br>33                               |  | <b>Parallel session :</b><br><b>1- Grading DR in the Spectra</b><br><br><i>Objective: Practical experience on DR grading</i>  | Practical training                  | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and UK graders              | -4 accounts of Spectra<br>4 Groups: make sure every participant is able to practice at least one time. |
| 34<br>35   | <b>DAY 3</b><br><b>Part 1 (Final practical training)</b> |   |                                     |   |  |
| 36<br>37   | <b>08:45-9:00</b>  | Recap of day 2/ introduction to day 3   |                                     |   |  |
| 38<br>39<br>40<br>41<br>42<br>43<br>44<br>45<br>46<br>47                         | <b>09:00-09:30</b>                                       | <b>Quality Assurance in Diabetic Screening</b><br><i>Objective : Understand why quality assurance is important and the correct steps required to ensure good quality assurance procedures are in place</i>  | Presentation                        | UK graders  | UK graders   |
| 48<br>49<br>50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59<br>60       | <b>09:30-10:30</b>                                       | <b>Practice : Grading in iTAT</b><br><i>Objective: Know the Online training for DR grading and the importance of lifelong learning for DR grading</i>   | Practice                            | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and support from UK graders | ITAT accounts for practicing   |

|  |   |   |                       |  |   |
|--|---|---|-----------------------|--|---|
| 1<br>2<br>3<br>4<br>5<br>6<br>7<br>8<br>9<br>10<br>11<br>12<br>13<br>14  | <b>10:45-12:30</b>                              | <b><u>Assessment</u></b><br><br><b>Parallel session:</b><br><b>1- Taking retina images of the patients following the procedure ,and provide counseling to the patients</b><br><b>2- Grading DR in the Spectra</b>     | Practical Training    | Current graders of Tien Giang and Ho Chi Minh Eye Hospital, and UK graders | 2 fundus cameras<br>4 Group and 4 accounts of Spectra<br>4 Groups |
| 15<br>16   | <b>DAY 3</b><br><b>Part 2 (Future planning)</b> |   |                       |  |   |
| 17<br>18<br>19   | <b>13:30-14:15</b>                              | <b>Post course Quiz and results</b>   |                       | UK graders   | UK graders  |
| 20<br>21<br>22<br>23<br>24   | <b>14:15-15:00</b>                              | <b>Teaching Methodology for adults</b><br><i>Objective : How to train new graders effectively</i>   | Think aloud work shop | Orbis VN Supported by UK graders   | UK graders  |
| 25<br>26<br>27<br>28<br>29<br>30<br>31<br>32<br>33   | <b>15:15-16:00</b>                              | <b>Supportive Supervision Methodology: Developing quality improvement</b><br><i>Objective: How to plan, implement the supervision trips to correct / improve other graders' performances. Provide checklist tools</i> | Think aloud work shop | Orbis VN Supported by UK graders   | UK graders  |
| 34<br>35<br>36   | <b>16:00-16:45</b>                              | <b>Feedbacks and Plan for next steps</b>  | Discussion            |  |   |
| 37<br>38<br>39<br>40<br>41<br>42<br>43<br>44<br>45<br>46<br>47<br>48<br>49<br>50<br>51<br>52<br>53<br>54<br>55<br>56<br>57<br>58<br>59<br>60 | <b>16:45-17:00</b>                              | <b>Certificates for Vietnamese graders in attendance</b>  |                       |  |   |

Table S3: UK DR Grading Classification Scale

| NSC        | International Term                                     | Symptoms                         | Features  | Action                 |
|------------|--|----------------------------------|---|------------------------|
| <b>R0</b>  | No DR  | None                             | No signs of diabetic retinopathy  | Annual rescreen        |
| <b>R1</b>  | Mild non-proliferative (mild pre-proliferative)        | None                             | Haemorrhages & microaneurysms, only   | Annual rescreen        |
| <b>R2</b>  | Moderate non-proliferative, moderate pre-proliferative | None                             | Extensive Microaneurysms, intraretinal haemorrhages, hard exudates, venous abnormalities, large blot haemorrhages, cotton wool spots (small infarcts), venous beading, venous loop, venous reduplication.   | Refer routinely to HES |
| <b>R3s</b> | Stable proliferative diabetic retinopathy              |                                  | No haemorrhages or exudates or new vessels, laser scars   | Annual rescreen        |
| <b>R3a</b> | Active proliferative diabetic retinopathy              | Floaters, central loss of vision | New vessel formation either at the disc (NVD) or elsewhere (NVE). Extensive fibrovascular proliferation, retinal detachment, pre-retinal or vitreous haemorrhage.   | Urgent referral to HES |
| <b>M0</b>  |  |                                  | No maculopathy  | Annual rescreen        |
| <b>M1</b>  | Diabetic maculopathy                                   | Blurred central vision           | <p>The macula is defined as a circle centred on the fovea, with a radius of the distance to the disc margin.</p> <p>If the leakage involves or is near the fovea the condition is termed clinically significant macular oedema (CSME).</p> <p>Exudative maculopathy presents with leakage, retinal thickening, microaneurysms, hard exudates at the macula. Ischaemic form can have a featureless macular with NVE and poor vision.</p> <p>Milder forms:</p> <ul style="list-style-type: none"> <li>• exudate &lt; or = 1DD of centre of fovea</li> <li>• circinate or group of exudates within macula</li> <li>• any microaneurysm or haemorrhage &lt; or = 1DD of centre of fovea only is associated with a best VA of &lt; or = 6/12 retinal thickening &lt; or =</li> </ul> | Refer to HES           |

|   |                  |                                |   |                                       |
|---|------------------|--------------------------------|---|---------------------------------------|
|   |                  |                                | 1DD of centre of fovea<br>(if stereos available)  |                                       |
| <b>P</b>  | Photocoagulation | Reduced night<br>vision, glare | Small retinal scars throughout<br>the peripheral retina.  |                                       |
| <b>U</b>  | Ungradable       |                                | Ungradable is usually due to<br>cataract, small pupils, other<br>lesions usually referred for<br>assessment | Refer for slit<br>lamp<br>examination |
| <b>Abbreviations:</b> DR = diabetic retinopathy, NPDR = none-proliferative retinopathy, NVE = new vessels elsewhere, IRMAs = intraretinal microvascular abnormalities (part of severe pre-proliferative retinopathy, vessels will not leak with angiogram, otherwise they would be 'new vessels' making the condition 'proliferative'), MO=macular oedema, MA= microaneurysm, DD=disc diameter, HES= hospital eye service |                  |                                |   |                                       |

**Table S4: Reference standards intra-rater agreement score using kappa statistic (first attempt versus second attempt)**

|  | <b>Intra-rater agreement<br/>(reference standard, UK),<br/>k (95% CI)<br/>(by eyes, n=106)</b> | <b>Intra-rater agreement<br/>(reference standard, UK), k<br/>(95% CI)<br/>(by worst eye, n=53)</b> |
|--|--|--|
| <b>Overall Diabetic Retinopathy Grading:</b>   |  |  |
| <b>Any DR</b>  | 0.96 (0.91, 1.00)  | 0.92 (0.82, 1.00)  |
| <b>Treatable DR</b>  | 0.81 (0.60, 1.00)  | 0.74 (0.47, 1.00)  |
| <b>Referable Maculopathy</b>   | 0.97 (0.92, 1.00)  | 1.00 (1.00, 1.00)  |
| Abbreviations: CI=confidence interval, k=kappa, DR=Diabetic retinopathy, DMO=diabetic macular oedema<br>Any DR defined as R1, R2, R3s, R3a and U<br>Treatable DR defined as R3a<br>Referable DMO defined as M1 and U |  |  |

**Table S5: Using kappa statistic to determine the inter-rater agreement between the reference standard and one senior grader from QUB grading centre**

|  | <b>Inter-rater agreement<br/>(reference standard vs a<br/>senior grader QUB), k<br/>(95% CI)<br/>(by eyes, n=106)</b> | <b>Inter-rater agreement<br/>(reference standard vs a<br/>senior grader QUB (by<br/>worst), k (95% CI)<br/>(by worst eye, n=53)</b> |
|--|---|---|
| <b>Overall Diabetic Retinopathy Grading:</b>   |   |   |
| <b>Any DR</b>  | 0.79 (0.67, 0.91)   | 0.74 (0.55, 0.92)   |
| <b>Treatable DR</b>  | 0.71 (0.48, 0.95)   | 0.68 (0.39, 0.97)   |
| <b>Referable Maculopathy</b>   | 0.75 (0.61, 0.90)   | 0.74 (0.55, 0.93)   |
| Abbreviations: CI=confidence interval, k=kappa, DR=Diabetic retinopathy, DMO=diabetic macular oedema<br>Any DR defined as R1, R2, R3s, R3a and U<br>Treatable DR defined as R3a<br>Referable DMO defined as M1 and U |   |   |

**Intra and inter-grader agreement**

To ensure there was good intra-grader reliability as a reference standard, a stratified random sample of images were regraded. There was approximately one month between the first and second attempts to reduce the possibility of bias caused by memory. Additionally, inter-grader agreement was calculated using kappa to ensure there was good grading agreement between the reference standard and one senior grader from the Ophthalmic Reading Centre at QUB, Belfast. Any disagreements were discussed with

1  
2  
3 a retinal specialist until consensus was reached. Overall, the intra-grader agreement and inter-grader  
4  
5 agreement ranged from substantial to almost perfect.  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only



**Table S6. Diagnostic test accuracy of DR graders in Vietnam against a reference standard from the UK, excluding ungradable images (Phase I)**

|   | <b>Level 1 graders<br/>(n=373 patient<br/>images)</b> | <b>Level 2 graders<br/>(n=373 patient<br/>images)*</b> | <b>Level 3 graders<br/>(n=235 patient<br/>images)†</b> |
|---|---|--|--|
| <b>Any DR</b>   |   |  |  |
| Sensitivity (%) (95% CI)  | 47.9 (38.8, 57.2)                                     | 50.8 (41.6, 60.0)                                      | 49.0 (38.7, 59.3)                                      |
| Specificity (%) (95% CI)  | 89.7 (85.1, 93.0)                                     | 98.8 (96.3, 99.7)                                      | 100 (96.8, 100)  |
| PPV (%) (95% CI)  | 69.0 (57.9, 78.4)                                     | 95.3 (86.2, 98.8)                                      | 100 (90.6, 100)  |
| NPV (%) (95% CI)  | 78.2 (72.9, 82.7)                                     | 80.9 (76.0, 85.0)                                      | 74.6 (67.8, 80.5)                                      |
| <b>Referable DR</b>   |   |  |  |
| Sensitivity (%) (95% CI)  | 38.1 (19.0, 61.3)                                     | 28.6 (12.2, 52.3)                                      | 22.2 (7.4, 48.1)                                       |
| Specificity (%) (95% CI)  | 98.9 (96.9, 99.6)                                     | 100 (98.7, 100)  | 99.5 (97.1, 99.9)                                      |
| PPV (%) (95% CI)  | 66.7 (35.4, 88.7)                                     | 100 (51.7, 100)  | 80.0 (29.9, 98.9)                                      |
| NPV (%) (95% CI)  | 96.4 (93.8, 97.9)                                     | 96.0 (93.3, 97.6)                                      | 94.0 (89.9, 96.6)                                      |
| <b>Referable DMO</b>  |   |  |  |
| Sensitivity (%) (95% CI)  | 9.3 (3.0, 23.1)                                       | 37.2 (23.4, 53.3)                                      | 26.5 (13.5, 44.7)                                      |
| Specificity (%) (95% CI)  | 99.1 (97.0, 99.8)                                     | 99.4 (97.5, 99.9)                                      | 100 (97.6, 100)  |
| PPV (%) (95% CI)  | 57.1 (20.2, 88.2)                                     | 88.9 (63.9, 98.1)                                      | 100 (62.9, 100)  |
| NPV (%) (95% CI)  | 88.9 (85.1, 92.0)                                     | 92.3 (88.9, 94.7)                                      | 88.6 (83.5, 92.4)                                      |
| <p><b>Abbreviations:</b> UK = United Kingdom, DR = Diabetic Retinopathy, DMO = Diabetic Macular Oedema, CI = Confidence Intervals,<br/>           Grading criteria: UK National Diabetic Eye Screening Programme (NDESP) classification system (See supplementary material, Table S1 for more details).<br/>           Any DR, is defined as grades R1, R2, R3s and R3a.<br/>           Referable DR is defined as grades R2 and R3a.<br/>           Referable DMO is defined as grades M1<br/>           Sensitivity is the ability of a test to correctly identify patients with a disease and specificity is the ability of a test to correctly identify people without the disease<br/>           Positive predictive value (PPV) is the proportion of those who test positive who have the condition (true positives) and negative predictive value (NPV) is the proportion of those who test negative who do not have the condition (true negatives).</p> |   |  |  |

**Table S7: Diagnostic test accuracy of DR graders in Vietnam against a reference standard from the UK after additional DR training was delivered, excluding ungradable images (Phase II)**

|   | Level 1 graders<br>(n=115 patient<br>images) | Level 2<br>graders (n=115<br>patient images) | Level 3 graders<br>(n=62 patient<br>images) |
|---|--|--|---|
| <b>Any DR</b>   |  |  |   |
| Sensitivity (%) (95% CI)  | 97.6 (85.6, 99.9)                            | 72.5 (55.9, 84.9)                            | 55.6 (38.1, 72.1)                           |
| Specificity (%) (95% CI)  | 95.6 (86.8, 99.8)                            | 100 (93.5, 100)                              | 100 (84.6, 100)                             |
| PPV (%) (95% CI)  | 93.0 (79.9, 98.2)                            | 100 (85.4, 100)                              | 100 (80.0, 100)                             |
| NPV (%) (95% CI)  | 98.5 (90.7, 99.9)                            | 85.5 (75.2, 92.2)                            | 57.9 (10.9, 73.2)                           |
| <b>Referable DR</b>   |  |  |   |
| Sensitivity (%) (95% CI)  | 88.9 (50.7, 99.4)                            | 55.6 (22.7, 84.7)                            | 77.8 (40.0, 97.2)                           |
| Specificity (%) (95% CI)  | 90.0 (81.9, 94.8)                            | 96.9 (90.5, 99.2)                            | 100 (92.8, 100)                             |
| PPV (%) (95% CI)  | 44.4 (22.4, 68.7)                            | 62.5 (25.9, 89.8)                            | 100 (56.1, 100)                             |
| NPV (%) (95% CI)  | 98.9 (93.4, 99.9)                            | 95.9 (89.2, 98.7)                            | 96.1 (87.8, 98.8)                           |
| <b>Referable DMO</b>  |  |  |   |
| Sensitivity (%) (95% CI)  | 90.0 (54.1, 99.5)                            | 60.0 (26.4, 86.3)                            | 80.0 (44.4, 97.5)                           |
| Specificity (%) (95% CI)  | 97.0 (91.8, 99.2)                            | 97.9 (91.9, 99.6)                            | 100 (92.6, 100)                             |
| PPV (%) (95% CI)  | 75.0 (42.8, 93.3)                            | 75.0 (35.6, 95.5)                            | 100 (59.8, 100)                             |
| NPV (%) (95% CI)  | 99.0 (93.6, 99.9)                            | 95.9 (89.2, 98.7)                            | 96.0 (87.4, 99.6)                           |
| <p><b>Abbreviations:</b> UK = United Kingdom, DR = Diabetic Retinopathy, DMO = Diabetic Macular Oedema, CI = Confidence Intervals,<br/>           Grading criteria: UK National Diabetic Eye Screening Programme (NDESP) classification system (See supplementary material, Table S1 for more details).<br/>           Criteria:<br/>           Any DR is defined as grades R1, R2, R3s, and R3a.<br/>           Referable DR is defined as grades R2 and R3a.<br/>           Referable DMO is defined as grades M1.<br/>           Sensitivity is the ability of a test to correctly identify patients with a disease and specificity is the ability of a test to correctly identify people without the disease<br/>           Positive predictive value (PPV) is the proportion of those who test positive who have the condition (true positives) and negative predictive value (NPV) is the proportion of those who test negative who do not have the condition (true negatives).</p> |  |  |   |

**Table S8: The prevalence of any diabetic retinopathy (DR), referable DR, any maculopathy and ungradable cases with the reference grader from Phase I and Phase II**

| <b>Diabetic Retinopathy grades</b>   | <b>Phase I</b> | <b>Phase II (post remedial training)</b> | <b>P-Value</b> |
|--|----------------|--|----------------|
| R0 (n,%)   | 257 (62.68)    | 68 (59.13)                               | P=0.347        |
| R1 (n,%)   | 100 (24.39)    | 32 (27.83)                               |                |
| R2 (n,%)   | 11 (2.68)      | 2 (1.74)                                 |                |
| R3a (n,%)  | 10 (2.44)      | 7 (6.09)                                 |                |
| R3s (n,%)  | 1 (0.24)       | 0 (0.00)                                 |                |
| U (n,%)  | 31 (7.56)      | 6 (5.22)                                 |                |
| <b>Any DR</b>  |                |  |                |
| - Yes (n,%)  | 153 (37.32)    | 47 (40.87)                               | P=0.488        |
| - No (n,%)   | 257 (62.68)    | 68(59.12)                                |                |
| <b>Referable DR</b>  |                |  |                |
| - Yes (n,%)  | 52 (12.68)     | 15 (13.04)                               | P=0.918        |
| - No (n,%)   | 358 (87.32)    | 100 (86.96)                              |                |
| <b>Any DMO</b>   |                |  |                |
| - M0 (n,%)   | 324 (79.02)    | 99 (86.09)                               | P=0.173        |
| - M1 (n,%)   | 43 (10.49)     | 10 (8.70)                                |                |
| - U (n,%)  | 43 (10.49)     | 6 (5.22)                                 |                |
| <b>Abbreviations:</b> DR=diabetic retinopathy, DMO=Diabetic Macular Oedema, U=ungradable, Chi-Squares used to test significance. |                |  |                |

| Section & Topic          | No  | Item   | Reported on page #          |
|--------------------------|-----|--|-----------------------------|
| <b>TITLE OR ABSTRACT</b> |     |  |                             |
|                          | 1   | Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)  | 1                           |
| <b>ABSTRACT</b>          |     |  |                             |
|                          | 2   | Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)                                 | 3-4                         |
| <b>INTRODUCTION</b>      |     |  |                             |
|                          | 3   | Scientific and clinical background, including the intended use and clinical role of the index test   | 5-6                         |
|                          | 4   | Study objectives and hypotheses  | 5                           |
| <b>METHODS</b>           |     |  |                             |
| <i>Study design</i>      | 5   | Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)     | 5                           |
| <i>Participants</i>      | 6   | Eligibility criteria   | 6-7                         |
|                          | 7   | On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)                 | 6-7                         |
|                          | 8   | Where and when potentially eligible participants were identified (setting, location and dates)   | 6-7                         |
|                          | 9   | Whether participants formed a consecutive, random or convenience series  | 7                           |
| <i>Test methods</i>      | 10a | Index test, in sufficient detail to allow replication  | 6-7                         |
|                          | 10b | Reference standard, in sufficient detail to allow replication  | 6                           |
|                          | 11  | Rationale for choosing the reference standard (if alternatives exist)  | 6                           |
|                          | 12a | Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory         |                             |
|                          | 12b | Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory |                             |
|                          | 13a | Whether clinical information and reference standard results were available to the performers/readers of the index test                                 | 6-7                         |
|                          | 13b | Whether clinical information and index test results were available to the assessors of the reference standard  | 6-7                         |
| <i>Analysis</i>          | 14  | Methods for estimating or comparing measures of diagnostic accuracy  | 12-14                       |
|                          | 15  | How indeterminate index test or reference standard results were handled  | 12-14<br>Supplementary file |
|                          | 16  | How missing data on the index test and reference standard were handled   | 12-14                       |
|                          | 17  | Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory  | 12-14                       |
|                          | 18  | Intended sample size and how it was determined   | NA                          |
| <b>RESULTS</b>           |     |  |                             |
| <i>Participants</i>      | 19  | Flow of participants, using a diagram  | 10-11                       |
|                          | 20  | Baseline demographic and clinical characteristics of participants  | 9                           |
|                          | 21a | Distribution of severity of disease in those with the target condition   | Supplementary file          |
|                          | 21b | Distribution of alternative diagnoses in those without the target condition  | Supplementary file          |
|                          | 22  | Time interval and any clinical interventions between index test and reference standard   | Supplementary file          |
| <i>Test results</i>      | 23  | Cross tabulation of the index test results (or their distribution) by the results of the reference standard  | 12-14 and<br>supplementary  |
|                          | 24  | Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)  | 12-14                       |
|                          | 25  | Any adverse events from performing the index test or the reference standard  | NA                          |
| <b>DISCUSSION</b>        |     |  |                             |
|                          | 26  | Study limitations, including sources of potential bias, statistical uncertainty, and generalisability  | 17                          |
|                          | 27  | Implications for practice, including the intended use and clinical role of the index test  | 17-18                       |
| <b>OTHER INFORMATION</b> |     |  |                             |
|                          | 28  | Registration number and name of registry   | NA                          |
|                          | 29  | Where the full study protocol can be accessed  | NA                          |

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

