



## Automated grading for diabetic retinopathy: a large-scale audit using arbitration by clinical experts

Alan D Fleming, Keith A Goatman, Sam Philip, et al.

*Br J Ophthalmol* 2010 94: 1606-1610 originally published online September 21, 2010

doi: 10.1136/bjo.2009.176784

---

Updated information and services can be found at:

<http://bjo.bmj.com/content/94/12/1606.full.html>

---

### References

*These include:*

This article cites 20 articles, 8 of which can be accessed free at:

<http://bjo.bmj.com/content/94/12/1606.full.html#ref-list-1>

### Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

---

### Notes

---

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://journals.bmj.com/cgi/ep>

# Automated grading for diabetic retinopathy: a large-scale audit using arbitration by clinical experts

Alan D Fleming,<sup>1</sup> Keith A Goatman,<sup>1</sup> Sam Philip,<sup>2</sup> Gordon J Prescott,<sup>3</sup> Peter F Sharp,<sup>1</sup> John A Olson<sup>2</sup>

<sup>1</sup>Biomedical Physics, University of Aberdeen, Foresterhill, UK  
<sup>2</sup>Diabetes Retinal Screening Service, Aberdeen, UK  
<sup>3</sup>Section of Population Health, University of Aberdeen, Foresterhill, UK

## Correspondence to

Dr John A Olson, Diabetes Retinal Screening Service, NHS Grampian, David Anderson Building, Foresterhill Road, Aberdeen AB25 2ZP, UK; [john.olson@nhs.net](mailto:john.olson@nhs.net)

Accepted 31 January 2010  
 Published Online First 21 September 2010

## ABSTRACT

**Background/aims** Automated grading software has the potential to reduce the manual grading workload within diabetic retinopathy screening programmes. This audit was undertaken at the request of Scotland's National Diabetic Retinopathy Screening Collaborative to assess whether the introduction of automated grading software into the national screening programme would be safe, robust and effective.

**Methods** Automated grading, performed by software for image quality assessment and for microaneurysm/dot haemorrhage detection, was carried out on 78 601 images, obtained from 33 535 consecutive patients, which had been manually graded at one of two regional diabetic retinopathy screening programmes. Cases where the automated grading software assessment indicated gradable images with no disease but the screening programme indicated ungradable images or disease more severe than mild retinopathy were arbitrated by seven senior ophthalmologists.

**Results** 100% (180/180) of patients with proliferative retinopathy, 100% (324/324) with referable background retinopathy, 100% (193/193) with observable background retinopathy, 97.3% (1099/1130) with referable maculopathy, 99.2% (384/387) with observable maculopathy and 99.8% (1824/1827) with ungradable images were detected by the software.

**Conclusion** The automated grading software operated to previously published results when applied to a large, unselected population attending two regional screening programmes. Manual grading workload reduction would be 36.3%.

## INTRODUCTION

Systematic screening for diabetic retinopathy using retinal photography has been shown to reduce the incidence of blindness among people with diabetes.<sup>1–4</sup> Diabetic retinopathy screening programmes are challenged by the rising prevalence of diabetes,<sup>5</sup> the costs of implementation,<sup>6, 7</sup> the maintenance of an effective quality assurance system<sup>8</sup> and the repetitious nature of grading. As a result, means for improving the efficiency of screening are being sought such as optimisation of the screening interval,<sup>9</sup> reducing the number of photographic fields used<sup>10</sup> and automation of image grading,<sup>11</sup> the topic of this paper.

The performance of automated grading software is often measured in terms of sensitivity and specificity for detecting retinopathy.<sup>12, 13</sup> However, absolute values for these may not be immediately useful when deciding whether an automated grading system would be effective at improving

screening programme efficiency. It is also important to know the relative sensitivities and costs between alternative grading systems, one of which is performed only by human graders while the other includes automated grading software. It is also important to know whether any missed cases have equivocal or unequivocal retinopathy.

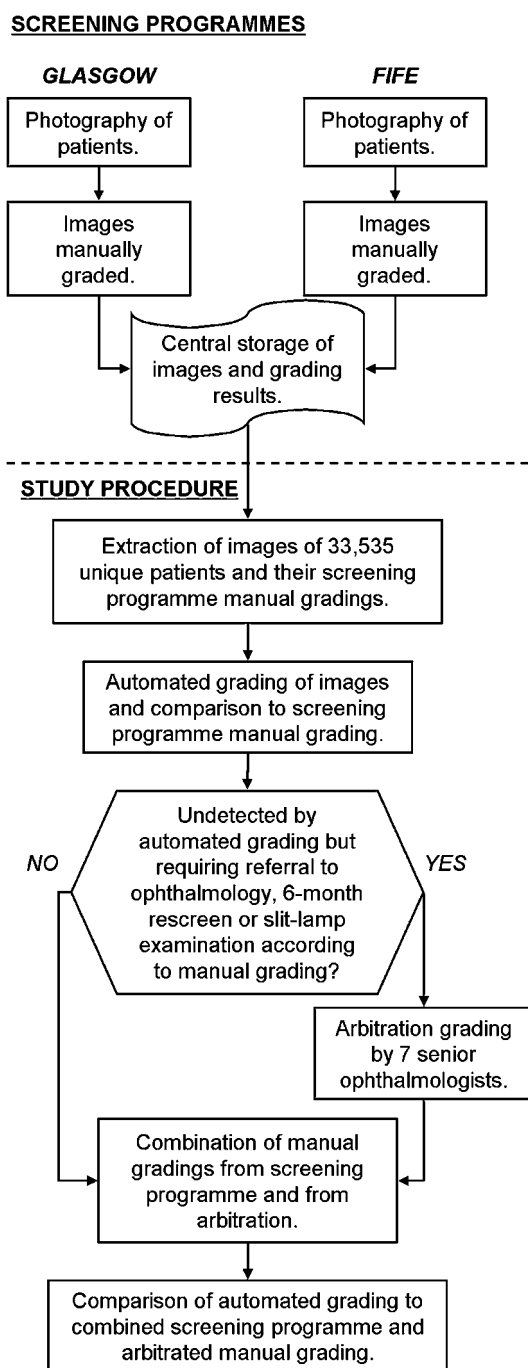
In a previous study, involving 6722 patients from Grampian, we compared automated and manual grading systems with a reference standard grading.<sup>7, 14</sup> The automated grading software had a better detection rate, 90.5% (2283/2523), than manual grading, 86.5% (2182/2523), for any retinopathy (mild or more severe retinopathy) and there was no significant difference between the detection rates, 97.9% (323/330) by automated grading and 99.1% (327/330) by manual grading, for observable/referable retinopathy (more severe than mild retinopathy). With 45.7% of cases not requiring manual grading, an annual saving of £200 000 would be made for the 160 000 people with diabetes in Scotland, estimated for the financial year 2005/2006. However, before the software could be implemented into the Scottish national diabetic retinopathy screening programme, it was necessary to demonstrate that its performance was maintained when used in other Scottish screening centres.

## MATERIALS AND METHODS

The study outline is shown in figure 1. Images were obtained from two Scottish screening centres, Glasgow and Fife, working in accordance with the requirements of National Health Quality Improvement Scotland. Caldicott Guardian approval was obtained for the study.

Photography and grading were performed by graders employed within the Scottish screening programmes according to the recommendations of Scotland's National Diabetic Retinopathy Screening Collaborative.<sup>15</sup>

A single 45° macula-centred photograph was taken of each eye, with additional photographs obtained as required, for example to obtain better image quality or a better view of pathology. Mydriasis was used where small pupils prevented the capture of adequate quality images; a process known as 'staged mydriasis'. Images were obtained with seven fundus cameras in Glasgow (Canon EOS D30 and EOS 20D digital cameras with Canon CR6-45NM and CR-DGi non-mydratiac retinal cameras) and three in Fife (Canon EOS 20D digital cameras with Canon CR-DGi non-mydratiac retinal cameras). Image sizes were 2336×3504 (10 950



**Figure 1** Diabetic retinopathy screening programme and study procedures.

images), 1696×2544 (25 378 images) and 1440×2160 (42 273 images) pixels.

Manual grading was performed by the screening centres following the Scottish grading scheme illustrated in table 1. This table defines the lesions present at each grade of retinopathy, explains the terms observable and referable maculopathy/retinopathy and illustrates the associated patient outcomes. The outcome for the majority patients is 12 month recall. There are three more urgent possible outcomes: referral to ophthalmology, 6 month recall or referral for slit-lamp examination.

There were 84 658 consecutive images and grading results from 34 785 anonymised patient screening episodes attending screening centres in Glasgow or Fife between 1 January 2007 and 31 January 2008. In the case of patients with repeat

photographs, only the earliest screening episode was used. This set contained 78 601 images from 33 535 patients.

### Automated grading software

All images were processed by the automated grading software which assessed image quality and disease using previously described algorithms,<sup>17 18</sup> summarised here. The locations of the optic disc and the fovea were determined to check that the image shows a macula-centred view and to determine whether the image is of a left or a right eye. Image clarity was assessed by checking that sufficient small vessels were visible in the macula region. If an adequate quality macula-centred image is present for each eye then disease is assessed in these images by performing microaneurysm/dot haemorrhage detection. The first stage in the detection of these was performed on the green plane of the image, after correction for uneven illumination, using a non-specific morphological filter that determines candidate lesions by separating dot-like dark objects from the linear vasculature.<sup>18</sup> The second stage performs more detailed analysis of the candidate lesions, measuring such features as their area, contrast and likelihood of lying on a vessel. An automated classifier, which had been previously trained using a set of 35 images containing 198 individually annotated microaneurysms/dot haemorrhages, separates true lesions from background objects based on the values of these features.

Artefacts caused by dirt on camera internal surfaces or by faulty photosensors can have a very similar appearance to microaneurysms/dot haemorrhages and hence cause frequent false-positive detections. Therefore, an automated artefact-masking algorithm was developed that noted the location of potential microaneurysms/dot haemorrhages in images taken with the same camera. Repeated detection of a potential microaneurysm/dot haemorrhage at the same location and in the same camera was judged to indicate the presence of an artefact.

The automated grading software produces a binary output: a patient is software positive if the software assessed quality to be inadequate or if it detected microaneurysms/dot haemorrhages. The automated grading software performance was evaluated by comparing this binary output against the screening programme manual grading results.

### Arbitration

The most important discrepancies between the binary output of the automated grading software and the screening programme were those involving patients who were not detected by the automated grading software but who the screening programme referred to ophthalmology, recalled at 6 months or referred for slit-lamp examination. Arbitration grading was performed on these discrepancies.

The arbitration grading was performed by seven senior ophthalmologists working at screening centres in Scotland. The discrepancies were mixed with 40 control images and presented to each grader in random order. As the control images had been graded by the arbitrators at least 6 months earlier, during a quality assurance test, an intragrader comparison was made.

If five or more arbitrators allocated a grade requiring referral to ophthalmology, recall at 6 months or slit-lamp examination, then this was defined as consensus for an outcome more urgent than the standard 12 month recall. In these cases, a consensus grade was allocated. This was necessary in order to evaluate the automated grading software performance for each grade. The grade assigned by the greatest number of arbitrators was used as the consensus grade. Where there was a choice between possible

**Table 1** The Scottish grading scheme<sup>16</sup>

Grade	Features	Location	Outcome
<b>Retinopathy</b>			
No visible retinopathy (R0, M0)	No visible lesions of diabetic retinopathy		
Mild background retinopathy (R1)	Microaneurysms/dot haemorrhages Hard exudates Cotton wool spots Blot haemorrhages Superficial/flame-shaped haemorrhages	Anywhere	Re-screen after 12 months
Observable background retinopathy (R2)	Four or more blot haemorrhages ( $\geq$ Airlie House standard photograph 2a)	In one hemi-field only	Re-screen after 6 months
Referable background retinopathy (R3)	Four or more blot haemorrhages ( $\geq$ Airlie House standard photograph 2a) Venous beading ( $\geq$ Airlie House standard photograph 6a) IRMA (Airlie House standard photograph 8a)	In both inferior and superior hemi-fields Anywhere	Refer to ophthalmology
Proliferative referable retinopathy (R4)	Active new vessels Vitreous haemorrhage	Anywhere	
<b>Maculopathy</b>			
Observable maculopathy (M1)	Hard exudates	$>1$ and $\leq 2$ disc diameters from the centre of the fovea	Re-screen after 6 months
Referable maculopathy (M2)	Blot haemorrhages Hard exudates	$\leq 1$ disc diameter from the centre of the fovea	Refer to ophthalmology
<b>Gradability</b>			
Ungradable (R6)	Retina not sufficiently visible for assessment		Refer for slit-lamp examination

Inferior and superior hemi-fields are delineated by a line passing through the centres of the fovea and the optic disc. Designations in the Scottish grading scheme are given (R0...R6, M0...M2). IRMA, intraretinal microvascular abnormalities.

grades, due to more than one grade being associated with the same maximum number of arbitrators, the most severe of these grades was chosen as the consensus grade.

If three or four arbitrators allocated a grade which would imply that either referral to ophthalmology, recall at 6 months or slit-lamp examination was required, then this was defined as no consensus. The patient was deemed to be indeterminate regarding whether or not the outcome was more urgent than the standard 12 month recall. In this case the consensus grade was indeterminate.

If two or fewer arbitrators allocated a grade which would imply that referral to ophthalmology, recall at 6 months or slit-lamp examination was required, then this was defined as consensus that the patient required 12 month recall. The consensus grade was mild retinopathy or no retinopathy.

### Statistical methods

The intergrader agreement between the assigned grades of the seven senior ophthalmologists was summarised using a  $\kappa$  statistic. A  $\kappa$  statistic was also used to assess intragrader agreement, comparing the earlier and later gradings of the 40 control images. The detection rates within different ethnic groups were compared using Pearson  $\chi^2$  tests or Fisher exact tests if required due to small numbers not being detected. This was done separately for those patients recalled at 12 months and those having a more urgent outcome. Stata version 10 (StataCorp, College Station, Texas, USA) was used for these analyses.

### RESULTS

The diabetic retinopathy screening centres referred 5.0% of patients to ophthalmology, recalled 1.8% at 6 months and referred 5.5% for slit-lamp examination.

Arbitration grading was performed on the images of 127 patients referred to ophthalmology, recalled at 6 months or referred for slit-lamp examination but who were not detected by the automated grading software. The  $\kappa$  statistic for intergrader agreement was  $\kappa=0.37$ .

Table 2 summarises the consensus grades resulting from the arbitration grading and shows the number of arbitrators who considered referral to ophthalmology, 6 month recall or slit-lamp examination was required for each case. In 63 cases the output of the automated grading software was appropriate and in 37 cases (31 with referable maculopathy, three with observable maculopathy and three ungradable) it was inappropriate according to the consensus of the arbitrators. Twenty-seven cases were indeterminate due to lack of consensus.

Considering the assigned grades, all seven arbitrators gave identical grades in only 20 cases (five cases of no retinopathy, one case of mild background retinopathy, one case of observable maculopathy, 12 cases of referable maculopathy and one case ungradable).

For the 40 control images, there was good intragrader agreement, with  $\kappa=0.77$ . Eighty-three per cent of the grades allocated to these images were identical to those allocated during a quality assurance test  $>6$  months earlier, 10.7% were higher than the previous ones and 6.3% were lower.

Table 3 displays the automated system results for each grade of retinopathy and for each possible patient outcome. The software was positive in 100% (193/193) of cases of proliferative retinopathy, 100% (324/324) of cases of referable background

**Table 2** Results of the arbitration grading. The second column groups the cases according to the definition used for consensus

Number of arbitrators, out of 7, who assigned a grade implying referral to ophthalmology, 6-month recall or slit-lamp examination (number of cases)	Consensus grade (number of cases)
7 arbitrators (19)	Referable maculopathy (31) Observable maculopathy (3) Ungradable (3)
6 arbitrators (8)	
5 arbitrators (10)	
4 arbitrators (11)	Indeterminate (27)
3 arbitrators (16)	
2 arbitrators (13)	Mild retinopathy (17) No retinopathy (46)
1 arbitrator (22)	
0 arbitrators (28)	

**Table 3** Proportions of patients, with CIs, for which the automated grading software output was positive

Grade	Software positive %	95% CI
No visible retinopathy	49.6 (10668/21503)	48.9 to 50.3
Mild background retinopathy	83.9 (6678/7964)	83.0 to 84.6
Observable maculopathy	99.2 (384/387)	97.8 to 99.7
Observable background retinopathy	100 (180/180)	97.9 to 100
Referable maculopathy	97.3 (1099/1130)	96.1 to 98.1
Referable background retinopathy	100 (324/324)	98.8 to 100
Proliferative referable retinopathy	100 (193/193)	98.1 to 100
Ungradable	99.8 (1824/1827)	99.5 to 99.9
<b>Outcome</b>		
12 month recall	58.9 (17346/29467)	58.3 to 59.4
6 month recall	99.5 (564/567)	98.5 to 99.8
Refer to ophthalmology	98.1 (1616/1647)	97.3 to 98.7
Slit-lamp examination	99.8 (1824/1827)	99.5 to 99.9

Twenty-seven cases have been omitted since they had an indeterminate grade.

retinopathy and 100% (180/180) of cases of observable background retinopathy. Other software-positive rates were 97.3% (1099/1130) for referable maculopathy, 99.2% (384/387) for observable maculopathy and 99.8% (1824/1827) for patients classified as ungradable. The error rate for cases requiring referral to ophthalmology, 6 month recall or slit-lamp examination was 0.11% (37/33535).

A total of 51.7% (17346/33535) of all patients required 12 month recall but were software positive. Out of the 33535 patients screened, 12185 would not require manual grading since they were assessed by the automated grading software as having adequate quality and no microaneurysms. Automated grading would therefore reduce the number of patients requiring manual grading by 36.3%.

Ethnicity was recorded for 12100 patients: 10068 (83%) were Caucasian, 1915 (16%) were Asian and 117 (1%) were Afro-Caribbean. The mean age of all patients was 62 years and 54.6% of all patients were male.

Table 4 displays the proportions which were software positive for the three ethnic categories for each patient outcome. No statistically significant differences were found between the ethnic groups in detection rates either for patients with non-referable retinopathy (Pearson  $\chi^2$  test  $p=0.38$ ) or for patients requiring referral to ophthalmology, 6 month recall or slit-lamp examination (Fisher exact test  $p=0.82$ ).

## DISCUSSION

This audit assessed the performance of an automated system based on microaneurysm/dot haemorrhage detection and image quality assessment operating on a large, unselected population of people with diabetes participating in Scotland's systematic diabetic retinopathy screening programme. Out of 33535 patients, all 697 patients with observable or referable retinopathy, other than maculopathy, were detected. In addition, the detection rates for observable and referable maculopathy and

technical failures are compatible with published standards and quality assurance protocols, and are very high compared with the sensitivity of retinal photography.<sup>10 19 20</sup> They are also higher than previously reported manual grading rates for detection of referable retinopathy, which were 96% by Olson *et al*<sup>21</sup> and a maximum of 85% by Abramoff *et al*<sup>22</sup>

The scale of the audit means that there are sufficient arbitration results to warrant detailed consideration. It has been shown that for cases where the automated grading software and the diabetic retinopathy screening programme results are conflicting, the grading software result was correct more frequently than incorrect; only 29% of the 127 arbitrated discrepancies had five or more arbitrators (defined as consensus) assigning an outcome requiring more urgent action than the standard 12 month recall. There was much disagreement between the senior ophthalmologists; full agreement occurred in only 16% of arbitrated discrepancies. The good concordance in the grading of the control images suggests that there was no tendency by the arbitrators to give the images a less or more severe grade than they would have done outside this evaluation.

The automated system performed equally well for images from the three ethnic categories tested: Caucasian, Asian and Afro-Caribbean.

While the design of our earlier study made possible a direct comparison between manual and automated grading, this audit did not look for cases of referable retinopathy/maculopathy that were missed by manual grading. It is therefore not known if there were patients who were disadvantaged by manual grading.

All cases that required referral to an ophthalmologist but were not detected by the automated grading software were of referable maculopathy. This may be explained by these cases having clear exudates but no obvious microaneurysms. However, in another study, also involving people attending the Scottish diabetic retinopathy screening programme, only 13.2% of patients with signs of referable maculopathy received laser treatment.<sup>23</sup> This implies that only a few of the cases of referable maculopathy not detected by the automated grading software would be considered for laser treatment and some of these may be detected on their next screening attendance.

It has been confirmed that the automated system achieves similar performance to that found in our previous study,<sup>14</sup> when applied to a larger multicentre data set. The current audit and our previous study achieved, respectively, detection rates of 87.0% and 87.9% for any retinopathy, 98.5% and 97.9% for observable retinopathy, and 99.8% and 99.5% for technical failures. While the earlier study indicated that 45.7% (3070/6722) of patients would be removed from the manual grading workload, in this audit 36.3% (12185/33535) would have been removed. This may be due to the presence of artefacts, which affected eight out of the 10 cameras used in the audit.

In this audit the automated grading software achieved 100% detection of proliferative, referable background and observable background retinopathy. It has been shown that the performance of an automated screening system for diabetic

**Table 4** Proportions of patients, with CIs, for which the automated grading software gave a positive result for each possible patient outcome, split by ethnicity

Outcome	Caucasian		Asian		Afro-Caribbean	
	Software positive %	95% CI	Software positive %	95% CI	Software positive %	95% CI
12 month recall	65.0 (5560/8555)	64.0 to 66.0	63.3 (989/1563)	60.9 to 65.6	66.3 (67/101)	56.7 to 74.8
6 month recall	99.6 (257/258)	97.8 to 99.9	100 (27/27)	87.5 to 100	100 (1/1)	20.7 to 100
Slit-lamp examination	100 (542/542)	99.3 to 100	99.3 (136/137)	96.0 to 99.9	100 (6/6)	61.0 to 100
Refer to ophthalmology	98.3 (695/707)	97.1 to 99.0	98.4 (183/186)	95.4 to 99.4	100 (9/9)	70.1 to 100

retinopathy is maintained at safe and effective levels over a large unselected set of images obtained from regional grading centres. Following this audit, Scotland's National Diabetic Retinopathy Screening Collaborative has concluded that automated grading should be introduced into Scotland's diabetic retinopathy screening programme.

**Acknowledgements** We thank Dr Alison Bow, Dr Brian Power, Dr Cynthia Santiago, Dr Anne Sinclair, Dr William Wykes, Dr Sonia Zachariah and Dr Usha Zamvar for voluntarily performing the arbitration grading. We thank staff of NHS Greater Glasgow & Clyde and NHS Fife retinal screening programmes, and their directors Dr William Wykes and Dr Caroline Styles, for their support. We thank Dr Nigel McLean of Scottish Health Innovations Limited and Scotland's National Caldicott Guardian, Dr Adam Bryson, for their support.

**Funding** Medalytix.

**Competing interests** Implementation in Scotland and elsewhere is being considered. If this occurs it is likely that there will be some remuneration for the University of Aberdeen, NHS Grampian, ADF, JAO and PFS. KAG, SP and GJP have no financial conflict of interest other than by association with the institutions mentioned above.

**Contributors** JAO was the principal investigator. JAO, PFS, KAG, ADF, and GJP contributed to the study design. ADF developed the automated methods, set up the analysis and generated the results. GJP performed statistical analysis. All participated in the interpretation of the data. ADF wrote the first draft of the paper. All authors reviewed and revised the paper for important intellectual content. JAO takes responsibility for the content.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## REFERENCES

- Hutchinson A, McIntosh A, Peters J, *et al*. Effectiveness of screening and monitoring tests for diabetic retinopathy—a systematic review. *Diabet Med* 2000;**17**:495–506.
- Arun CS, Al-Bermani A, Stannard K, *et al*. Long-term impact of retinal screening on significant diabetes-related visual impairment in the working age population. *Diabet Med* 2009;**26**:489–92.
- Stefansson E, Bek T, Porta M, *et al*. Screening and prevention of diabetic blindness. *Acta Ophthalmol* 2000;**78**:374–85.
- Zoega GM, Gunnarsdottir T, Bjornsdottir S, *et al*. Screening compliance and visual outcome in diabetes. *Acta Ophthalmol* 2005;**83**:687–90.
- Wild S, Roglic G, Green A, *et al*. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* 2004;**27**:1047–53.
- James M, Turner DA, Broadbent DM. Cost effectiveness analysis of screening for sight threatening diabetic eye disease. *Br Med J* 2000;**320**:1627–31.
- Scotland GS, McNamee P, Philip S, *et al*. Cost-effectiveness of implementing automated grading within the national screening programme for diabetic retinopathy in Scotland. *Br J Ophthalmol* 2007;**91**:1518–23.
- Garvican L, Scanlon PH. A pilot quality assurance scheme for diabetic retinopathy risk reduction programmes. *Diabet Med* 2004;**21**:1066–74.
- Olafsdottir E, Stefansson E. Biennial eye screening in patients with diabetes without retinopathy: 10-year experience. *Br J Ophthalmol* 2007;**91**:1599–601.
- Williams GA, Scott IU, Haller JA, *et al*. Single-field fundus photography for diabetic retinopathy screening—a report by the American Academy of Ophthalmology. *Ophthalmology* 2004;**111**:1055–62.
- Sharp PF, Olson JA, Strachan F, *et al*. The value of digital imaging in diabetic retinopathy. *Health Technol Assess* 2003;**7**:30.
- Larsen M, Gott J, Larsen N, *et al*. Automated detection of fundus photographic red lesions in diabetic retinopathy. *Invest Ophthalmol Vis Sci* 2003;**44**:761–6.
- Usher D, Dumskiy M, Himaga M, *et al*. Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. *Diabet Med* 2004;**21**:84–90.
- Philip S, Fleming AD, Goatman KA, *et al*. The efficacy of automated “disease/no disease” grading for diabetic retinopathy in a systematic screening programme. *Br J Ophthalmol* 2007;**91**:1512–7.
- Leese GP, Morris AD, Olson JA. A national retinal screening programme for diabetes in Scotland. *Diabet Med* 2003;**20**:962–4.
- Scottish Diabetic Retinopathy Screening Collaborative. Scottish diabetic retinopathy grading scheme. 2007 v1.1. <http://www.ndrs.scot.nhs.uk/ClinGrp/Docs/Grading%20Scheme%202007%20v1> (accessed Nov 2009).
- Fleming AD, Philip S, Goatman KA, *et al*. Automated assessment of retinal image quality based on clarity and field definition. *Invest Ophthalmol Vis Sci* 2006;**47**:1120–5.
- Fleming AD, Philip S, Goatman KA, *et al*. Automated microaneurysm detection using local contrast normalization and local vessel detection. *IEEE T Med Imaging* 2006;**25**:1223–32.
- British Diabetic Association. *Retinal photography screening for diabetic eye disease*. London: British Diabetic Association 1997:1–19.
- NHS Quality Improvement Scotland. *Diabetic retinopathy screening*. NHS Quality Improvement Scotland, 2004:1–46.
- Olson JA, Strachan FM, Hipwell JH, *et al*. A comparative evaluation of digital imaging, retinal photography and optometrist examination in screening for diabetic retinopathy. *Diabet Med* 2003;**20**:528–34.
- Abramoff MD, Niemeijer M, Suttorp-Schulten MSA, *et al*. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. *Diabetes Care* 2008;**31**:193–8.
- Schofield CJ, Ellis JD, Ellingford A, *et al*. Macular oedema is not predicted by periferoveal single blot haemorrhages. *Diabet Med* 2008;**25**:129–33. Epub 2008 Jan 4.133.