



# UNIVERSITY OF ABERDEEN

## **Should automated grading of retinal photographs be used in the Scottish Diabetic Retinopathy Screening Service?**

### **Origins**

I was invited to provide an independent review of the case for automated grading in the SDRSS. Published and unpublished documents were provided by Dr Rod Harvey, and Dr John Olson and colleagues. These were supplemented with other literature.

### **Recommendation**

My conclusion is that automated grading should replace level 1 manual grading of retinal screening photographs in Scotland.

Norman Waugh May 2009.

Professor Norman Waugh, MB ChB, DA, MRCP (UK), MPH, FRCP (Edin), FFPHM.

Group

Aberdeen Health Technology Assessments

Section of Population Health  
School of Medicine & Dentistry  
Polwarth Building  
Foresterhill  
Aberdeen AB25 2ZD  
[http://www.abdn.ac.uk/public\\_health](http://www.abdn.ac.uk/public_health)

## **Should automated grading of retinal photographs be used in the Scottish Diabetic Retinopathy Screening Service?**

### **Background.**

People with diabetes aged 12 years and over in Scotland are offered annual eye screening by digital photography. The photographs then go through a grading process.

Level 1 grading is done by less experienced graders, whose task is to identify anything which looks like retinopathy. They divide photographs into two groups: group 1 being those with no retinopathy at all, and group 2 those with either any retinopathy or suspected retinopathy, or considered of inadequate quality for grading. All those in group 2 are referred to level 2 grading by experienced graders. So level 1 is about dividing the photos into no disease, disease or ungradable.

The grading includes two aspects – retinopathy and maculopathy, with the codes and actions as shown in the table below. The macula is a small (5 mm diameter) area near the centre of the retina and contains the fovea, the most sensitive area of the retina. The fovea is responsible for central vision such as reading and writing. Damage to the macula can result in loss of central vision. This is usually immediately obvious, unlike loss of peripheral vision.

Maculopathy itself cannot be identified by digital photography (because the usual sign is oedema of the macula, and oedema is transparent), so the codes for that are based on surrogate markers - signs of retinopathy close to the macula.

Retinopathy	Description	Outcome
R0	No retinopathy anywhere	Routine re-screening at 12 month
R1	Mild background retinopathy	Rescreen at 12 months
R2	Background retinopathy requiring monitoring for progression	Rescreen at 6 months
R3	Background retinopathy sufficient to require referral	Refer to Ophthalmology, probably for surveillance rather than laser treatment
R4	Proliferative retinopathy	Refer to Ophthalmology, probably for laser treatment
R6	Inadequate photos	Arrange alternative screening.
Maculopathy		
M0	No features predictive of maculopathy	Rescreen 12 months
M1	Any hard exudates within 1 -2 disc diameters of the centre of the macula	Rescreen in 6 months
M2	Any hard exudates or blot haemorrhages within one disc radius of the centre of the macula	Refer to ophthalmology, probably for surveillance rather than laser treatment

Available from <http://www.ndrs.scot.nhs.uk/>

Note that eyes get both an R code and an M code.

So there are four possible outcomes after full grading;

- Eyes are deemed normal or with only mild background retinopathy, groups R0, R1 or M0. People in this group are rescreened at the normal interval of 12 months.
- Retinopathy or maculopathy is considered sufficiently serious for patients to be referred to Ophthalmology. This group is hereafter referred to as the referable group. This includes the M2 group, with possible maculopathy.
- An intermediate group, with either R2 retinopathy, or M1 maculopathy, where no urgent action is required, but patients are re-screened at 6 months instead of 12 – hereafter referred to as the observable group.
- Inadequate photos in which that the retina is not visible enough for assessment. Alternative screening is arranged for this group.

Software developed in Aberdeen can be used for grading of retinal photographs. The software and its algorithms are still being developed, but at present, the issue is whether automated grading could replace level 1 human grading, by reliably dividing photographs into entirely normal, or not.

### **The first Aberdeen study**

Two papers have been published from an Aberdeen study in which photographs from 6722 patients were sent for independent automated grading and manual human grading, and also for reference grading by highly experienced graders. Both types of grading were then compared with the reference grading, and hence with each other. One paper dealt with the clinical effectiveness and the other with cost-effectiveness. The main results were that;

- Automated grading was more sensitive than manual for the presence of any retinopathy (Philips et al 2007)
- Automated was slightly less sensitive than manual for the surrogate markers of possible macular oedema (i.e. the M1 and 2 groups).
- This led to automated grading having slightly poorer sensitivity for referable retinopathy – detection of 99.1% for manual versus 98.0% for automated, but the confidence intervals for these figures overlapped, so the difference could be due to chance. If we combine the observables and the referables, into a group called “actionables”, then the proportions detected are 99.1% and 97.9 % respectively, again not statistically different.
- Automated grading would save a large amount of money – about £200,000 for a region the size of Scotland at the costs and prevalence in the year 2005/6 (Scotland et al 2007). Savings would now be higher due to the increase in recorded prevalence. The costs in the paper were based on estimated prevalence of 160,000. By the end of 2008, there were about 220,000 people known to have diabetes in Scotland (Scottish Diabetes Survey 2008). The current saving would be about £275,000.

So from the study, we appear at first to have a trade-off – automated grading at level 1 would save a lot of money but at the cost of missing a few cases of macular oedema. In the study, with 6722 patients, seven with actionable retinopathy were missed by automated grading and three by manual grading – a difference of four patients, which was not statistically significant. None of the missed cases had sight-threatening disease.

However two aspects need consideration.

Firstly, the human graders were more experienced than the average, and their results might not be reproducible in routine screening.

Secondly, we need to remember that digital photography does not detect macular oedema, but only background retinopathy which is associated with and used to predict macular oedema – namely any blot haemorrhages or hard exudates with a radius of 1 or less disc diameter of the centre of the fovea. Most patients graded M2 do not have macular oedema. So the slightly lower sensitivity with automated grading is based on “referable” or suspected maculopathy of which only 12 to 14% actually have macular oedema requiring treatment.

When considering automated grading, sensitivity is more important than specificity. A false negative case might come to harm, if referable retinopathy was not referred. A false positive would come to no harm, because it would go forward to the second level of human grading. The only harm would be cost. But compared to an entirely manual system, automated grading would still provide large savings, because almost 40% of final grading would be done by the automated system.

### **The Glasgow and Fife study.**

As mentioned above, the graders in the Aberdeen study were more expert than is the average in routine screening. A much larger (unpublished) study compared the results of automated grading with manual

grading in 33535 patients from routine screening in Glasgow and Fife. Patients with photographs where the results of manual grading suggested action was required, but automated grading did not, were assessed independently by seven ophthalmologists, all classed as level 3 graders, with the final grade allocated only when five of the seven agreed.

When comparing the results of routine manual grading with automated grading amongst the 33535 patients, there were only 127 cases where the manual grading reported action was required, but the automated system did not. The results from the “gold standard” final grading showed that in most (71%) cases, action was not required. So in terms of sensitivity, automated grading missed 37 of 4041 actionable cases, a sensitivity of 99.1%. A few of these were technical failures where the photographs needed to be repeated, rather than having referable retinopathy. There were four cases needing only observation, with repeat photography at 6 months; they would have had it at the routine screening interval of 12-15 months instead, and are unlikely to have come to any harm.

This is an important point. Screening is repeated at regular intervals, so the main problem would be if it was missing rapidly developing retinopathy which could cause visual loss without warning within 12-15 months. But all of the referable cases missed had only M2 changes (i.e. no serious R grading), 90% of whom would be deemed not to have macular oedema after referral to ophthalmology. In the serious retinopathies, R3 and R4, sensitivity was 100%.

In terms of specificity, manual grading reported that 88% of cases need not go to level 2 grading, whereas automated grading ruled out only 59%. So as before, automated grading would result in more cases going on to second level grading than human grading.

So in summary, if we focus only on patients who might come to harm because they were not referred if automated grading was used at level 1, there might be about four patients (10% of the 37) at risk out of 33535 (0.012%). But not all people with macular oedema are at serious risk of visual loss. And some who are, might present with symptoms at a stage where treatment would preserve vision.

Other aspects to bear in mind are;

- patients in whom automated grading suggested action, but manual did not, were not studied, so there could be a different group which would be disadvantaged by manual grading.
- progression to visual loss is slow
- with gradual loss, there is the possibility of detection of changes suggestive of macular oedema at the next annual screening round.
- there are different mechanisms for macular oedema – focal leaks from microaneurysms, diffuse leaks, and hypoxia due to ischaemia. Only the first of these responds well to laser. The benefit of screening and early treatment is only seen in those with focal leaks, in whom deterioration is usually very gradual – a line of vision every 6-12 months. Rapid loss due to ischaemia cannot be prevented by photocoagulation.
- Many patients (about 50%?) cannot be treated because the lesions are under the fovea which cannot be lasered without causing visual loss.
- laser therapy is reserved until there is a reduction in vision perceived by the patient (partly because it can cause visual loss)

So even if there was a slightly lower sensitivity for referable M2 with automated grading, the difference in terms of visual loss would be much smaller, because most referables don't have macular oedema; treatment would not benefit all that do; the ones in which lasering is beneficial are the focal leaks with slow progression; some of whom would be detected by symptoms or at the next round of screening.

Though not all of those with referable retinopathy would survive – the risk of cardiovascular events (cardiovascular death, non-fatal MI, non-fatal stroke) is increased almost four-fold in those with sufficient retinopathy to require laser treatment (Targher et al 2008). (One question which therefore arises is whether the presence of referable retinopathy triggers measures to reduce cardiovascular risk.)

So if we consider the 220,000 patients with diabetes in Scotland, there will be about 3% with M2, which gives 6,600. The 3% is based on an approximate average between the Aberdeen prevalence of 2.7% and the Glasgow and Fife one of 3.4%.

If we assume that the observed difference in the Grampian study by Philips et al was true (remembering that the 95% CIs for manual and automated sensitivities overlapped – there was no statistically significant difference), then the sensitivity difference of 2.2% (99.1% minus 97.9%) would result in a difference of about 80 people being referred with M2, of whom about 8 would be treated, not all of whom would benefit. Some would not benefit because they would have diffuse leaks or ischaemia; others would have slowly progressive focal leaks which would have led to detection in a later screening round, and which might not have been treated at first referral anyway.

Nevertheless, it is possible that a few might suffer modest loss of vision (which would trigger symptoms and referral). The resulting disutility might be around 0.21 (as for maculopathy in the CORE model), and if, for example, we assumed that four patients might have had that loss avoided, and that the benefit lasted for five years, then we would usually be prepared to pay between £21,000 and £31,500 for the resulting QALYs (assuming affordable costs per QALY of £20,000 and £30,000 respectively). This is far less than the extra cost of £275,000 per annum.

### **The second Aberdeen study.**

Fleming et al (2009, in press) used a more sophisticated algorithm for automated grading. This used blot haemorrhages and exudates in addition to microaneurysms. In this study, in 633 people with M2, using the new algorithm increased the sensitivity for referable maculopathy from 93% to 96%. This still meant that there were 4% false negatives, but this was less than the 6% seen with manual grading.

### **Other considerations**

The real cost is of course the opportunity cost – if we save £275,000 a year by using automated grading for level 1, what can we achieve by alternative use of that money? It could be invested in improving control of diabetes, by insulin pumps, or DAFNE courses, or more DSN time, thereby preventing more retinopathy.

There is another assumption in these figures – that attendance is 100%, but we know that it is not. We also know that poor attenders for diabetes care have poorer outcomes, so we might expect those who do not attend to have more retinopathy than those who do.

Studies of trends over time usually show a reduction in the cumulative incidence of retinopathy, no doubt reflecting better control of glycaemia and blood pressure. In Danish people with type 1 diabetes, Hovind et al (2003) reported a decline in maculopathy over 20 years from 19% to 7%. Nordwall et al from Sweden reported a drop in severe retinopathy after 25 years of diabetes from 47% in a cohort with onsets of diabetes in 1961-65 to 24% in a cohort diagnosed in 1971-5. Backlund et al from Stockholm reported a 50% drop in blindness due to diabetes over a 15-year period.

However, not all centres have reported declines in blindness. Trautner and colleagues in Germany found that there had been no change.

Some of the improvement will be due to earlier detection and treatment, but much will be due to better control of glycaemia and blood pressure. In the DCCT, photocoagulation for macular oedema was required for 2.2% of those in the intensive arm versus 5.4% in the conventional arm. With longer follow up to 10 years beyond the end of DCCT, more patients developed clinically significant macular oedema but the frequency in the former intensively treated group was still under half that in the conventional group.

Vallance et al in Tayside noted that over the period 2001 to 2006, the number of patients with type 2 diabetes receiving laser treatment fell by about half. Since screening has been organised on a systematic basis in Tayside since the early 1990s, this cannot be explained by a “prevalence round” of newly detected retinopathy in the early 2000s. Vallance et al note improvements in control of hyperglycaemia (mean HbA1c fell from 7.9% in 2001 to 7.4% in 2006), blood pressure (mean SBP 142 in 2001 to 137 in 2006) and cholesterol (mean 5.0 mmol/l 2001 to 4.3 mmol/l in 2006). (These changes may reflect the use of the excellent computerised database in Tayside to alert doctors in the need for treatment intensification.)

Another study from Tayside (Rhatigan et al 1999) found that blindness in people with diabetes screened by the mobile camera unit, was low at 53 per 100,000. Only 26% of blindness was due to diabetes, the leading cause being age-related macular degeneration (46%), with most of the rest being due to glaucoma (22%). Of those with diabetic retinopathy treated with laser photocoagulation, 19% still went blind.

It is known from the quality assurance system in the DRS programme that the effectiveness of manual grading varies amongst operators. However at level 1 grading (which is what would be partially replaced by automated grading), it is very rare for referable retinopathy to be missed.

## **Conclusions**

At present, automated grading has very good sensitivity but poor specificity. The second tier manual grading provides the specificity, making the combination much more cost-effective than a purely manual system.

Automated grading could replace about 40% of first level manual grading and save around £275,000 a year.

Neither manual nor automated grading can detect all cases of retinopathy predictive of macular oedema. A few percent will be false negatives, but most such patients do not have macular oedema requiring laser treatment, and in some, laser would not be effective. Research is underway on new forms of screening (optical coherence tomography) to detect maculopathy. Meanwhile, patients should be educated to report any problems with central vision and so present as early as possible.

## **Recommendation**

Automated grading should replace level 1 manual grading in Scotland.

## **Research needs**

1. We need to develop better methods for screening for macular oedema. A review by Virgili et al concluded that comparing optical coherence tomography (OCT) looked promising, and a trial comparing OCT with referral based on existing criteria is already underway (HTA project 06/402/49: PI J Olson, Aberdeen).

2. The progression rate amongst those with the M1 level of presumed maculopathy is not known, but is presumed to be very low, but the OCT trial should produce some data. It should be possible to determine from the DRS data, what proportion of those with M1 progress to M2 and how long it takes.

3. The DRS data should also provide data on speed of progression of background retinopathy, which could feed into economic modelling of different screening intervals. For many patients, annual screening may be unnecessarily frequent.

4. A review of the usefulness of colour vision testing is underway (HTA 07/28/02: PI M Westwood, CRD).
5. Retinal arteriolar narrowing is associated with a higher risk of stroke, but the current software does not measure that. If it could be developed to do so, it might be possible to reduce cardiovascular mortality and morbidity by targeted interventions. However this assumes that those with retinal arteriolar narrowing are not already indentified as high risk by other means.
6. It would be good to see if all regions of Scotland have had declines in the incidence of severe retinopathy, and blindness due to it.
7. The quality of photographs is graded on a scale from 1 to 5, with 5 being classed as a technical failure. It would be worth checking on whether intermediate quality photographs cause more problems for automated or manual grading.

Norman Waugh, May 4th 2009.

#### References.

- Philips et al. The efficacy of automated “disease/no disease” grading for diabetic retinopathy in a systematic screening programme. *Br J Ophthalmology* 2007;91:1512-17
- Scotland et al. Cost-effectiveness of implementing automated grading within the national screening programme for diabetic retinopathy in Scotland. *Br J Ophthalmology* 2007;91:1518-23
- Scottish Diabetes Survey 2008. Unpublished (data being collated by Dept of PH, Univ of Aberdeen)
- Targher et al. Diabetic retinopathy is associated with an increased risk of cardiovascular events in type 2 diabetic patients. *Diabetic Med* 2008;25:45-50.
- Hovind et al. Decreasing incidence of severe diabetic microangiopathy in type 1 diabetes. *Diabetes care* 2003;26:1258-64.
- Nordwall et al. Declining incidence of severe retinopathy and persisting decrease of nephropathy in an unselected population of type 1 diabetes. *Diabetologia* 2004;47:1266-72
- Backlund et al. New blindness in diabetes reduced by more than one-third in Stockholm County. *Diabetic Med* 1997;14:732-740
- Trautner et al. Incidence of blindness in southern Germany between 1990 and 1998. *Diabetologia* 2001;44:147-150
- DCCT. Prolonged effect of intensive therapy on the risk of retinopathy complications in patients with type 1 diabetes mellitus. *Arch Ophthalmology* 2008;126:1707-15
- Vallance et al. Diabetic retinopathy: more patients, less laser. *Diabetes Care* 2008;31:126-1131.
- Rhatigan et al. Blindness in patients with diabetes who have been screened for eye disease. *Eye* 1999;13:166-9
- Virgili et al, reported in DARE. Original review 2007.