An interinstitutional comparative study and validation of computer aided drusen quantification

V Sivagnanavel, R T Smith, G B Lau, J Chan, C Donaldson and N V Chong

Updated information and services can be found at:
http://bjo.bmjjournals.com/cgi/content/full/89/5/554

These include:

References	This article cites 22 articles, 2 of which can be accessed free at: **http://bjo.bmjjournals.com/cgi/content/full/89/5/554#BIBL**
Rapid responses	You can respond to this article at: **http://bjo.bmjjournals.com/cgi/eletter-submit/89/5/554**
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 article
Topic collections	Articles on similar topics can be found in the following collections Vision Research (596 articles) Other ophthalmology (2112 articles)

Notes

To order reprints of this article go to:
http://www.bmjjournals.com/cgi/reprintform

To subscribe to *British Journal of Ophthalmology* go to:
http://www.bmjjournals.com/subscriptions/

SCIENTIFIC REPORT

An interinstitutional comparative study and validation of computer aided drusen quantification

V Sivagnanavel, R T Smith, G B Lau, J Chan, C Donaldson, N V Chong

....

Aims: To assess the portability and clinical applicability of a software program based on Photoshop (Adobe Systems Inc, San Jose, CA, USA) for digital drusen quantification.

Methods: Independent graders from the Digital Fundus Photo Reading Center of Columbia University and King's College Hospital used macular background levelling software to quantify the percentage of drusen in the central and middle Wisconsin subfields. 100 images of consecutive patients with choroidal neovascularisation in one eye and significant drusen in the other eye were analysed to determine suitability, and 10 were chosen for assessment by this software.

Results: Of the 10 images used in the interinstitutional validation, the random effects ANOVA for the central and middle subfields showed a high degree of interobserver agreement. The ICC for interobserver reliability was 0.83 (95% CI: 67 to 95) for the central subfield and 0.84 (95% CI: 69 to 99) for the middle subfield. Overall agreement with the manual grading results was good and the within patient coefficient of variation was about 20% for all the pairwise comparisons between observers and the manual stereo gradings. Of the 100 images used to assess practical applicability of the software, 79 were suitable for semiautomated analysis. 13 had extensive mixed retinal pigment epithelial (RPE) changes limiting drusen identification, five had a significant number of reticular drusen, which are poorly identified by the software, and three had multiple small areas of RPE atrophy, which are difficult to distinguish from drusen.

Conclusions: The software was successfully used by two institutions demonstrating portability, with good correlation between graders and to the manual stereo grading. Digital drusen quantification was possible in 79% of the images analysed.

A ge related macular degeneration (AMD) is the leading cause of blindness in the developed world.[1–7] A hallmark feature is the presence of drusen, and increases in drusen load have been correlated with advanced stages of AMD.[1–4] Several studies have attempted drusen reduction as a means of preventing visual loss.[5–10] Manual drusen quantification is laborious and costly.[11][12] Automated drusen quantification has value in furthering our understanding of the natural history of AMD and in trials of drusen reduction. We used a Photoshop (Photoshop 5.5, Adobe Systems Inc, San Jose, CA, USA) based semiautomatic drusen quantification software developed at Columbia University,[13] and evaluated it for interinstitutional portability and clinical applicability.

METHODS

Colour fundus images (Topcon TRC 50IX retinal camera) from 10 patients were selected at random from the digital database at Kings College Hospital (KCH) with stage 2 or 3 age related maculopathy (ARM) (as defined by the international grading system). Extensive hyperpigmentary or hypopigmentary abnormalities were excluded. Images were analysed by the methods described previously.[13] Briefly, the images used had minimum resolutions of 2700 pixels/inch. The images were saved as 24 bit RGB TIFF files, with 256 levels of intensity value for each colour channel. Images were then resized so that the distance from the centre of the macula to the temporal disc edge was 490 pixels, allowing uniformity of processing. The regions studied were the central 1000 μm diameter circular subfield and the 1000–3000 μm diameter annular subfield centred on the fovea, the central, and middle subfields defined by the Wisconsin grading template. Drusen area was measured as a percentage of the 1000 μm and 3000 μm subfield, and was unaffected by variable image size. The variation in brightness found in most fundus photographs was normalised using the red, green, and blue channels to create a standardised image in Photoshop, with nearly identical mean background colours, establishing a uniform basis for drusen segmentation. Contrast enhanced versions of the images (Photoshop/autolevels) were created for ease of visual recognition of drusen. Drusen analysis was carried out on the green channel of the standardised image using a digital template.[13]

After background levelling,[13] the optimum threshold level for drusen segmentation in the selected subfield is chosen by flicker comparison with the contrast enhanced image. For a given threshold, the drusen image is segmented such that pixels with brightness intensities above the threshold are coloured green, to label as drusen, and the rest darkened. Each such drusen image is superimposed on the contrast enhanced image. The optimised threshold is selected by visually inspecting the correspondence of the boundaries of the segmented drusen objects to those of the contrast enhanced objects. The threshold is then adjusted so that this visual fit is optimum in the aggregate as judged by the user (fig 1A–C). The total drusen area as a percentage of the selected subfield is then read directly (Photoshop/Histogram).

As part of the interinstitutional study, one expert and one non-expert grader from each institution (Eye Institute, Columbia University, USA and Kings College Hospital, London (KCH)) independently performed drusen quantification on the 10 images. A random effect ANOVA was used to assess the interobserver agreement in terms of the intraclass correlation coefficient (ICC). The interobserver and interinstitution effects were fitted in a random intercept (mixed)

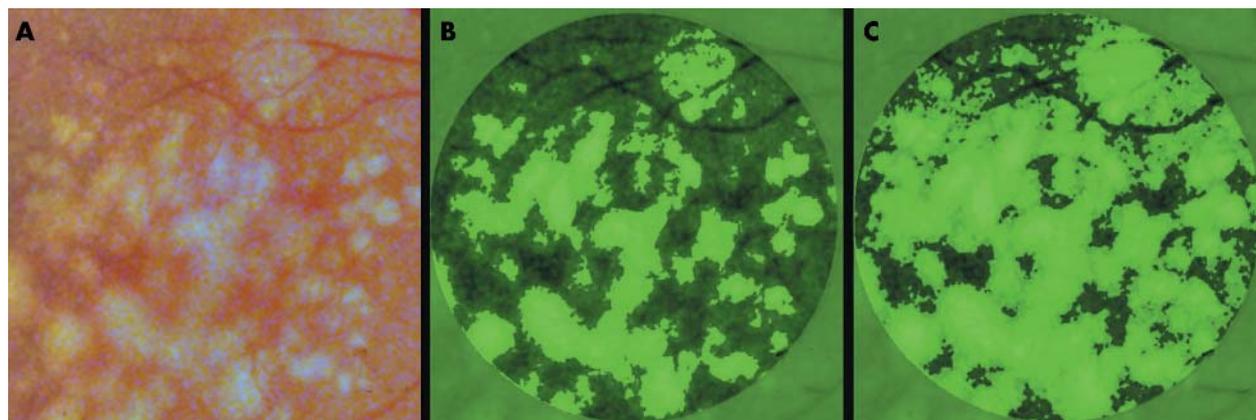...

Figure 1 Images illustrating drusen segmentation. Contrast enhanced layer for drusen identification (A), selection of best fit threshold (B), alternative threshold over representing drusen load (C).

linear model in order to determine if the two institutions were related to any measure disagreement. The automated measurements were also compared against the stereo manual grading and the difference was estimated using the method suggested by Bland and Altman.[14]

Secondly, as part of a pretrial assessment for a potential drusen reduction randomised controlled trial, 100 consecutive fluorescein angiograms taken at KCH between April 1999 and November 2002 were reviewed. Patients included had choroidal neovascularisation as a result of AMD in one eye and significant drusen in the fellow eye (defined as five large drusen or more than 20 small drusen in the macular area). Colour images of the fellow eye were analysed to determine whether they were suitable to be assessed by this software based on its previously determined limitations.[13]

RESULTS

Interinstitutional validation

The most labour intensive process in our method was in background levelling. For simple images this took about 1 minute and for more complicated images about 7 minutes. The total time taken for complete image evaluation and drusen segmentation varied from 4–10 minutes compared to 20–30 minutes per image with manual tracing.

There was good consensus between graders in the selection of the final threshold for drusen quantification. The random effects ANOVA showed a high degree of interobserver agreement as most of the variability was due to the interpatient variation ($F_{(9,30)} = 20$; $p = 0.00001$ and $F_{(9,30)} = 22$; $p = 0.00001$ for the central and middle subfields). Although the results were rather similar for the middle and central subfields, the middle subfield showed better agreement in general. The ICC for interobserver reliability was 0.83 (95% CI: 67 to 95) for the central subfield and 0.84 (95% CI: 69 to 99) for the middle subfield. The random effects linear mixed model confirmed good interobserver agreement (mean difference of 4.7; 95% CI: −7 to 17.6; $p = 0.44$ 3.6; 95% CI: −2.4 to 9.6; $p = 0.24$, for the central and middle subfields).and, in addition, it showed a non-significant disagreement between the two countries.

When the automated grading results were compared to the manual stereo grading results, we found that the automated measures tended to underestimate for large drusen values in both subfields. In addition, in the central subfield, the automated measures tended to overestimate for smaller drusen values. Optimum agreement with manual grading was obtained when the percentage of drusen in the measured area was 25%. Overall agreement with the manual grading results remained good and the within patient coefficient of

variation was about 20% for all the pairwise comparisons. Figure 2A shows the plot of the automated versus manual measurements for each observer for the middle subfield, with the line of equality for comparison. The Bland and Altman plots of the difference versus mean of the automated and manual measurements for each observer are presented in figure 2B. The estimates of the disagreements between
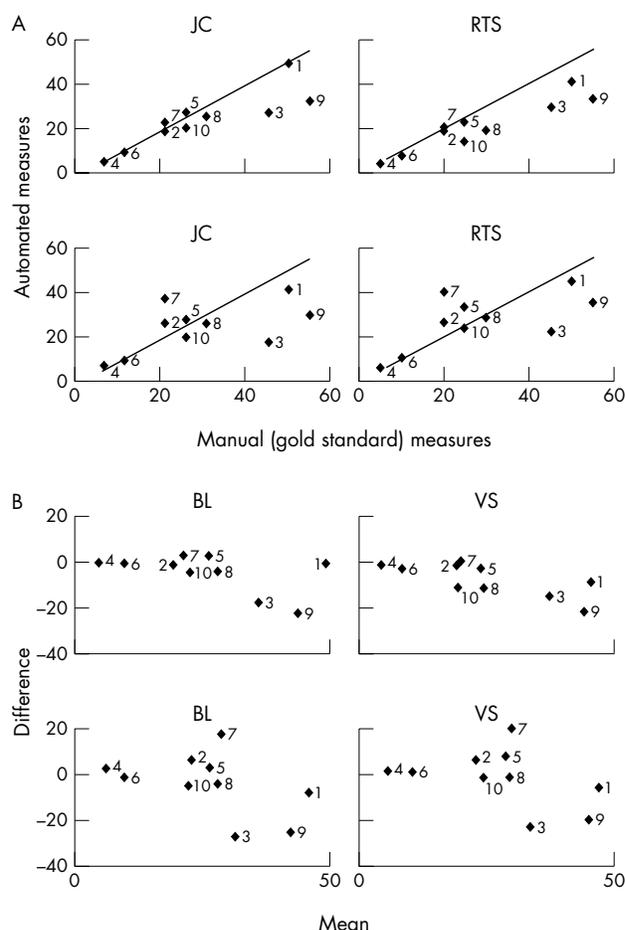

Figure 2 (A) Plots of the automated *v* manual measurements in relation to the line of equality for each observer (middle subfield). (B) Difference *v* mean of automated and manual measurements for each observer (middle subfield).

Table 1 Mean deviation for each observer from manual grading (graders JC and RTS from Columbia University, USA and graders VS and BL from Kings College Hospital, UK)

Observer	Central subfield			Middle subfield		
	Disagreement in relation to stereo grading (observer manual)	95% CI	p Value (two sided)	Disagreement in relation to stereo grading (observer manual)	95% CI	p Value (two sided)
JC	6.3	−4.7 to 17	0.23	−4.6	−10.7 to 1.5	0.12
RTS	5.2	−6.9 to 17	0.36	−7.5	−12.7 to −2.4	0.01
BL	8.5	−8 to 25	0.27	−4.1	−13.7 to 5.6	0.36
VS	6.8	(−9 to 23)	0.37	−1.3	−10.3 to 7.6	0.73

automated and manual gradings for each observer are shown in table 1, together with the test of whether the disagreement was significantly different from zero, either overestimating or underestimating the true drusen value. There was no significant deviation from the manual gradings for the central subfields for all four graders. Underestimation of drusen levels in the middle subfield reached significance for grader RTS.

Practical applicability
Seventy nine images were found to be suitable for analysis by the software. Of the 21 considered unsuitable, 13 had extensive mixed retinal pigment epithelial (RPE) changes limiting drusen identification (fig 3A). Five had a significant number of reticular drusen, which are poorly identified (fig 3B), and three had multiple small areas of RPE atrophy, which are difficult to distinguish from drusen (fig 3C). Significant thinning of RPE with baring of choroidal vessels can also make drusen recognition difficult (fig 3D).

CONCLUSION
Previous attempts at automated quantification have had limitations.[15–17] Shin et al described a method of computer assisted, interactive image processing which afforded higher accuracy.[18] They achieved an ICC of 0.92 and 0.93 for comparison of expert manual grading with automated supervised grading by two observers. However, major problems included identification of soft drusen with indistinct borders, large size drusen, and contrast confusion from darker blood vessels.

Comparison of the results of our semiautomated method with stereo manual grading and intraobserver reproducibility have been reported previously.[13] Good interobserver reproducibility has been demonstrated in the present study by graders from two institutions. Our digital method requires two supervised steps which are potential sources of interobserver variation—firstly, in background levelling for uniformity of drusen analysis and, secondly, in the selection of the threshold for drusen quantification. Levelling of the macular background is an approximation that may make a given section too bright or too dim. Consequently, drusen would be over-represented or under-represented. This was not a significant source of variation in this study. Disagreements between graders were predominantly the result of the subjective choice of final threshold selection. Large amounts of soft drusen with indistinct borders were more likely to be underestimated. Also, drusen underlying mixed RPE changes could potentially be excluded. Poor image quality and lack of stereo caused a tendency to include RPE atrophy as drusen. These confounding factors would have to be removed manually or by additional software and are a source of potential interinstituional and interobserver variation.

Although a semiautomated method requires greater time from the grader than a fully automatic system, it is an acceptable compromise for improved accuracy and reproducibility in relation to some published fully automated methods.[19] Rapantzikos et al have had greater success using a histogram based adaptive local thresholding.[20] However the limitations of confounding lesions has not been explored. Our semiautomated software has the potential to assess the change of drusen area in the majority of high risk patients with AMD. It has value in trials of drusen dynamics and reduction.
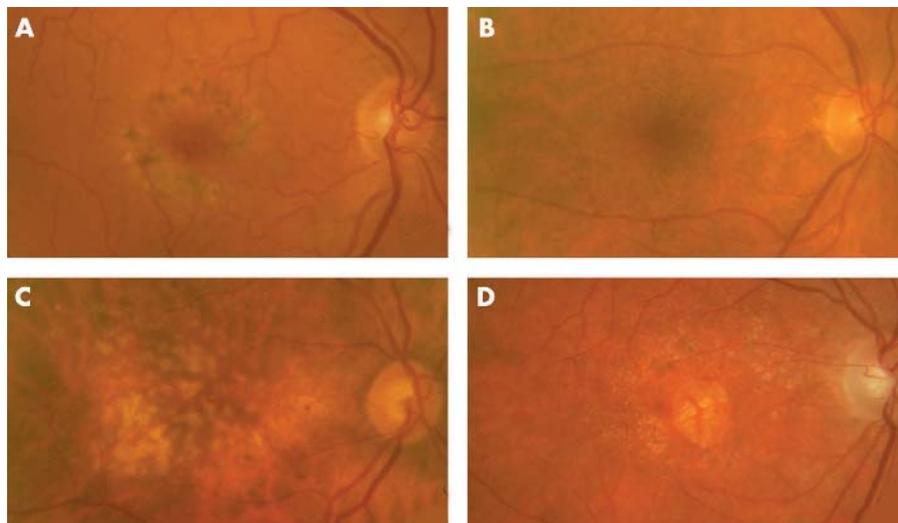

Figure 3 Images illustrating limitations of the practical application of software. Extensive RPE changes (A), reticular drusen (B), mixed drusen and RPE atrophy (C), baring of choroidal vessels (D).

....

Authors' affiliations

V Sivagnanavel, G B Lau, N V Chong, Retinal Research Unit, King's College Hospital, University of London, UK
R T Smith, J Chan, Department of Ophthalmology, Columbia University, New York, NY, USA
C Donaldson, Department of Biostatistics, Research and Development, Kings College Hospital, University of London, UK

Correspondence to: Miss V Sivagnanavel, Retinal Research Unit, King's College Hospital, University of London, UK; vasuki_siva1@yahoo.co.uk

REFERENCES

1 **Bressler SB**, Maguire MG, Bressler NM, *et al.* Relationship of drusen and abnormalities of the retinal pigment epithelium to the prognosis of neovascular macular degeneration. The Macular Photocoagulation Study Group. *Arch Ophthalmol* 1990;**108**:1442–7.
2 **Bressler NM**, Bressler SB, Seddon JM, *et al.* Drusen characteristics in patients with exudative versus non-exudative age-related macular degeneration. *Retina* 1988;**8**:109–14.
3 **Holz FG**, Wolfensberger TJ, Piguet B, *et al.* Bilateral macular drusen in age-related macular degeneration. Prognosis and risk factors. *Ophthalmology* 1994;**101**:1522–8.
4 **Klein R**, Klein BEK, Jensen SC, *et al.* The five-year incidence and progression of age related maculopathy. The Beaver Dam Eye Study. *Ophthalmology* 1997;**104**:7–21.
5 **Abdelsalam A**, Del Priore L, Zarbin MA. Drusen in age-related macular degeneration: pathogenesis, natural course, and laser photocoagulation-induced regression. *Surv Ophthalmol* 1999;**44**:1–29.
6 **Frennesson C**, Nilsson SEG. Prophylactic laser treatment in age related maculopathy reduced the incidence of exudative complications. *Br J Ophthalmol* 1998;**82**:1169–74.
7 **Figueroa MS**, Regueras A, Bertrand J, *et al.* Laser photocoagulation for macular soft drusen. Updated results. *Retina* 1997;**17**:378–84.
8 **The Choroidal Neovascularization Prevention Trial Research Group**. Laser treatment in eyes with large drusen. Short-term effects seen in a pilot randomized clinical trial. *Ophthalmology* 1998;**105**:11–23.
9 **The Choroidal Neovascularization Prevention Trial Research Group**. Laser treatment in fellow eyes with large drusen: updated findings from a pilot randomized clinical trial. *Ophthalmology* 2003;**110**:971–8.
10 **Scorolli L**, Corazza D, Morara M, *et al.* Argon laser vs subthreshold infrared (810-nm) diode macular grid photocoagulation in nonexudative age related macular degeneration. *Can J Ophthalmol* 2003;**38**:489–95.
11 **Klein R**, Davis MD, Magli YL, *et al.* The Wisconsin age-related maculopathy grading system. *Ophthalmology* 1991;**98**:1128–34.
12 **Bird AC**, Bressler NM, Bressler SB, *et al.* An international classification and grading system for age-related maculopathy and age-related macular degeneration. *Surv Ophthalmol* 1995;**39**:367–74.
13 **Smith RT**, Nagasaki T, Sparrow JR, *et al.* A method of drusen measurement based on the geometry of fundus reflectance. *Biomedical Engineering Online* 2003;**2**:10.
14 **Bland**, Altman D. (1086)., Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307–10.
15 **Kirkpatrick JNP**, Spencer T, Manivannan A, *et al.* Quantitative image analysis of macular drusen from fundus photographs and scanning laser ophthalmoscope images. *Eye* 1995;**9**:48–55.
16 **Berger JW**. Quantitative, spatio-temporal image analysis of fundus features in age related macular degeneration. *Proc SPIE* 1998;**3246**:48–53.
17 **Peli E**, Lahav M. Drusen measurement from fundus photographs using computer image analysis. *Ophthalmology* 1986;**93**:1575–80.
18 **Shin DS**, Javornik NB, Berger JW. Computer assisted, interactive fundus image processing for macular drusen quantification. *Ophthalmology* 1999;**106**:1119–25.
19 **Morgan WH**, Cooper RL, Constable IJ, *et al.* Automated extraction and quantification of macular drusen from fundal photographs. *N Z J Ophthalmol* 1994;**22**:7–12.
20 **Rapantzikos K**, *et al.* Detection and segmentation of drusen deposits on human retina: potential in the diagnosis of age-related macular degeneration. *Medical Image Analysis* 2003;**7**:95–108.