

# Effect of video display on the grading of diabetic retinopathy

MTJ Costen<sup>1</sup>, RSB Newsom<sup>1</sup>, B Parkin<sup>1</sup>, CS Marsh<sup>1</sup>,  
RL Mehta<sup>2</sup>, AJ Luff<sup>1</sup> and CR Canning<sup>1</sup>

CLINICAL STUDY

## Abstract

**Purpose** Digital images are increasingly being used in ophthalmology. These may be viewed either on thin-film transistor (TFT) or on cathode-ray tube (CRT) displays. However, there is little data showing which is superior. In this study, we compared the performance of CRT and TFT displays for grading of both compressed and uncompressed images of diabetic retinopathy.

**Methods** A total of 49 35 mm transparencies of diabetic retinopathy were scanned and compressed. The images comprised 17 with no retinopathy, eight with background, five with preproliferative, and 19 with proliferative retinopathy. Four levels of compression were used: 0, 70, 80, and 90%. A total of 196 randomised images were presented to two masked graders using both TFT and CRT displays under uniform lighting conditions, 2 months apart. The grade of retinopathy was assessed. Statistical analysis of grading accuracy was performed using receiver operator characteristic curves of sensitivity and specificity and the Stuart–Maxwell test for paired, nonparametric data.

**Results** Both displays showed high sensitivity and specificity for the detection of any retinopathy. For the specific grade of retinopathy, the CRT performed slightly better with a sensitivity of 0.80 for uncompressed images, compared with 0.75 using the TFT. Compression reduced these sensitivities to 0.73 on the CRT and 0.63 on the TFT. Grading of uncompressed images magnified to four times their original size was more accurate on the TFT.

**Conclusions** Grading on both displays met sensitivity and specificity criteria proposed by Diabetes UK (formerly British Diabetic Association) for screening of diabetic retinopathy. The CRT generally performed slightly better than the TFT in relation to the detection of the specific grade of retinopathy.

*Eye* (2004) 18, 169–174. doi:10.1038/sj.eye.6700595

**Keywords:** diabetic retinopathy screening; digital image compression; JPEG; thin-film transistor (TFT); cathode-ray tube (CRT)

## Introduction

Digital imaging systems are increasingly being used in place of conventional photography in the screening and grading of retinopathy. Advances in technology mean that the resolution of digital colour fundus cameras approaches that of 35 mm film.<sup>1</sup> In imaging diabetic retinopathy, such systems have been shown to have sensitivities and specificities approaching that of conventional film or polaroid,<sup>2–5</sup> and long-term running costs are less.<sup>2</sup> Image acquisition is faster with digital systems and other advantages include instant image feedback and lower flash intensities.<sup>5</sup> Detailed retinal images require large amounts of memory and are thus inefficient to transfer and store. Compression of such images into smaller file sizes can alleviate this problem.

There are two main video display types in current usage, thin-film transistor (TFT) and cathode-ray tube (CRT). TFT displays are smaller, they have a lower power consumption, and have a brighter display than their CRT counterparts. Such displays are growing in popularity in the hospital setting. These advantages, together with their portability, mean that they would also be the obvious choice in the mobile retinal screening environment.

There is currently little data on the effect of computer display type on the quality of retinal images. Given these facts, we felt that it was important to compare these two screen types in the grading of diabetic retinopathy. Accurate screening for retinopathy according to Diabetes UK guidelines depends on the ability to detect

<sup>1</sup>Southampton Eye Unit  
Southampton General  
Hospital, Tremona Road  
Southampton, UK

<sup>2</sup>Health Care Research Unit  
Southampton General  
Hospital, Southampton, UK

Correspondence:  
MTJ Costen  
Eye Unit Southampton  
General Hospital, Tremona  
Road, Southampton  
SO16 6YD, UK  
Tel: +44 2380 7988 36  
E-mail: mtcosten@  
doctors.org.uk

Received: 12 August 2002  
Accepted in revised form:  
21 March 2003

'any retinopathy' and 'sight-threatening' retinopathy.<sup>6</sup> In order to investigate the effect of display type on image quality more closely in this study, both compressed and uncompressed digital images of diabetic retinopathy were graded as to the specific level of retinopathy using both TFT and CRT displays.

## Methods

A total of 49 high-quality 35 mm transparencies (17 with no retinopathy, eight with background, five with preproliferative, and 19 with proliferative retinopathy) were selected in order to give a range of stages of retinopathy. Images were chosen mainly to determine the effect of screen type and image compression on the threshold for the detection of sight-threatening retinopathy. Sight-threatening retinopathy in this study was defined as the presence of neovascularisation, or features consistent with 'preproliferative' retinopathy.<sup>7</sup> The images were digitised to tagged information file format (TIFF) files using the Umax 'powerlook' 3000 scanner at a resolution of 3000 dots per inch (dpi) as described in a previous study.<sup>8</sup>

Comparable grading performance for both 35 mm transparencies of diabetic retinopathy and their scanned counterparts has been reported previously.<sup>9</sup> The images were converted into JPEG files at 0, 70, 80 and 90% compression using 'Paintshop Pro'. These levels of compression were chosen so as to demonstrate the change in image quality across a range of image file sizes.<sup>8</sup>

The original 35 mm transparency was projected and graded by three retinal specialists to produce a 'gold standard' for each image. The retinal specialists were not involved in the study grading.

## Image grading

All the images at their various compression levels ( $n = 196$ ) were randomised. Each image was presented independently to two masked specialist registrar observers under uniform lighting conditions, firstly on a 14 in, 32 bit true colour, superextended graphics array (SVGA) TFT screen (1024 × 768 pixels, 0.28 mm dot pitch). In an attempt to prevent any memory effect in the grading, the same images were presented using a 17.5 in, 32 bit true colour CRT display (1024 × 768 pixels, 0.26 dot pitch) after a 2-month 'washout period'. To investigate the effect of image magnification on grading performance, randomised uncompressed images ( $n = 49$ ) were also graded on each screen type at separate grading sessions.

Both uncompressed and compressed images were graded as 'no retinopathy', 'background', 'preproliferative' and 'proliferative' retinopathy, based on

European Grading Protocols.<sup>7</sup> Where there was disagreement between the graders, the images were regraded by a third grader.

The grade for each image at each compression level was compared to the gold standard for both the TFT and CRT images. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated both in relation to the specific grade of retinopathy and for the detection of 'any retinopathy' and 'sight-threatening retinopathy'.

## Statistical methods

Receiver operator characteristic curves (ROC) were calculated to demonstrate the association between sensitivity and specificity for the detection of both 'sight-threatening' and 'any retinopathy' on both TFT and CRT displays. The ROC curves give a measure of the accuracy of the grading system. The accuracy is represented by the area under the curve (AUC). Perfect agreement with the gold standard would be indicated by  $AUC = 1$ . The closer the curve converges towards the top left-hand corner, the more accurate the grading.

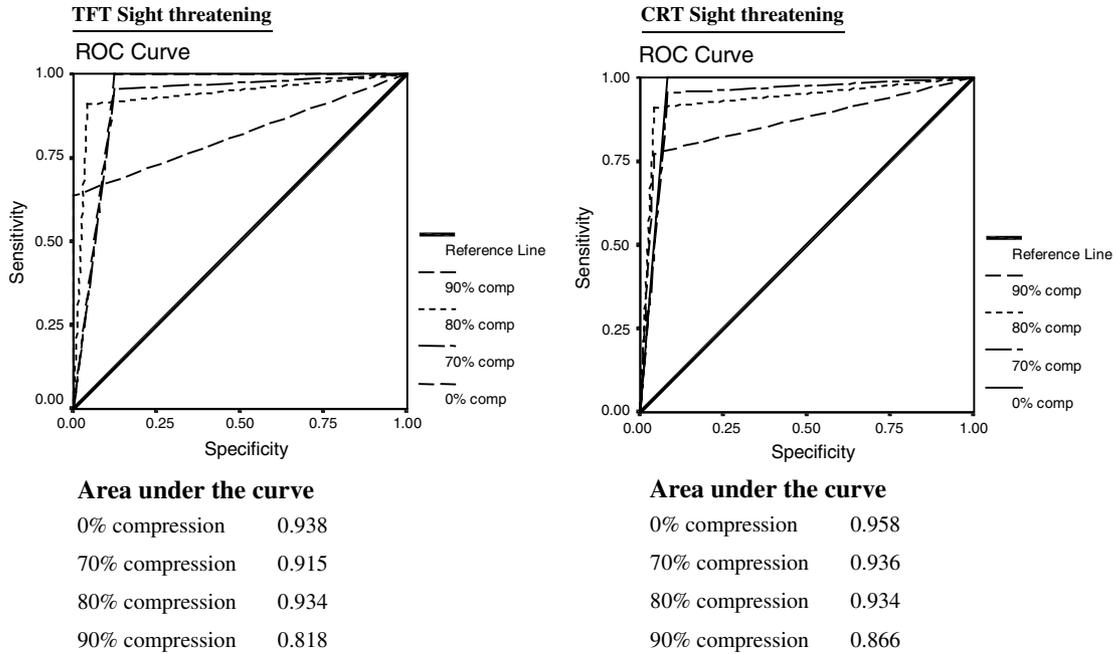
In considering the sensitivity and specificity for detection of the actual grade of retinopathy on both TFT and CRT displays, the Stuart–Maxwell test was used. This is a modification of the McNemar's test based on a  $2 \times 2$  contingency table. The Stuart–Maxwell test is suitable for nonparametric, paired, ordered nominal data where there are more than two ordered categories<sup>10,11</sup> (ie grade of retinopathy).

The number of true and test positives and negatives was analysed in the form of a  $4 \times 4$  table for each of the screen types. The null hypothesis of 'no difference between the proportions within each group' was tested for. With a  $P$  value of  $< 0.05$ , the null hypothesis was rejected assuming a difference between the proportions greater than would be expected by chance. Such a difference between the proportions would imply inaccurate grading, with more misclassifications, for that screen type.

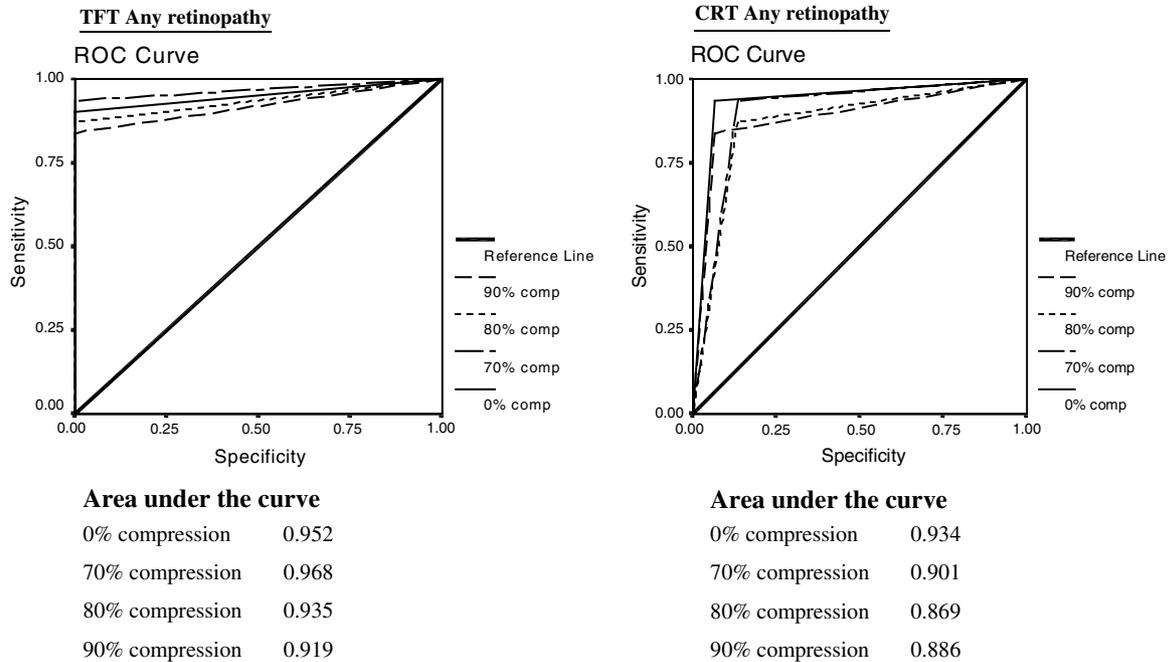
Both the AUC data and the  $P$ -statistic derived from the Stuart–Maxwell test yield information only for the screen type in question and do not enable statistical inference to differences between the two screen types. Apparent differences may suggest trends in performance and are discussed as such. Statistical analysis was performed using SAS for Windows™ (v8) and SPSS for Windows™ (v11).

## Results

The retinopathy grading results using both TFT and CRT displays are summarised in Table 1. In relation to the



**Figure 1** ROC curves for 'sight-threatening retinopathy'. The broken lines represent ROC curves for each level of image compression. 'Comp' = level of image compression.



**Figure 2** ROC curves for 'any retinopathy'.

detection of sight-threatening retinopathy, sensitivity and specificity were slightly greater for CRT when compared to TFT with images from 0 to 80% compression. The ROC curves for 'sight-threatening retinopathy' (Figure 1) demonstrate that the AUC is greater for the CRT at all compression levels, implying greater grading accuracy.

For the detection of any retinopathy, greater specificity was obtained using the TFT with 0.94 at 80% compression, compared to 0.81 at 80% for the CRT. This is reflected in the ROC curves for 'any retinopathy' (Figure 2). The AUC for the TFT at 80% compression is 0.935 compared to 0.869 for the same images viewed on

CRT. However, with uncompressed images and images compressed to 90%, sensitivity was slightly greater on the CRT. At other compression levels, the sensitivities were the same.

In grading the specific level of retinopathy ('grade of retinopathy'), the sensitivity using both displays is reduced, but remains slightly greater on the CRT even with highly compressed images (Table 1). For the images at 80% compression, sensitivity was 0.73 on the CRT and 0.63 on the TFT. Grading of images compressed to 70% or more of their original size yielded greater specificity on TFT with a specificity of 0.88 at 80% compression on the TFT and

0.76 for the same level of compression on the CRT.

The results of the Stuart–Maxwell test to investigate any association between the proportions are displayed in Table 1. It can be seen that for the CRT, the *P*-value remains >0.05 for all but the 90% images.

For the TFT, *P*-values for all images are less than their CRT counterparts suggesting greater grading inconsistency with the TFT.

When uncompressed images were magnified to ×4 their original size, greater sensitivity and specificity were achieved using the TFT rather than the CRT (Table 1). This is also reflected in the *P*-values for the ×4 data.

**Table 1** Sensitivity and specificity data for both TFT and CRT displays

	TFT, % compression				CRT, % compression			
	0%	70%	80%	90%	0%	70%	80%	90%
<b>Sight-threatening retinopathy</b>								
Sensitivity	1.0	0.96	0.91	0.65	1.0	0.96	0.91	0.78
Specificity	0.88	0.88	0.92	1.0	0.92	0.92	0.92	0.96
PPV	0.88	0.88	0.91	1.0	0.92	0.92	0.91	0.95
NPV	0.88	0.88	0.92	1.0	0.92	0.92	0.92	0.96
<b>Any Retinopathy</b>								
Sensitivity	0.91	0.94	0.87	0.84	0.94	0.94	0.87	0.87
Specificity	1.0	1.0	0.94	1.0	0.94	0.88	0.81	0.83
PPV	1.0	1.0	0.96	1.0	0.97	0.94	0.9	0.9
NPV	1.0	1.0	0.94	1.0	0.94	0.88	0.81	0.83
<b>Grade of retinopathy</b>								
Sensitivity	0.75	0.63	0.63	0.47	0.8	0.73	0.73	0.56
Specificity	0.81	0.81	0.88	1.0	0.84	0.79	0.76	0.88
PPV	0.84	0.83	0.9	1.0	0.89	0.85	0.85	0.9
NPV	0.81	0.81	0.88	1.0	0.84	0.79	0.76	0.88
Stuart Maxwell test <i>P</i>	0.0296	0.168	0.049	<0.0001	0.14	0.526	0.512	0.0074
<b>Sight-threatening retinopathy (X4 magnification)</b>								
Sensitivity					0.95			
Specificity					0.88			
PPV					0.87			
NPV					0.88			
<b>Any retinopathy (X4 magnification)</b>								
Sensitivity					0.93			
Specificity					0.94			
PPV					0.97			
NPV					0.94			
<b>Grade of retinopathy (X4 magnification)</b>								
Sensitivity					0.92			
Specificity					0.70			
PPV					0.76			
NPV					0.70			
Stuart–Maxwell test <i>P</i>					0.6895			
					0.13			

PPV: positive predictive value; NPV: negative predictive value. Results of the Stuart–Maxwell test are displayed for 'Grade of retinopathy' and the ×4 images.

## Discussion

In this study, we aimed to investigate the effect of video display properties on the grading of diabetic retinopathy. For screening purposes, both display types met the sensitivity (0.80) and specificity (0.95) requirements of the Diabetes UK guidelines for the detection of retinopathy.<sup>6</sup> For the 90% compressed images viewed on both TFT and CRT, the Diabetes UK requirements for detection of sight-threatening retinopathy would not be met. For such highly compressed images, the image quality was judged to be very poor and unsuitable for any grading purposes.

Sensitivity in detecting both 'any retinopathy' and 'sight-threatening' retinopathy was comparable for both TFT and CRT displays. There was a trend towards slightly greater sensitivity using the CRT. When the specific grade of retinopathy was assessed, sensitivity was greater using the CRT for all but the magnified images. Applying the Stuart–Maxwell test, lower *P*-values for the TFT grading data would suggest that the grading using CRT is more consistent when considering the actual grade of retinopathy in question. Magnification of the images to  $\times 4$  their original size resulted in a trend towards better grading performance on the TFT rather than the CRT.

When considering these apparent differences in performance, there are a number of technical attributes related to screen design, which may help to explain these trends.

The CRT display typically has a higher 'dot pitch' (0.26 *vs* 0.28 for the TFT) hence greater image resolution. This may enable greater accuracy of grading, by being able to resolve fine image details.

TFT displays are prone to artefacts such as ghosting and latent images,<sup>12</sup> which can affect image quality. CRT displays however are prone to 'veiling glare' from the glass front, which can also reduce image quality.<sup>13</sup> TFT displays tend to have a higher luminance (typically about 178 *vs* 118 cd/m<sup>2</sup> for CRT),<sup>12</sup> which may result in lower contrast-detection thresholds.<sup>14</sup> Indeed, this may explain why when images were magnified to  $\times 4$  their original size, the TFT display performed slightly better. This may also explain why the TFT display tended to have a higher specificity for all except the images of sight-threatening retinopathy, specificity being a measure of the ability to reliably identify truly negative images, in other words images with either no retinopathy or retinopathy below the threshold in question.

Constable *et al*,<sup>15</sup> looking at remote transfer of both compressed and uncompressed diabetic fundal images obtained with a portable fundus camera and viewed on a TFT screen, found that although transmission times were reduced by compression, even uncompressed images were inadequate for the diagnosis of diabetic retinopathy.

Another group also reported that real-time fundal images obtained using a direct ophthalmoscope attached to a digital camera and transmitted to a remote site also proved to be inadequate for screening for retinopathy.<sup>16</sup> In the latter study, it is not clear what type of monitor was used however. Owing to the remote nature of the screening in both these studies, portable imaging systems were used. Better results may have been obtained using more sophisticated retinal cameras.

Digital images can be enhanced, and manipulation of digital images with dedicated software has been used to increase grading performance.<sup>9</sup> We did not investigate such manipulation of our images and this may have improved grading performance even further. There is as yet little data on the clinical effect of software enhancement of compressed images.

In our study, we used preselected, high-quality images of diabetic retinopathy which were scanned from 35 mm transparencies; and although agreement between high-quality scanned images and originals is high,<sup>9</sup> clearly, in the screening environment most digital diabetic images would be obtained directly by a digital camera. This could have an additional effect on image quality and ultimate grading performance. In considering the grade of retinopathy, we did not investigate maculopathy in our study and further studies into the effect of screen type on detection of maculopathy would be needed in order to apply these findings more precisely to the clinical setting. In addition, the selection of specific features, such as new vessels, in the study images means that they would not be truly representative of the diabetic population. Therefore, larger studies 'in the field' are needed, whereby images from the screening environment are compressed using a variety of different compression techniques and graded remotely on different display devices.

In conclusion, grading of diabetic retinopathy yielded comparable sensitivities for both display types with nonmagnified images. Specificity tended to be slightly higher for the TFT than for the CRT. Images seemed to be less affected by compression when viewed on the CRT, and there was a trend towards more accurate grading as to the specific level of retinopathy using this display.

When images were magnified, performance was slightly greater for the TFT in all categories, and this factor together with the greater specificity when using TFT may be a feature of the greater display luminance characteristic of this type of monitor.

Both display types were adequate for screening for retinopathy at compression levels equal to or less than 80% of the original image file size. Any differences in performance between the two displays would appear to

be small. There may be some circumstances where one display type slightly outperforms the other, but from our study this would appear not to be of any clinical significance. Both display types would appear to offer similar image quality. When acquiring a digital imaging system for diabetic retinopathy screening, factors such as the cost, size and portability of the viewing system are probably the main determinants in choosing the type of monitor.

## References

- 1 Jensen PK, Scherfig E. Resolution of retinal digital colour images. *Acta Ophthalmol Scand* 1999; **77**: 526–529.
- 2 George LD, Halliwell M, Hill R, Aldington SJ, Lusty J, Dunstan F *et al*. A comparison of digital retinal images and 35 mm colour transparencies in detecting and grading diabetic retinopathy. *Diabet Med* 1998; **15**: 250–253.
- 3 Kerr D, Cavan DA, Jennings B, Dunnington C, Gold D, Crick M. Beyond retinal screening: digital imaging in the assessment and follow-up of patients with diabetic retinopathy. *Diabet Med* 1998; **15**: 878–882.
- 4 Prasad S, Bannon P, Clearkin LG, Phillips RP. Digital fundus imaging: a quality and cost comparison with 35-mm film. *Acta Ophthalmol Scand* 1999; **77**: 79–82.
- 5 Taylor DJ, Fisher J, Jacob J, Tooke JE. The use of digital cameras in a mobile retinal screening environment. *Diabet Med* 1999; **16**: 680–686.
- 6 British Diabetic Association. *Guidelines on Screening for Diabetic Retinopathy*. British Diabetic Association: London, 1999.
- 7 Retinopathy Working Party. A protocol for screening for diabetic retinopathy in Europe. *Diabet Med* 1991; **8**: 263–267.
- 8 Newsom RS, Clover A, Costen MT, Sadler J, Newton J, Luff AJ *et al*. Effect of digital image compression on screening for diabetic retinopathy. *Br J Ophthalmol* 2001; **85**: 799–802.
- 9 George LD, Lusty J, Owens DR, Ollerton RL. Effect of software manipulation (Photoshop) of digitised retinal images on the grading of diabetic retinopathy. *Br J Ophthalmol* 1999; **83**: 911–913.
- 10 Bhapkar VP, Koch GG. Hypotheses of 'no interaction' in multidimensional contingency tables. *Technometrics* 1968; **10**: 107–122.
- 11 Fleiss JL. *Statistical Methods for Rates and Proportions*, 1st edn. Wiley and Sons: New York, 1973, pp. 77–78.
- 12 Pavlicek W, Owen JM, Peter MB. Active matrix liquid crystal displays for clinical imaging: comparison with cathode ray tube displays. *J Digit Imaging* 2000; **13**: 155–161.
- 13 Roehrig H, Blume H, Ji TL, Browne M. Performance tests and quality control of cathode ray tube displays. *J Digit Imaging* 1990; **3**: 134–145.
- 14 Wang J, Anderson J, Lane T, Stetson C, Moore J. Contrast-detail characteristic evaluations of several display devices. *J Digit Imaging* 2000; **13**: 162–167.
- 15 Constable IJ, Yogesan K, Eikelboom R, Barry C, Cuypers M. Fred Hollows lecture: digital screening for eye disease. *Clin Exp Ophthalmol*. 2000; **28**: 129–132.
- 16 Marcus DM, Brooks SE, Ulrich LD, Bassi FH, Laird M, Johnson M *et al*. Telemedicine diagnosis of eye disorders by direct ophthalmoscopy. A pilot study. *Ophthalmology* 1998; **105**: 1907–1914.