Journal of Digital Imaging

SCAR R&D Symposium 2003: Comparing the Efficacy of 5-MP CRT Versus 3-MP LCD in the Evaluation of Interstitial Lung Disease

Steve Langer, PhD,[1] Brian Bartholmai, MD,[1] Ken Fetterly, MS,[1] Scott Harmsen, PhD,[1] William Ryan, MS,[1] Brad Erickson, MD, PhD,[1] Kathy Andriole, PhD,[2] and John Carrino, MD[3]

The efficacy of two medical-grade, self-calibrating, gray scale displays were compared with regard to impact on sensitivity and specificity for the detection of interstitial lung disease (ILD) on computed radiographs (CR). The displays were a 5-megapixel (MP) cathode ray tube (CRT) device and a 3-MP liquid crystal display (LCD). A sample consisting of 230 anteroposterior (AP), posteroanterior (PA), and lateral views of the chest with CT-proven findings characteristic for ILD as well as 80 normal images were compared. This double-blinded trial produced a sample sufficient to detect if the sensitivity of the LCD was 10% or more reduced (one-sided) from the ''gold standard'' CRT display. Both displays were calibrated to the DICOM gray scale standard and the coefficient of variation of the luminance function varied less than 2% during the study. Five board-certified radiologists specializing in thoracic radiology interpreted the sample on both displays and the intraobserver *Az* (area under the ROC curve) showed no significant correlation to the display used. In addition, an interobserver kappa analysis showed that the relative disagreement between any observer pair remained relatively constant between displays, and thus was display invariant. This study demonstrated there is no significant change in observer performance sensitivity on 5-MP CRT versus 3-MP LCD displays for CR examinations demonstrating ILD of the chest.

KEY WORDS: ROC, kappa, image quality, displays, interstitial lung disease, receiver operating characteristic

For nearly the first century of its existence, the practice of radiology has relied upon film as its primary detector, archive, and display device. With the advent of broadly accepted softcopy reading solutions since the mid-1990s, there has been a gradual shift toward reading diagnostic images on computer displays. Historically, those displays have been cathode ray tubes (CRT). However, in less than 10 years the prevalence of CRTs has been challenged by the LCD (liquid crystal display). It is clear that the display market is moving toward the broader use of LCD flat panel displays. The reasons for adoption of LCD over CRT are multifaceted, but is implementation of this new display technology outpacing proof of efficacy? In particular, the authors have embarked on a study to answer this question in the case of interstitial lung disease (ILD) findings on 5-megapixel (MP) CRT versus 3-MP LCD.

A standard tool of display analysis is the receiver operating characteristic (ROC), which seeks to quantify an observer's sensitivity and specificity for a given imaging task. In addition, to measure the agreement among observers for a given finding, the kappa statistic is used.[1] However, there is a Catch-22 here: To adequately size a sample to detect a given sensitivity difference among two displays, one has to have an a priori estimate of what the sensitivity gap is. A literature survey shows that the sensitivity for ILD findings for chests on film is in the range of 67%–89%.[2-4] On CRT softcopy reading, the literature quotes sensitivity ranges from 55% (early 1990s CRTs) to >83%.[4-8]

[1]*From the Department of Radiology, Mayo Clinic and Foundation, Rochester, MN.*

[2]*From the Department of Radiology, University of California at San Francisco, San Francisco, CA.*

[3]*From the Department of Radiology, Brigham and Women's Hospital, Boston, MA.*

Correspondence to: Steve Langer, PhD, tel: 507-266-4418; e-mail: langer.steve@mayo.edu

Online publication 29 June 2004

doi: 10.1007/s10278-004-1013-7

LCD results are sparse in the formal literature, but presentations at ARRS, RSNA, and SCAR seem to support that the LCD at 3 MP can compete with the 5-MP CRT.[9] This work aims to strengthen the case.

METHODS

Image Acquisition, Processing, and Display Platform

All images used in the study were acquired from Fuji CR systems and archived without persistent application of any vendor-specific lookup tables (LUTs), edge enhancement algorithms, or annotations. The images were stored on a McKesson PACS (Version 4.6) (McKesson Medical Imaging, Richmond, BC, Canada), with the Window/Level optimized for default appearance to minimize the need for the radiologist observers to manipulate or process the images. The images were displayed to fit the screen using the default downsampling algorithm of the PACS application. The CRT display was a Barco 521 (2048 × 2560); the LCD (1200 × 1600 matrix) was a Barco Coronis 3 MP (Barco Corporation, Kortrijk, Belgium).

Sample Selection

Computed radiographs (CR) of the chest were utilized in this study to maximize the potential to discover differences in display capabilities. In particular, the high spatial frequency abnormalities present in ILD potentially allow for the differentiation in diagnostic sensitivity between display devices of different matrix size.

IRB (Internal Review Board) approval was obtained to perform a retrospective records search and to utilize anonymized image data for the study. Specifically, a retrospective review of a radiology database which tracks diagnoses on computed tomography (CT) scans was performed for all studies from 1997 to 2003. This search specifically targeted chest CT scans coded for ILD. From the results of this search, a secondary correlative search of the radiology information system (RIS) was performed to extract a subset of CT-proven cases of ILD which also had CR of the chest performed within 6 months of the CT scan. The CT studies, which were discovered as a result of these searches, were pulled from the radiology archive and reviewed by a board-certified radiologist specializing in thoracic radiology (BJB). This radiologist confirmed the presence, ascertained the type, and determined the extent of ILD on each of these cases. For the purposes of this study, studies with classic features of usual interstitial pneumonitis (UIP), nonspecific interstitial pneumonitis (NSIP), chronic hypersensitivity pneumonitis (HSP), and other diffuse atypical fibrotic lung disease were included. Focal fibrotic changes, diffuse alveolar processes, or studies with predominantly ground-glass opacities were excluded from the study. In addition, where possible, cases of mild or early manifestations of ILD were preferred for inclusion to optimize the possibility of detecting differences in the two types of displays utilized in the study. Studies with significant pathology unrelated to ILD were also excluded from the study.

Each of the CR images for patients with CT-proven ILD were also analyzed by a board-certified thoracic radiology specialist (BJB). As with the selection of CT examinations, CR images with the most subtle manifestations of disease were favored for inclusion in the study, and any images with significant pathology unrelated to the proven ILD were excluded. For some patients, multiple CR studies obtained on different dates were utilized. A similar method was utilized to obtain normal CR images for inclusion in the study. A search of the radiology diagnostic database over the course of one year was used to discover negative CTs of the chest. For each of these results, a correlative search for chest CR examinations within 6 months of the CT was performed in the RIS. For each of the CR studies found, the images were evaluated by a board-certified thoracic radiology specialist, and any images with pathology or transient abnormalities such as atelectasis were excluded.

For the study, the individual normal and abnormal images were deidentified, pooled, and randomized.

Experimental Design, Data Collection, and Analysis

The sample size was based on the comparison of sensitivity estimates for the 3-MP LCD display and the 5-MP CRT display. Assuming a test of paired proportions with independent results, for a sample size of 229 patients with CT-proven findings there would be at least 80% power for a one-sided test of 10% sensitivity degradation at $\alpha = 0.05$. Also, with this sample size the 95% confidence intervals for 80% and 90% estimates of sensitivity would be 74%–85% and 86%–94%, respectively. In addition, 81 patients with normal CT images have been included in the readings. These "normal" patients were randomly assigned to be read along with the images from the 229 patients with CT-proven findings. These normal exams were mainly included to keep the readers "honest." However, estimates of specificity have also been calculated, with less precision than for sensitivity, at 80% and 90% the 95% confidence intervals would be 70%–88% and 81%–96%, respectively.

A total of five observers were used in the study: two chest imaging fellows and 3 staff radiologists with at least 12 years of postboards experience. The observer experience was split into four reading sessions of approximately 1 h each. The observer's first session was randomly assigned to either the LCD or the CRT display to avoid any training bias. Then the observer read half of the 310 image set. Within a short time (one to several days), the observer read the balance of the image set on the other display. A period of time (not less than 12 days and typically 3 or more weeks) was then used to wash out memory effects and the observer read the images again on the complementary displays.

For the first session for a given observer, the study coordinator (SC) provided instructions both verbally and in writing. In addition, the SC explained the viewing software and the observer was given a couple of training images to become comfortable with the system. The SC sat with the

observer for every session to capture the observer's ROC score for each image. The SC also recorded the observer's use of imaging tools and time per image. Figure 1 shows the data collection form: by entering the Confidence score cell, a timer is started in the time cell. As soon as the Confidence score is entered, the timer is frozen and the observer is time per image is captured.

ROC curves were constructed to summarize the discriminatory ability of an observer's confidence score in distinguishing a scan depicting lung disease from a normal scan. ROC curves were constructed for both the 5-MP CRT and 3-MP LCD displays, for each of the five observers. The performances of the two displays (5 MP vs. 3 MP) were compared, within each of the five observers, by testing for a significant difference in the paired areas under the curve (AUC). This was done using a method suggested by DeLong et al.[10] to compare correlated curves.

Sensitivity and specificity estimates, along with 95% confidence intervals, were calculated for each observer separately. Confidence levels of "likely lesion" and "certainly lesion" were considered to be calls of positive for lung disease, while levels of "uncertain," "likely normal," and "certainly normal" were considered to be a normal scan judgment.

Agreement between each pair of observers was estimated using a weighted kappa statistic, along with a 95% confidence interval, of the five ordered confidence judgments with: 1 = "certainly lesion" to 5 = "certainly normal".[11] These estimates were then interpreted.[1]

Display Setup, Calibration, and Measurements

The displays used in this study were a Barco MGD521 CRT and a Barco Coronis 3-MP LCD. The CRT was controlled by a Barco 5MP2 video card and the LCD was controlled by a Barco BarcoMed 3MP2FH display controller. The pixel matrix of the CRT was 2048 × 2560 (nominal 5 MP) and that of the LCD was 1536 × 2048 (nominal 3 MP). The pixel pitch of the CRT was 0.148 mm and that of the LCD was 0.207 mm. The luminance response of the displays was calibrated to the DICOM Grayscale Standard Display Function (DGSDF) using the vendor-provided MediCal calibration software and photometer.[12] The ambient light in the CRT and LCD environments averaged 7.5 lx, ranging from a minimum of 5 to 10 lx. Given these low ambient light levels in combination with the reflective properties of the displays, it is not expected that the difference in ambient lighting conditions affected the results of this experiment.

After calibration, the displays were evaluated subjectively using the procedures and images provided by the American Association of Physicists in Medicine Task Group 18 report *Assessment of Display Performance for Medical Imaging Systems* (further referred to as TG18).[13] At the time of this study, there were a pair of the CRTs and a pair of the LCDs in our laboratory. As each member of a display pair performed equally well with regard to the TG18 evaluation, it was concluded that all displays were functioning properly. However, only a single display of each pair was used for this study.

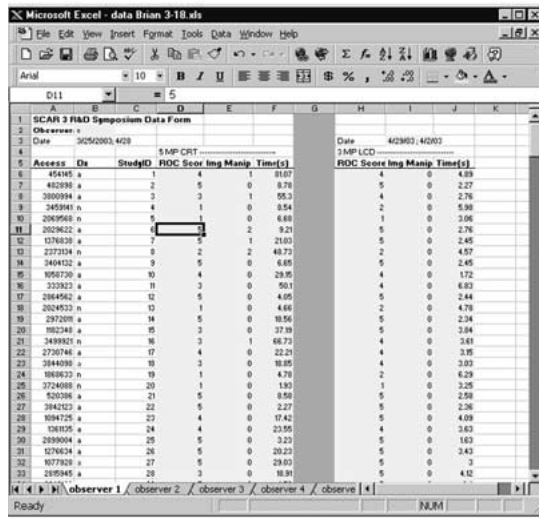

Fig 1. A sample of the data collection spreadsheet. Timing information for each image view was collected automatically; the clock started when the study coordinator entered the cell "ROC Score" and stopped as soon as a value was entered and the cell was exited.

After calibration, the TG18-LN images were used to measure the gray scale response of the displays. These images provided 18 unique input digital driving levels (DDL) equally spaced in the 8-bit range provided by the display driver cards. The 18-step luminance response of the displays was measured weekly for the duration of the project to ensure consistency of the gray scale rendition of the displayed images.

The resolution of the displays was characterized by measuring the modulation transfer function (MTF). As measured, the MTF included the effects of the finite pixel element size and the blur inherent to the display. The MTF was calculated from a digital photograph of a bright spot (an individual pixel element with a DDL of 182 placed upon a uniform background of DDL 110).

Digital photographs were acquired with a Hamamatsu Orca-ER charge coupled device (CCD) camera with a Nikon PK-13 extension tube and a Nikon 105-mm photographic lens (35-mm format) (Hamamatsu Corporation, Bridgewater, NJ). The pixel value output of the camera is linearly proportional to the incident luminance within the range of luminance levels used in this work. The pixel matrix of the camera was 1344 × 1024. The optical lens of the camera was positioned 11 cm from the display face and the lens was focused. The effective pixel pitch of the camera images was 0.005 mm, resulting in photographs of the displays that were 30× oversampled for the CRT and 42× oversampled for the LCD (calculated linearly).

A background image and a four-spot image were acquired. The background image contained only pixels with DDL of 110. The four-spot images had a background DDL of 110 and four pixels arranged on a 2 × 2 matrix with a DDL of 182. The four spots were separated by a distance of 14 and 12 display pixels for the CRT and LCD, respectively.

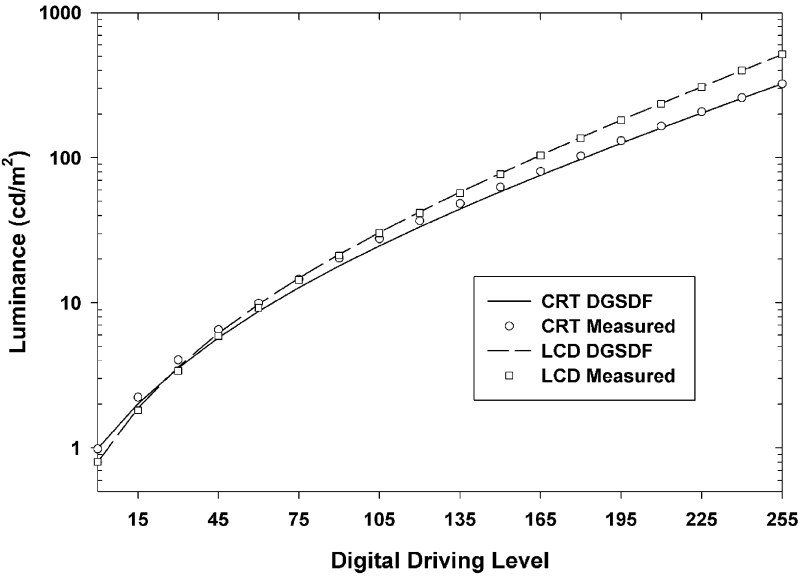

Fig 2. Measured gray scale display functions and ideal DICOM Grayscale Standard Display Function (DGSDF) for the 5-MP CRT and 3-MP LCD.

The photographic image of the background was subtracted from that of the spots. This subtraction effectively removed pixel and raster line noise that would be expected to interfere with accurate MTF characterization. Next, a $2 \times 2\text{-mm}^2$ region surrounding each of the spots was extracted and the 2D Fourier transform (FT) of each region was calculated. The MTF was defined as the modulus of the FT, normalized by the zero-frequency components. This MTF characterized the combined resolution response of the finite displayed pixel element and the blur inherent to the displays. While the 2D MTF was calculated, only the average of the horizontal and vertical MTF components are presented here.

RESULTS

Display Measurements

The calibrated luminance response of the displays is shown in Figure 2. Also shown in Figure 2 is the ideal DGSDF for the luminance range of each display. The luminance of the CRT ranged from 0.98 to 324 cd/m^2, corresponding to a gray scale just noticeable difference (JND) range of 571. The luminance of the LCD ranged from 0.8 to 517 cd/m^2, corresponding to a JND range of 649. Based on these values, the LCD demonstrated superior gray scale dynamic range. Both displays demonstrated good agreement with the DGSDF model. The weekly survey of the luminance response of the displays demonstrated good luminance stability over the 7-week duration of the experiment. For any given DDL, the maximum coefficient of variation of the luminance output was 2%.

The average 1D MTF of the two displays is shown in Figure 3. The 1D MTF of a rectangular pixel is expected to be a sinc function. The sinc function corresponding to an ideal 5-MP CRT and 3-MP LCD is also shown in Figure 3. Comparison of the display MTFs to the ideal response functions demonstrates that the LCD provides an MTF more similar to its ideal than does the CRT. That the 3-MP LCD had a greater MTF than the 5-MP CRT is an important result. This indicates that although the CRT has a greater number of addressable pixels, its ability to faithfully reproduce the information contained in a digital image is more severely compromised by its inherent blur.

Finally, Figure 4 demonstrates the number of JND/DDL (which can be thought of as the available contrast at a given DDL) versus the averaged histogram of AP and lateral views. The LCD display enjoys an advantage in all areas where there is significant image information.

In general, the quality of an electronic display is a combination of luminance response, reso-

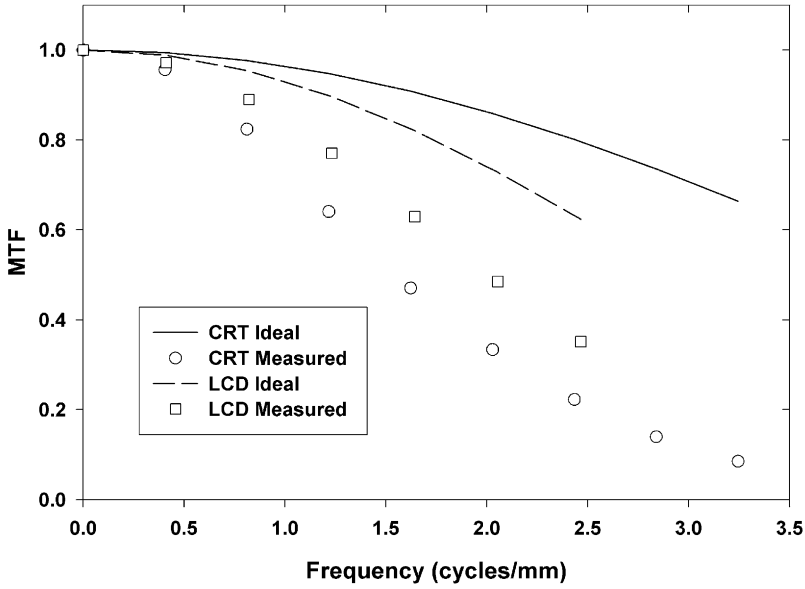

Fig 3. Modulation transfer function of the 5-MP CRT and the 3-MP LCD.

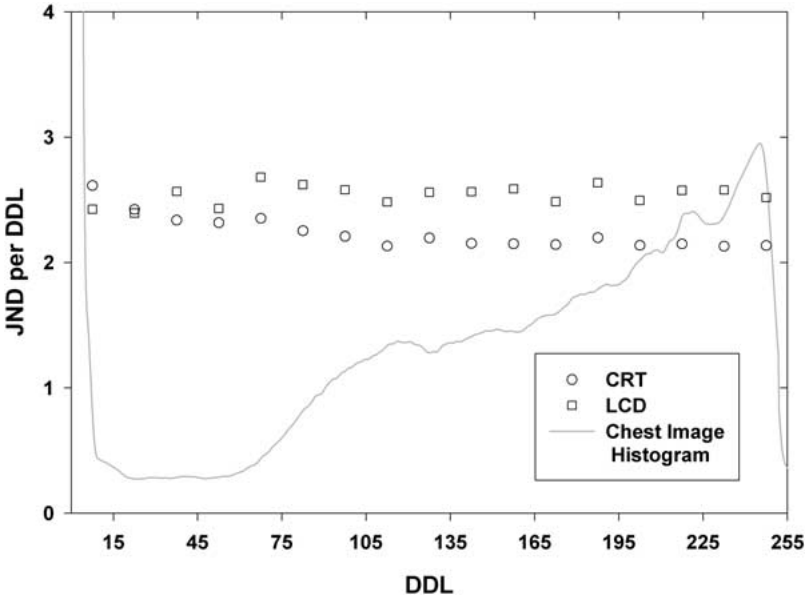

Fig 4. Comparison of the JND/DDL of the CRT and LCD displays versus a histogram constructed by averaging the AP and lateral views of both female and male patients. Note that the LCD enjoys an advantage in all areas of significant image information.

lution, and noise. Note that the noise of these displays was not measured directly. However, subjective evaluation of TG18-AFC test patterns, designed to be sensitive to display noise, did not demonstrate notable differences between these displays, (see Table 1).

Observer Results

Tool Usage

As mentioned in the Methods section, all images in the study were archived to the PACS

Table 1. Summary of Key Display Metrics

	Matrix (pitch)	Min (cd/m^2)	Max (cd/m^2)	JND
CRT	2048 × 2560 (170 dpi)	0.98	323	571
LCD	1536 × 2048 (123 dpi)	0.8	517	648

Table 2. Fraction of Cases Each Observer Spent Using Imaging Tools

	5-MP CRT	3-MP LCD
Observer 1	0.04	0.02
Observer 2	0.02	0.04
Observer 3	<0.001	0.003
Observer 4	<0.001	<0.001
Observer 5	<0.001	0.006

Observer 1 used the Window/Level or Magnification function on the LCD display 0.02 × 310 = 6 times. Observers 1 and 2 started the study on the CRT display first, while the remaining observers started on the LCD display.

system with the Window and Level optimized. Thus, observers in this study spent relatively little effort in further image-processing tasks in contrast to other studies.[9] As a result, Table 2 shows that tool usage was insignificant for observers with both displays.

Time/Image

To reduce the effect of a task acclimatization bias in the timing results, each observer was shown two or three test cases (on whatever display he was assigned to start on) before the actual timing per case commenced. Nevertheless, there was an 80% correlation of observers performing faster on the second display used, as seen in Table 3. This pattern can best be explained by noting that the observers learned certain "tricks" as they progressed (e.g., parking the cursor over the "Next Case" button while the current case loaded).

To eliminate this bias, Table 4 demonstrates the timing of observers as computed from only their performance on the second half of the survey sample for each display. In other words, the timing information from reading the first 155 exams on each display was discarded in computing the following. Note that in this example all observers (except 2) seemed able to reach a diagnosis faster on the LCD display,

Table 3. Seconds to Reach a Decision per Image

	5-MP CRT	3-MP LCD
Observer 1	7.96	5.52
Observer 2	5.79	5.64
Observer 3	3.06	3.30
Observer 4	3.65	3.45
Observer 5	3.95	4.41

Observers 1 and 2 started on the CRT. All others started on LCD. Note that except for Observer 4, all observers performed faster on the second display regardless of which display they started on.

Table 4. Average Numbers of Seconds per Image as Computed from Only the Second Half of the Sample Survey on Each Display

	5-MP CRT	3-MP LCD
Observer 1	5.58	4.51
Observer 2	4.96	5.59
Observer 3	2.96	2.43
Observer 4	3.09	3.01
Observer 5	3.80	3.71

Observers 1 and 2 started on CRT, all others started on LCD. With the exception of Observer 2, all observers rendered a decision faster on the LCD display.

which would seem to indicate greater confidence on that device.

Intraobserver ROC

To compare the observer's performance against self as a function of the display, a ROC curve was built for each observer and for each display. The ROC curve is a plot of false positive fraction (ordinate) versus true positive fraction, and a perfect curve would have an area underneath it of unity. As Figure 5 shows for a single observer, the areas of the two displays (Az) are actually very close from simple visual inspection.

The computed Az is given in Table 5. As can be seen, the Az differences are quite small, and, as the P-values demonstrate, the difference in display performance within each observer was not statistically significant.

Table 6 illustrates the intraobserver sensitivity/specificity performance, along with 95% confidence intervals, regarding confidence levels of "likely lesion" and "certainly lesion" as a

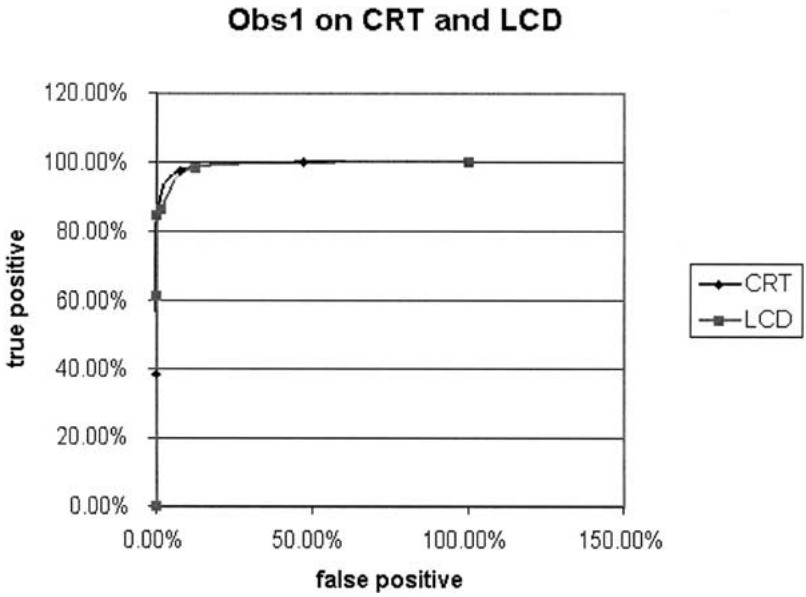

Fig 5. A representative plot of the 3-MP LCD ROC curve against that for the 5-MP CRT for a single observer.

positive judgment, while levels of "uncertain," "likely normal," and "certainly normal" were considered to be a scan as normal judgment. A point of interest is that while there is never more than a 5% intraobserver variance in sensitivity or specificity, there is a great difference in interobserver performance. This is most highly correlated to the observer's time per image, with those who rushed showing a concomitant reduction in performance.

Interobserver Kappa

Finally, we performed a pairwise kappa test to assess interobserver agreement. A kappa of unity represents perfect agreement, a kappa between 0.80 and 1.0 indicates excellent agreement, values between 0.6 and 0.8 represent substantial agreement, while those between 0.4 and 0.6 represent fair agreement.[1] As the Table 7 shows, the agreement among observers is either excellent or very good unless one compares the most sensitive observers (1, 2, and 5) with the least sensitive ones (3 and 4). Furthermore, to within the 95% confidence limits, the kappa value between any two observers is independent of the display, indicating that it is an inherent property of the observers themselves.

Table 5. *Az* Results for Intraobserver Performance

	5-MP CRT	3-MP LCD	*P*-value
Obs 1	0.993	0.986	0.16
Obs 2	0.978	0.978	0.98
Obs 3	0.967	0.970	0.67
Obs 4	0.962	0.973	0.25
Obs 5	0.986	0.984	0.48

CONCLUSIONS

Within the 80% power of this study, there is no significant difference in observer performance on the 5-MP CRT versus the 3-MP LCD as measured by intraobserver ROC (sensitivity and specificity) scores. Furthermore, there is no significant difference in tool usage on either display. A definite interobserver sensitivity difference was noted, but those differences were relatively constant between displays and correlated more to the observer pair (kappa statistic). Rather than the display used, the best predictor of observer sensitivity in this study was the time spent studying the image, with a positive correlation being shown between more time and better sensitivity.

Future Directions

During this study, it became apparent that performing new ROC studies on each new dis-

Table 6. Sensitivity and Specificity of Each Observer Versus the Display

	5 Sensitivity	5 Specificity	3 Sensitivity	3 Specificity
Obs 1	88% (83%–92%)	99% (96%–100%)	92% (88%–95%)	98% (91%–100%)
Obs 2	80% (74%–85%)	100% (96%–100%)	85% (79%–89%)	99% (93%–100%)
Obs 3	62% (55%–68%)	100% (96%–100%)	58% (51%–65%)	100% (96%–100%)
Obs 4	68% (61%–74%)	99% (93%–100%)	66% (59%–72%)	100% (96%–100%)
Obs 5	92% (87%–95%)	98% (91%–100%)	89% (84%–93%)	98% (91%–100%)

Observers 1 and 2 began on the 5-MP CRT while the others began on the 3-MP LCD. It would appear that sensitivity always improved on the second display used, indicating that task acclimatization is more deterministic than the display.

Table 7. Pairwise kappa statistic (with 95% confidence intervals in parenthesis) for all observers among the two displays. Note that unless one compares the most sensitive readers (1, 2, and 5) with the least sensitive (3 and 4), the agreement is either very good or excellent

Reviewer	2	3	4	5
5-MP CRT				
1	0.70 (0.65–0.75)	0.60 (0.55–0.65)	0.58 (0.53–0.64)	0.67 (0.62–0.71)
2		0.66 (0.61–0.71)	0.67 (0.62–0.73)	0.70 (0.64–0.75)
3			0.73 (0.69–0.78)	0.52 (0.46–0.58)
4				0.58 (0.52–0.64)
3-MPLCD				
1	0.66 (0.61–0.71)	0.52 (0.47–0.58)	0.58 (0.52–0.63)	0.70 (0.65–0.74)
2		0.62 (0.56–0.67)	0.68 (0.62–0.73)	0.72 (0.66–0.78)
3			0.72 (0.68–0.77)	0.52 (0.46–0.58)
4				0.62 (0.57–0.68)

play to validate efficacy for a given imaging task is not the best use of observer effort. Numerous workers are pursuing objective computational methods to predict the human visual response to image quality.[14] One method that shows great promise uses an analysis of the "just noticeable differences" in the image (JND Metrix, Princeton, NJ). In a similar manner, we envision a new method for display evaluation based on the following strategy:

1. For a given imaging task (ie, evaluation of lungs for ILD) construct a full-fidelity image library for that finding.
2. Establish imaging software tools capable of injecting controlled noise, blur, downsampling, or reduced contrast "on the fly" as images are displayed on the workstation.
3. Conduct ROC experiments on state-of-the-art displays with optimal calibrations, but subject the observers to sessions with each of the above-mentioned variables adjusted until thresholds are found that reduce sensitivity/specificity by a significant amount (say 10% below the performance with optimal imaging parameters).
4. Repeat for all four parameters until threshold metrics are established for spatial resolution, contrast resolution, noise, and blur for that finding. In essence, establish the system signal-to-noise ratio needed for the human visual system to perceive the finding.

At the end of the process for a given finding, the output will be display requirements for a given imaging task. Ultimately, one could imagine measuring only four parameters on a display and then consulting a catalog to establish what imaging tasks the display would be suited for. The value of this approach is that future questions of display efficacy (whether the display is film, CRT, or LCD) will be answered by convening one expert panel that will establish the imaging requirements for a particular task. From then on one need only consult the results of that panel.

REFERENCES

1. Landis, RJ, Koch, GG: The measurement of observer agreement for categorical data Biometrics 33:159-174, 1977

2. Mathieson, JR, Mayo, JR, Staples, CA, et al: Chronic diffuse infiltrative lung disease: comparison of diagnostic

accuracy of CT and chest radiography Radiology 171:111-116, 1989

3. Padley, SP, Hansell, DM, Flower, CD, et al: Comparative accuracy of high resolution computed tomography and chest radiography in the diagnosis of chronic diffuse infiltrative lung disease Clin Radiol 44(4):222-226, 1991

4. Ackerman, SC, Gitlin, JN, Gayler, RW, et al: ROC analysis of fracture and pneumonia detection: comparison of laser digitized workstation images and conventional analog radiographs Radiology 186:263-268, 1993

5. Slasky, BS, Gur, WF, Good, WF, et al: ROC analysis of chest image interpretation with conventional, laser printed and high-resolution workstation images Radiology 174(3 Pt 1):775-780, 1990

6. Scott, WW, Bluemke, DA, Mysko, WK, et al: Interpretation of ED radiographs by radiologists and emergency medicine physicians: teleradiology workstation vs. radiograph Readings Radiology 195:223-229, 1995

7. Cox, GG, Cook, LT, McMillan, JH, et al: Chest radiography: comparisons of high-resolution digital displays with conventional and digital film Radiology 176(3):771-776, 1990

8. Siegel E, Reiner, B, Moffet, BR, et al.: Differences in Image Quality of CR Images Displayed on LCD and CRT Monitors. Presentation at ARRS, 2002, Atlanta, GA

9. Siegel E, Reiner B, Moffet, BR, et al.: Differences in Image Quality of CR Images Displayed on LCD and CRT Monitors. Presentation at SCAR 2002, Cleveland, OH

10. DeLong, ER, DeLong, DM, Clarke-Pearson, DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach Biometrics 44:837-845, 1988

11. Cohen, J: A coefficient of agreement of nominal scales Educ Psychol Measure 20:37-46, 1960

12. Digital Imaging and Communications in Medicine, Part 14: Grayscale Standard Display Function, Rosslyn, VA: National Electrical Manufacturers Association, 2001

13. American Association of Physicists in Medicine Task Group 18, Assessment of display performance for medical imaging systems (draft), available at http://deckard.mc.duke.edu/~samei/tg18; accessed July 8, 2003

14. Wang, Z, Bovic, AC: A universal image quality index IEEE Signal Process Lett 3:81-84, 2002