

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/262531800>

# DICOM Gray-Scale Standard Display Function: Clinical Diagnostic Accuracy of Chest Radiography in Medical-Grade Gray-Scale and Consumer-Grade Color Displays

Article in *American Journal of Roentgenology* · June 2014

DOI: 10.2214/AJR.13.11509 · Source: PubMed

CITATIONS

7

READS

260

5 authors, including:



**Antonio J Salazar**

Los Andes University (Colombia)

18 PUBLICATIONS 59 CITATIONS

[SEE PROFILE](#)



**Diego Andres Aguirre**

Fundación Santa Fe de Bogotá

72 PUBLICATIONS 1,053 CITATIONS

[SEE PROFILE](#)



**Juliana Ocampo**

Fundación Santa Fe de Bogotá

6 PUBLICATIONS 19 CITATIONS

[SEE PROFILE](#)



**Juan C. Camacho**

Memorial Sloan Kettering Cancer Center

87 PUBLICATIONS 383 CITATIONS

[SEE PROFILE](#)

# DICOM Gray-Scale Standard Display Function: Clinical Diagnostic Accuracy of Chest Radiography in Medical-Grade Gray-Scale and Consumer-Grade Color Displays

Antonio J. Salazar<sup>1,2</sup>  
 Diego A. Aguirre<sup>3</sup>  
 Juliana Ocampo<sup>3</sup>  
 Juan C. Camacho<sup>3,4</sup>  
 Xavier A. Díaz<sup>2</sup>

**Keywords:** consumer-grade display, medical display, radiography, ROC curve, teleradiology, x-ray film

DOI:10.2214/AJR.13.11509

Received July 8, 2013; accepted after revision September 6, 2013.

Supported by the University of Los Andes and the Fundación Santa Fe de Bogotá-University Hospital.

<sup>1</sup>Department of Electrical and Electronic Engineering, University of Los Andes, Carrera 1 Este no. 19A-40, Bogotá 11001, Colombia. Address correspondence to A. J. Salazar (ant-sala@uniandes.edu.co).

<sup>2</sup>Biomedical Engineering Group (GIB), Laboratory of Telemedicine and Electrophysiology, University of Los Andes, Bogotá, Colombia.

<sup>3</sup>Department of Radiology, Fundación Santa Fe de Bogotá University Hospital, Bogotá, Colombia.

<sup>4</sup>Abdominal Imaging Division, Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA.

AJR 2014; 202:1272–1280

0361–803X/14/2026–1272

© American Roentgen Ray Society

**OBJECTIVE.** The purpose of this study was to compare the diagnostic accuracy achieved with and without the calibration method established by the DICOM standard in both medical-grade gray-scale displays and consumer-grade color displays.

**MATERIALS AND METHODS.** This study involved 76 cases, six radiologists, three displays, and two display calibrations for a total of 2736 observations in a multireader-multicase factorial design. The evaluated conditions were interstitial opacities, pneumothorax, and nodules. CT was adopted as the reference standard. One medical-grade gray-scale display and two consumer-grade color displays were evaluated. Analyses of ROC curves, diagnostic accuracy (measured as AUC), accuracy of condition classification, and false-positive and false-negative rate comparisons were performed. The degree of agreement between readers was also evaluated.

**RESULTS.** No significant differences in image quality perception by the readers in the presence or absence of calibration were observed. Similar forms of the ROC curves were observed. No significant differences were detected in the observed variables (diagnostic accuracy, accuracy of condition classification, false-positive rates, false-negative rates, and image-quality perception). Strong agreement between readers was also determined for each display with and without calibration.

**CONCLUSION.** For the chest conditions and selected observers included in this study, no significant differences were observed between the three evaluated displays with respect to accuracy performance with and without calibration.

In radiology departments in which digital images are distributed, different types of displays are used, particularly medical-grade gray-scale displays. In the past, cathode-ray tube (CRT) displays were often used. However, studies showed that the diagnostic accuracy achieved with LCD medical-grade gray-scale displays [1] was the same as that achieved with CRT medical-grade gray-scale displays. Consequently, CRT medical-grade gray-scale displays were replaced by new LCD medical-grade gray-scale displays. Other research findings [2] showed that 3-megapixel (MP) LCD medical-grade gray-scale displays were equivalent to 5-MP LCD medical-grade gray-scale displays. Studies also have been conducted to evaluate the possibility of using color LCD medical- or consumer-grade displays instead of medical-grade gray-scale displays [3, 4]. In those studies, which were performed with patterns or clinical conditions, no differences be-

tween the defined variables were found. Other investigators [5–7] assessed the levels of gray required in medical-grade displays (e.g., 8-bit versus 11-bit depth) and found better sharpness and contrast in 8-bit images than in 11-bit images with 5-MP displays [5]. In contrast, in consumer-grade devices, there has been a trend toward incorporating LED displays. For example, one study compared the LED display of an iPod device (Apple) and a color LCD display [8].

Using all of these display types may result in inconsistent presentation of medical images. Therefore, part 14 of the DICOM gray-scale standard display function (GSDF) standard established a method of display calibration for generating consistent images [9, 10]. This technique is based on the Barten model [11] to ensure presentation of images of equal perceived contrast to observers independent of the luminance range of the display. This model entails a just-noticeable difference (JND) in luminance, defined as the

## DICOM in Chest Radiography

luminance change required for the average human observer to perceive the luminance change at a given luminance level. Each JND is associated with a JND index: One step in the JND index results in a luminance difference of 1 JND. This mapping is defined by the GSDF. When all of the luminance values produced by the display system are measured for all of the possible inputs into the display system—the digital driving level—the resultant map is the characteristic curve of the display system. In medical-grade displays the characteristic curve is close to the GSDF, and the range of luminance is larger than in consumer-grade displays. The GSDF calibration method distributes the luminance range of a display system according to the Barten model (i.e., constant contrast at any gray level). In medical-grade gray-scale displays, calibration is accomplished with a hardware lookup table (LUT) during the manufacturing process and may be adjusted at hospitals by use of a specialized photometer with feedback to the display. In consumer-grade displays, the calibration may be achieved with a software LUT implemented by the medical application software or by the operating system.

Neither of the aforementioned studies and to our knowledge no other studies have compared the real clinical effect of the presence or absence of GSDF calibration on the diagnostic accuracy of reading images on LCD or LED consumer-grade displays. In a previous study [12], we compared clinical accuracy in visualization of digitized images of chest radiographs on LCD and LED consumer-grade and medical-grade displays calibrated according to the DICOM standard. We found no differences in accuracy between the displays. As a next step, in the current study, we were interested in evaluating what would happen if the same displays were not calibrated, that is, used with the original factory calibration. Therefore, the purpose of this study was to compare the diagnostic accuracy achieved with and without the GSDF calibration method in both medical-grade gray-scale displays and consumer-grade color displays.

The clinical conditions evaluated were interstitial opacities, pneumothorax, and nodules. CT was adopted as the reference standard. In a multireader-multicase study design, comparisons of ROC curves [13], diagnostic accuracy measured as AUC [14–16], accuracy of condition classification, false-positive and false-negative rates, and the main factors affecting clinical accuracy and reading time

were evaluated. The degree of agreement between readers was also assessed.

### Materials and Methods

The methods used in this study were previously reviewed and approved by the two ethics committees at our institutions. This study involved 76 cases, six radiologists, three displays, and two display calibrations (i.e., factory calibration and DICOM GSDF calibration) for a total of 2736 observations in a multireader-multicase (MRMC) factorial design.

### Study Sample and Readers

Cases corresponded to digital chest radiographs acquired with computed radiography devices at a high-complexity university hospital in Bogotá, Colombia, between November 2007 and June 2009. Cases were randomly selected without repetition and were included in the sample if chest CT scans were available to establish the reference standard. CT was used to determine the true findings in normal and pathologic cases and to quantify pneumothorax size and nodule size. The readers were six hospital radiologists (2–10 years of experience after board certification) selected by the chief of the department of radiology according to their time availability and to balance their years of experience to achieve moderate variability among observers. Readers were treated as fixed effects in statistical analysis. With a ratio of pathologic to normal cases of 4:1, six radiologists, and expected differences in treatment AUCs of 0.1, this study was performed with 76 cases on the basis of the Obuchowsky table [17] for sample size selection in AUC comparisons. The distribution by condition was as follows: 20 cases of interstitial opacities (eight cases with a fine or reticular pattern, 12 cases with a nodular or reticulonodular pattern), 16 cases of pneumothorax (10 cases smaller than 25%, five cases, 25–50%, and one case larger than 50%), and 18 cases of nodules (five cases smaller than 7 mm; 11 cases, 7–15 mm; and two cases larger than 15 mm). There were 27 healthy subjects (i.e., without any of the selected conditions). This distribution was selected to achieve a sample distribution similar to the population distribution at our hospital. Conditions in cases were not exclusive; that is, multiple types of lesions per case were allowed. To avoid sensitivity bias, cases with obvious lesions were not included [18].

### Observed Variables

The conditions selected for this study were interstitial opacities, pneumothorax, and nodules. To compare the displays, several variables related to these conditions were defined, and the main effect factors (i.e., radiologist, display, and calibration) and their interactions were evaluated for the following variables.

*Image quality perception*—Image quality perception, a binary variable, was used to evaluate the primary perception of the observer to declare (before using image enhancement tools) whether the images were or were not amenable to interpretation.

*Diagnostic accuracy*—Diagnostic accuracy, measured as the AUC for each condition, was calculated for the level of confidence of each radiologist in the presence of each selected condition, that is, interstitial opacities, pneumothorax, and nodules. For each of these conditions, the observer selected one of the following scores: 0, definitely absent; 1, most likely absent; 2, cannot decide; 3, most likely present; or 4, definitely present.

*False-positive and false-negatives rates*—False-positive and false-negatives rates were calculated from the scores selected by observers in the diagnostic accuracy variables.

*Accuracy of condition classification*—The observers classified other aspects of the selected conditions: interstitial opacity patterns, largest nodule size, and pneumothorax size (as a percentage, quantified by the Rhea method) [19]. Accuracy of condition classification was calculated as the proportion of cases correctly classified.

*Agreement in condition classification*—Agreement between the six radiologists on condition classifications with or without use of GSDF calibration was measured with the kappa statistic [20] and then ranked as defined by Landis and Koch [21].

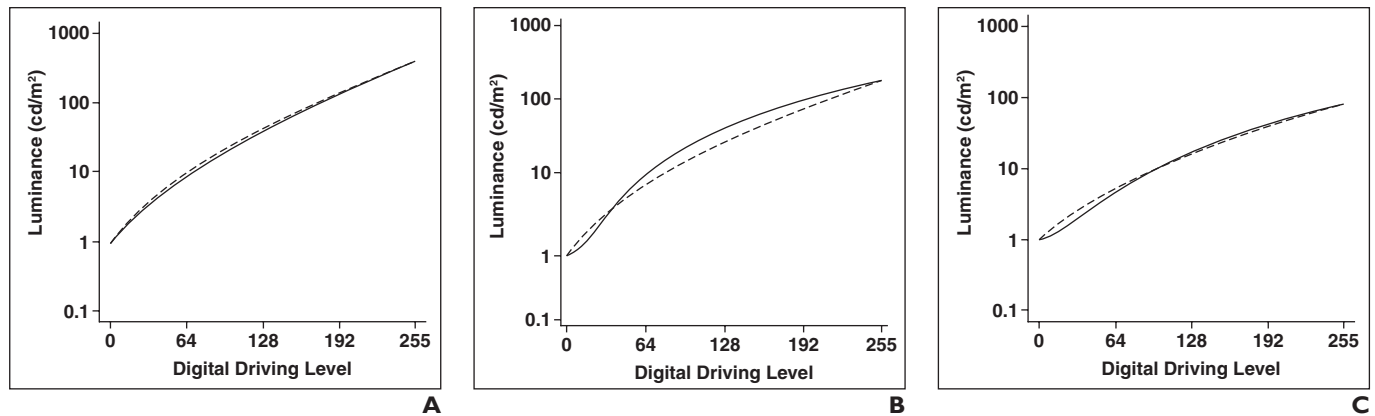
*Reading time*—The reading time spent by a radiologist on the observation of each case was measured to evaluate the main effect factors (radiologist, display type, and calibration) affecting the reading time.

### X-Ray Film Capture

The chest radiographs were printed on 35 × 43 cm film with a digital film printer (Agfa Drystar 5503, Agfa HealthCare) at 508-dpi resolution and 14-bit contrast. Each film was digitized with a film digitizer (iCR-612SL, iCRcompany) in previous studies [22, 23] at 375 dpi (6488 × 5248 matrix) in 8-bit gray scale. Larger resolutions were not used because they result in large files in the context of teleradiology services in rural areas with no high-speed broadband available. The corresponding images were stored in DICOM format without compression.

### Capture and Display Software

DICOM-compliant software developed for our previous studies [22–24] was used to review the images and to enter the observed data for each variable. This software entails image manipulation functions that can be used according to the reader's criteria: filters, zoom, brightness and contrast,



**Fig. 1**—Graphs show logarithmic display characteristic curves. Dashed line indicates transformed curve with gray-scale standard display function calibration; solid line, characteristic curve without calibration.

**A**, Medical-grade gray-scale display (3 MP).

**B**, LCD display.

**C**, LED display.

window and level, and negative and positive. The software blinded the radiologists to the patients' identities and conditions. Tools for measuring pneumothorax and nodule size were also incorporated.

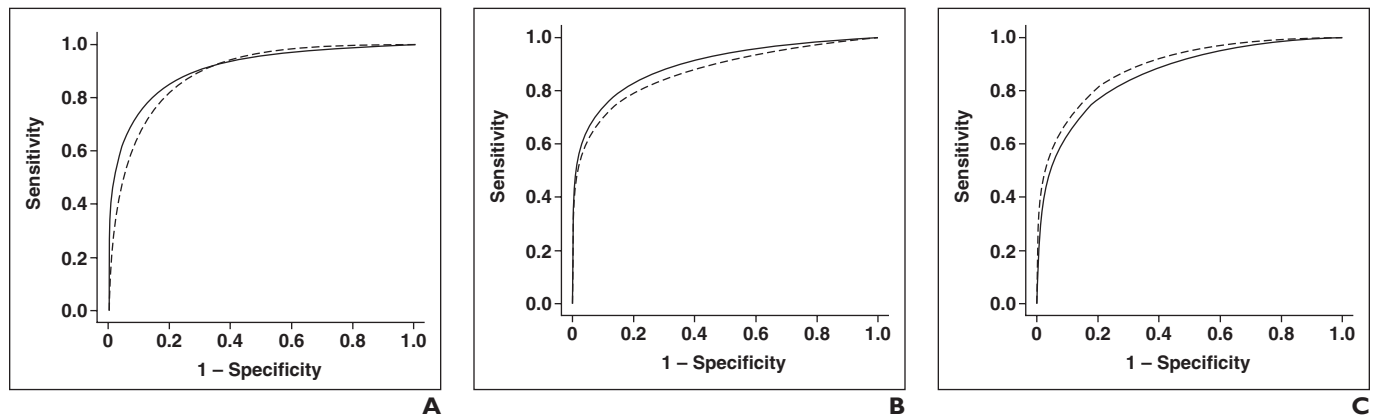
#### Displays and Calibration

Three different technology monitor displays were evaluated. The medical-grade display was a 3-MP gray-scale display (MD213MG, NEC Display Solutions), hereafter referred to as 3MP, with a dot pitch of 0.21 mm, spatial resolution of  $2048 \times 1536$  pixels, maximum luminance of  $1450 \text{ cd/m}^2$ , 10-bit gray-scale, and cost of \$15,000. The first consumer-grade display was an UltraSharp U2711 LCD display (Dell Computer Corporation), hereafter referred to as LCD, with a dot pitch of 0.23 mm, spatial resolution of  $2560 \times 1440$  pixels, maximum luminance of  $350 \text{ cd/m}^2$ , and cost of \$862. The other consumer-grade display was

the LED display of a Vostro 3750 laptop computer (Dell Computer Corporation), hereafter referred to as LED, with a dot pitch of 0.24 mm,  $1600 \times 900$  pixels, maximum luminance of  $220 \text{ cd/m}^2$ , and cost of \$780.

Each display was calibrated according to the DICOM GSDF. For the three displays, the first step in calibration was a contrast and brightness setup that allowed correct visualization of the low-contrast patterns of whites and blacks of the RP-133 standard pattern (RP indicating random practice) created by the Society of Motion Picture and Television Engineers [25–28]. The display characteristic curves were then obtained with a USB photometer (Mavo-Monitor, Gossen Foto- und Lichtmesstechnik), which has a measured range of  $0.01\text{--}19,990 \text{ cd/m}^2$ . For each display, luminosity was measured with this photometer for input digital driving levels between 0 (for

black) and 255 (for white) without ambient light. The minimum luminance values measured for the 3MP, LCD, and LED displays were 0.92, 0.24, and  $0.49 \text{ cd/m}^2$ . These luminance values corresponded to JND minimum values for 3MP, LCD, and LED of 68.13, 59.82, and  $71.09 \text{ cd/m}^2$ . The maximum luminance values measured for 3MP, LCD, and LED were 388.50, 178.17, and  $90.70 \text{ cd/m}^2$ . These maximum luminance values corresponded to JND maximum values for 3MP, LCD, and LED of 668.49, 556.19, and  $464.10 \text{ cd/m}^2$ . Using the maximum and minimum luminance values from the display characteristic curve, the numbers of JNDs ( $n_{JND}$ ) were also calculated and found to be 600.36, 496.36, and 393.02 for 3MP, LCD, and LED. The ambient light luminance reflecting on the display (with the display turned off) at the site read was  $0.5 \text{ cd/m}^2$  for the three displays with a controlled ambient luminosity of 20 lux. This val-



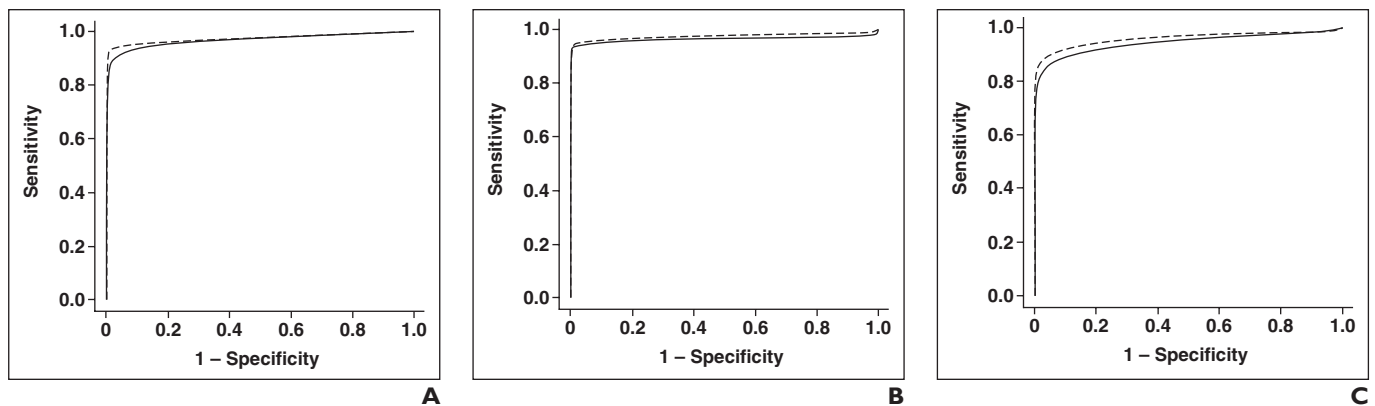
**Fig. 2**—Graphs show fitted binormal ROC curves for interstitial opacities. Dashed line indicates with gray-scale standard display function (GSDF) calibration; solid line, without GSDF calibration.

**A**, Medical-grade gray-scale display (3 MP).

**B**, LCD display.

**C**, LED display.

## DICOM in Chest Radiography



**Fig. 3**—Graphs show fitted binormal ROC curves for pneumothorax. Dashed line indicates with gray-scale standard display function (GSDF) calibration; solid line, without GSDF calibration.

**A**, Medical-grade gray-scale display (3 MP).

**B**, LCD display.

**C**, LED display.

ue was added to the display characteristic curve to obtain the characteristic curve for the reading set-up (Fig. 1). The transformed display curves, after DICOM GSDF calibration, for each of the three displays are shown in Figure 1. The calibration was accomplished with LUTs in the visualization software, even though the 3MP display includes a factory GSDF calibration.

### Procedure

This study was conducted with a treatment-by-reader-by-case factorial design. For each display (treatment), the six radiologists (readers) observed each digitized chest radiographic film (case). In each reading session, the radiologist verified the contrast and luminance settings of the display with the RP-133 pattern at a controlled ambient luminosity of approximately 20 lux (which produces a luminance of 0.5 cd/m<sup>2</sup>) for all readings. The images were interpreted over a 6-month period in 4-hour

sessions by each radiologist. The presentation of the cases was random for each display but assured a 76-case interval between two observations of the same case by one radiologist to avoid recall bias.

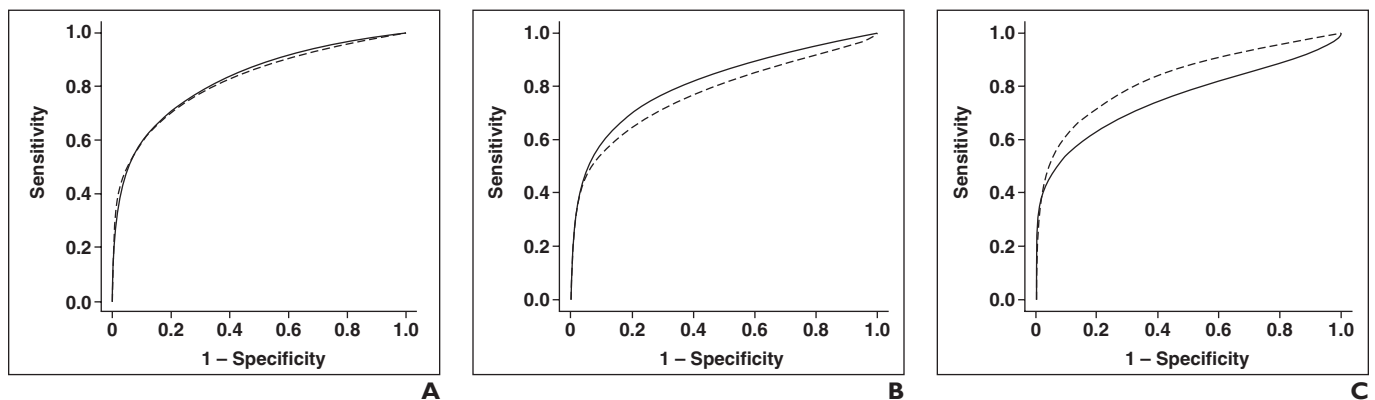
### Data Analysis

Statistical analysis on the observed variables and the factors affecting these variables (display, calibration, radiologist, and their interactions) was performed with SPSS statistics software (version 19, IBM SPSS). To evaluate the diagnostic accuracy and the reading time, MRMC ANOVA was performed with SPSS software. To estimate and compare the AUC values, we performed MRMC ANOVA on AUC pseudovalues generated by the jackknife method [29] using DBM MRMC software (version 2.3, Medical Image Perception Laboratory) [29–37]. A parametric binormal adjustment [38] with a contaminated binormal model [39, 40] was selected. To estimate and compare

image quality perception, the accuracy of condition classifications, and the false-positive and false-negative rates, which are proportions variables, we used generalized estimating equations in the SPSS program. Agreements with and without GSDF calibration were performed with Stata software (version 12.1, Stata).

### Results

The results were obtained from a total of 2736 observations (76 cases  $\times$  six radiologists  $\times$  3 displays  $\times$  2 calibrations). No differences in the image quality perception of the readers were observed in comparisons of the images yielded by each display with and without GSDF calibration: 3MP,  $p = 1.0$ ; LED,  $p = 0.35$ , LCD,  $p = 0.15$ . The proportions of cases marked as amenable to interpretation (i.e., the image quality perception variables) for displays with and without GSDF calibration



**Fig. 4**—Graphs show fitted binormal ROC curves for nodules. Dashed line indicates with gray-scale standard display function (GSDF) calibration; solid line, without GSDF calibration.

**A**, Medical-grade gray-scale display (3 MP).

**B**, LCD display.

**C**, LED display.

**TABLE 1: Comparison of AUC Values for Displays, Calibration, and Their Interactions**

Monitor	GSDF Calibration	AUC <sup>a</sup>	SE	95% CI		<i>p</i> <sup>b</sup>
				Lower Bound	Upper Bound	
Interstitial opacities						
3MP	Overall	0.91	0.03	0.85	0.97	0.69
LCD	Overall	0.90	0.03	0.84	0.96	
LED	Overall	0.90	0.03	0.83	0.96	
Overall	No	0.91	0.03	0.85	0.96	0.50
Overall	Yes	0.90	0.03	0.83	0.96	
Nodules						
3MP	No	0.94	0.03	0.88	0.99	0.32
	Yes	0.89	0.04	0.81	0.97	
LCD	No	0.89	0.04	0.81	0.97	
	Yes	0.91	0.03	0.85	0.96	
LED	No	0.90	0.03	0.84	0.97	
	Yes	0.89	0.03	0.82	0.96	
Pneumothorax						
3MP	Overall	0.83	0.04	0.76	0.90	0.95
	LCD	Overall	0.82	0.04	0.74	
LED	Overall	0.82	0.04	0.74	0.90	
Overall	No	0.83	0.04	0.75	0.90	0.53
Overall	Yes	0.82	0.04	0.75	0.89	
Interactions						
3MP	No	0.83	0.04	0.75	0.91	0.17
	Yes	0.83	0.04	0.75	0.90	
LCD	No	0.81	0.05	0.70	0.91	
	Yes	0.83	0.04	0.76	0.91	
LED	No	0.85	0.04	0.78	0.92	
	Yes	0.79	0.05	0.70	0.89	
Pneumothorax						
3MP	Overall	0.93	0.04	0.84	1.00	0.96
LCD	Overall	0.92	0.03	0.85	0.99	
LED	Overall	0.94	0.01	0.91	0.96	
Overall	No	0.91	0.02	0.87	0.95	0.36
Overall	Yes	0.95	0.03	0.89	1.00	
Interactions						
3MP	No	0.89	0.02	0.85	0.92	0.18
	Yes	0.97	0.09	0.79	1.00	
LCD	No	0.87	0.07	0.73	1.00	
	Yes	0.97	0.02	0.94	1.00	
LED	No	0.97	0.02	0.94	1.00	
	Yes	0.90	0.01	0.88	0.92	

Note—GSDF = gray-scale standard display function, SE = standard error of the mean, 3MP = 3-MP medical-grade gray-scale display. LCD and LED are consumer-grade displays.

<sup>a</sup>Each AUC was calculated for 456 observations.

<sup>b</sup>The hypothesis is as follows: The mean AUC values for the compared factors were equal. None of the differences were significant ( $p > 0.05$ ).

ranged from 0.884 (for LED without calibration) to 0.969 (for 3MP with calibration).

The shapes of the ROC curves (Figs. 2–4) were similar for all of the displays with and

without GSDF calibration for each condition. Table 1 shows the AUC values observed for each condition. The ranges of AUC values grouped by displays, calibration, and their

interactions were 0.89–0.94 for interstitial opacities, 0.87–0.97 for pneumothorax, and 0.79–0.85 for nodules. In addition, the AUC means of the monitor factor, the calibrations factor, and their interactions did not significantly differ for any condition.

The values obtained for the accuracy of condition classification are shown in Table 2. For pneumothorax the range of values was 0.80–0.84. Differences were not observed for the monitor factor ( $p = 0.11$ ), the calibration factor ( $p = 0.60$ ), or the monitor and calibration interaction ( $p = 0.98$ ). With respect to nodules, no differences were detected between calibrations or between the monitor and calibration combinations. In contrast, significant differences were detected between displays ( $p = 0.03$ ). Nevertheless, the maximum difference between displays was only 2%, values ranging from 0.87 to 0.89. In regard to interstitial opacities, the range of values was 0.72–0.78. Significant differences ( $p = 0.03$ ) were detected for the monitor and calibration interaction; no differences were detected between calibrations ( $p = 0.95$ ); and the display factor was significant ( $p = 0.01$ ) but with the higher proportion in favor of the LED display.

With respect to the false-positive and false-negative rate variables (Table 3), no significant differences were observed for the calibration factor in any of the three conditions. For interstitial opacities, the monitor factor and the monitor and calibration interaction were significant for false-positive rate. For monitors the false-positive rates were 13.8% (126/912) for 3MP, 13.8% (126/912) for LCD, and 10.1% (92/912) for LED. Even if there was a significant difference, this difference was in favor of the consumer-grade LED display, which had the lower false-positive rate. The same trend was observed for the monitor and calibration interaction. For false-negative rate, differences also were observed in the monitor factor and in the monitor and calibration interaction. However, the rates were low, ranging from 3.3% to 5.4% for monitors and 3.3% to 6.4% for the interaction. For nodules, the only significant factor in false-positive rate was monitor ( $p < 0.001$ ): 9.5% (87/912) for 3MP, 5.5% (50/912) for LCD, and 8.8% (80/912) for LED. For false-negative rate, there were no significant differences, the values being 6.6–9.4%. For pneumothorax, the only significant factor for false-negative rate was the monitor ( $p = 0.13$ ), the values being 2.1% (19/912) for 3MP, 4.2% (38/912) for LCD, and 1.9% (17/912) for LED. Nevertheless, rates were low for the three displays. For false-positive rate, there were no significant

## DICOM in Chest Radiography

differences, and values between 0% and 0.7% were detected.

The agreements between radiologists on condition classifications by display (with or without GSDF calibration) are shown in Table 4. The agreements were all ranked as moderate for interstitial opacities, the observed agreement being 74.5–80.1%. For nodule size, the agreements were all ranked as moderate for the 3MP and LED displays and as substantial for the LCD display, the observed agreement being 80.7–84.8%. Moreover, all of the agreements were ranked as almost perfect for pneumothorax size. Finally, the ANOVA of reading time showed that GSDF calibration had no effect on reading time ( $p > 0.36$ ).

### Discussion

High proportion values of image-quality perceptions (range, 0.88–0.97) were found with no differences between perceptions in the presence or absence of GSDF calibration. Similar form and high accuracy were determined for the three AUC condition variables for all of the tested displays with and without GSDF calibration. The range of AUC values observed was 0.79–0.97, and no differences were observed between the AUC means. High proportion values of correctly classified conditions were also noted (range, 0.72–0.89). In addition, no effect was noted in the false-positive and false-negative rates or in reading time for the calibration factor. Strong agreement between readers was also observed for each display with and without GSDF calibration. These results allow us to conclude that GSDF calibration had no effect on the performance achieved with the three tested displays and with the selected chest conditions and observers.

In previous studies of the GSDF, the researchers focused on the importance and method of calibration [6, 10, 41, 42], the need for an 11-bit depth in medical-grade gray-scale displays [5, 7], the effects of room illumination [43], the use of GSDF with film digitizers [44], and comparisons of color and gray-scale medical-grade LCD displays using GSDF calibration whenever possible [3, 4]. Nevertheless, none of these studies compared the effects of GSDF calibration on diagnostic accuracy, and none included LED displays, as our study did.

Uemura et al. [45] concluded that the calibration curve of the Commission Internationale de l'Eclairage is more suitable than the DICOM GSDF curve for calibrating diagnostic LCD monitors. Asai et al. [41] also found that the GSDF cannot achieve percep-

**TABLE 2: Comparisons of Proportions of Cases Correctly Classified for Displays, Calibration, and Their Interactions**

Monitor	GSDF Calibration	<i>n</i>	Proportion	SE	95% CI		<i>p</i> <sup>a</sup>
					Lower Bound	Upper Bound	
Interstitial opacities							
3MP	Overall	912	0.74	0.03	0.69	0.79	0.01 <sup>b</sup>
LCD	Overall	912	0.73	0.03	0.68	0.79	
LED	Overall	912	0.77	0.03	0.72	0.82	
Overall	No	1368	0.75	0.03	0.70	0.80	0.95
Overall	Yes	1368	0.75	0.03	0.70	0.80	
3MP	No	456	0.77	0.03	0.71	0.82	0.03 <sup>b</sup>
	Yes	456	0.72	0.03	0.66	0.78	
LCD	No	456	0.72	0.03	0.66	0.78	
	Yes	456	0.74	0.03	0.69	0.80	
LED	No	456	0.76	0.03	0.70	0.81	
	Yes	456	0.78	0.03	0.72	0.84	
Nodules							
3MP	Overall	912	0.88	0.03	0.83	0.94	0.03 <sup>b</sup>
LCD	Overall	912	0.87	0.03	0.80	0.93	
LED	Overall	912	0.89	0.03	0.84	0.95	
Overall	No	1368	0.88	0.03	0.82	0.94	0.36
Overall	Yes	1368	0.89	0.03	0.83	0.94	
3MP	No	456	0.88	0.03	0.82	0.94	0.42
	Yes	456	0.89	0.03	0.83	0.95	
LCD	No	456	0.86	0.03	0.80	0.93	
	Yes	456	0.88	0.03	0.81	0.94	
LED	No	456	0.89	0.03	0.84	0.95	
	Yes	456	0.89	0.03	0.83	0.95	
Pneumothorax							
3MP	Overall	912	0.80	0.03	0.75	0.85	0.11
LCD	Overall	912	0.83	0.03	0.78	0.89	
LED	Overall	912	0.82	0.03	0.77	0.87	
Overall	No	1368	0.82	0.03	0.77	0.87	0.60
Overall	Yes	1368	0.82	0.02	0.77	0.87	
3MP	No	456	0.80	0.03	0.74	0.86	0.98
	Yes	456	0.80	0.03	0.75	0.85	
LCD	No	456	0.83	0.03	0.77	0.89	
	Yes	456	0.84	0.03	0.78	0.89	
LED	No	456	0.82	0.03	0.77	0.87	
	Yes	456	0.82	0.03	0.77	0.88	

Note—GSDF = gray-scale standard display function, *n* = number of observations, SE = standard error of the mean, 3MP = 3-MP medical-grade gray-scale display. LCD and LED are consumer-grade displays.

<sup>a</sup>The hypothesis is as follows: The mean proportion values for the compared factors were equal.

<sup>b</sup>Statistically significant ( $p < 0.05$ ).

tual linearization for calibrating diagnostic LCD displays. Our results agree with the results of those studies in the sense that GSDF calibration does not improve the diagnostic

accuracy of medical displays. In addition, no effect was detected in the tested consumer-grade displays. In contrast, another study [44] concluded that the GSDF is adequate

**TABLE 3: Comparisons of Proportions of False-Positive and False-Negative Findings for Displays, Calibration, and Their Interactions**

Monitor	GSDF Calibration	n	False-Positive Rate					False-Negative Rate				
			Proportion	SE	95% CI		p <sup>a</sup>	Proportion	SE	95% CI		p <sup>a</sup>
					Lower Bound	Upper Bound				Lower Bound	Upper Bound	
Interstitial opacities												
3MP	Overall	912	0.138	0.021	0.10	0.18	< 0.001 <sup>b</sup>	0.036	0.013	0.01	0.06	0.03 <sup>b</sup>
LCD	Overall	912	0.138	0.022	0.10	0.18		0.048	0.017	0.01	0.08	
LED	Overall	912	0.101	0.018	0.07	0.14		0.054	0.018	0.02	0.09	
Overall	No	1368	0.131	0.020	0.09	0.17	0.26	0.043	0.015	0.01	0.07	0.15
Overall	Yes	1368	0.121	0.019	0.08	0.16		0.049	0.017	0.02	0.08	
3MP	No	456	0.123	0.020	0.08	0.16	0.02 <sup>b</sup>	0.039	0.015	0.01	0.07	0.02 <sup>b</sup>
	Yes	456	0.154	0.025	0.10	0.20		0.033	0.012	0.01	0.06	
LCD	No	456	0.151	0.025	0.10	0.20		0.046	0.017	0.01	0.08	
	Yes	456	0.125	0.021	0.08	0.17		0.050	0.019	0.01	0.09	
LED	No	456	0.118	0.021	0.08	0.16		0.044	0.015	0.01	0.07	
	Yes	456	0.083	0.018	0.05	0.12		0.064	0.021	0.02	0.10	
Nodules												
3MP	Overall	912	0.095	0.015	0.07	0.13	< 0.001 <sup>b</sup>	0.073	0.020	0.03	0.11	0.23
LCD	Overall	912	0.055	0.012	0.03	0.08		0.086	0.023	0.04	0.13	
LED	Overall	912	0.088	0.015	0.06	0.12		0.070	0.021	0.03	0.11	
Overall	No	1368	0.083	0.014	0.06	0.11	0.15	0.077	0.021	0.04	0.12	0.90
Overall	Yes	1368	0.075	0.013	0.05	0.10		0.076	0.021	0.04	0.12	
3MP	No	456	0.101	0.018	0.07	0.14	0.92	0.070	0.022	0.03	0.11	0.35
	Yes	456	0.090	0.016	0.06	0.12		0.077	0.021	0.04	0.12	
LCD	No	456	0.059	0.013	0.03	0.09		0.094	0.027	0.04	0.15	
	Yes	456	0.050	0.014	0.02	0.08		0.077	0.021	0.04	0.12	
LED	No	456	0.090	0.016	0.06	0.12		0.066	0.020	0.03	0.10	
	Yes	456	0.086	0.016	0.05	0.12		0.075	0.024	0.03	0.12	
Pneumothorax												
3MP	Overall	912	0.004	0.138	-0.27	0.27	1.00	0.021	0.008	0.01	0.04	0.01 <sup>b</sup>
LCD	Overall	912	0.002	0.000	0.00	0.00		0.042	0.015	0.01	0.07	
LED	Overall	912	0.000	0.104	-0.20	0.20		0.019	0.009	0.00	0.04	
Overall	No	1368	0.003	0.000	0.00	0.00	1.00	0.029	0.010	0.01	0.05	0.31
Overall	Yes	1368	0.001	0.000	0.00	0.00		0.026	0.010	0.01	0.05	
3MP	No	456	0.007	0.000	0.01	0.01	0.98	0.024	0.010	0.01	0.04	0.80
	Yes	456	0.002	0.000	0.00	0.00		0.018	0.007	0.00	0.03	
LCD	No	456	0.002	0.000	0.00	0.00		0.044	0.015	0.02	0.07	
	Yes	456	0.002	0.000	0.00	0.00		0.039	0.016	0.01	0.07	
LED	No	456	0.000	0.083	-0.16	0.16		0.018	0.009	0.00	0.04	
	Yes	456	0.000	0.000	0.00	0.00		0.020	0.010	0.00	0.04	

Note—GSDF = gray-scale standard display function, n = number of observations, SE = standard error of the mean, 3MP = 3-MP medical-grade gray-scale display. LCD and LED are consumer-grade displays.

<sup>a</sup>The hypothesis is as follows: The mean proportion values for the compared factors were equal.

<sup>b</sup>Statistically significant ( $p < 0.05$ ).

**TABLE 4: Agreements Between Radiologists on the Condition Classifications by Display (With or Without Gray-Scale Standard Display Function Calibration) and Condition**

Condition Classification	Display	Observed Agreement (%)	Expected Agreement (%)	$\kappa$	Agreement
Interstitial opacity pattern	3MP	74.5	46.0	0.527	Moderate
	LED	77.1	47.3	0.565	Moderate
	LCD	80.1	51.6	0.589	Moderate
Nodule size	3MP	80.7	58.0	0.541	Moderate
	LED	84.2	65.1	0.547	Moderate
	LCD	84.8	58.4	0.636	Substantial
Pneumothorax size	3MP	95.2	65.3	0.863	Almost perfect
	LED	95.0	68.5	0.842	Almost perfect
	LCD	96.5	65.8	0.899	Almost perfect

Note—Each proportion was calculated from 912 observations and 152 readings by the raters. 3MP = 3-MP medical-grade gray-scale display. LCD and LED are consumer-grade displays.

for uniformly displaying all radiologic images. No consensus has been achieved regarding differences between 8-bit and 10-bit displays. Hiwasa et al. [7], on the basis of their test patterns, concluded in favor of using 10-bit display because low-contrast objects were better discerned. Bender et al. [5], however, concluded that even if higher resolution resulted in more complete visualization of image information, radiologists judged this as a lack of sharpness and contrast and generally preferred the 8-bit display. Our study was conducted with 8-bit images because the remote referral sites in our teleradiology services are implanted with 8-bit digitizers.

An important limitation of our study was the way in which GSDF calibration was implemented: The luminance transformed characteristic curve obtained after calibration is an approximation of the GSDF curve because the calibrated values are selected from an 8-bit LUT, producing quantization errors [6]. These errors may be reduced by use of a LUT from the hardware driver of the display and use of more complex methods, such as the International Color Consortium color profile [10]. These profiles, however, are not always available for all display monitors.

## Conclusion

Performing or not performing display calibration according to the DICOM GSDF curve does not appear to be associated with a difference in diagnostic performance in chest radiography when interpretations are realized with both medical-grade gray-scale displays and consumer-grade color LCD or LED displays. The reason may be that current mon-

itors come with a factory calibration that somewhat resembles the GSDF, as shown in Figure 1, and that the GSDF is designed to model human gray-scale discrimination tasks rather than detection tasks.

Our results suggest that the tested displays may be recommended for reading digital chest radiographs. However, the displays evaluated in this study were not selected randomly (i.e., they were fixed factors in the MRMC analysis). Consequently our results apply only to the tested displays. To determine the diagnostic accuracy of other displays, especially consumer-grade displays, a new study similar to this one must be undertaken. In our case, this may be four times the cost of our medical-grade 3MP display. As stated by Geijer [4], the advantage of color display is the possibility of exchanging it four times as often as monochrome display within a fixed budget. Nevertheless, if a study to evaluate a different display is required, the cost of the study must be compared with the cost of all of the specialized medical displays required in a radiology service to determine the selection of a medical-grade display or a consumer-grade display.

The case readout at the monitors was performed in the following order: LED monitor, 3MP medical display, and LCD commercial display. The 3MP medical display was intentionally placed between the commercial displays to allow the same time interval between them and the medical display. Although the case readout sessions were conducted in the same order for all observers, no evident bias was found during the statistical analysis. For instance, the higher proportions of cases cor-

rectly classified were achieved with the first display used (LED commercial display), but the reading time increased significantly when the 3MP medical display was used. A potential limitation of the study includes the possible bias introduced by the readers themselves given the overall perception of low-quality images when they used the commercial display, even though the digitized image was exactly the same. When the same image was visualized simultaneously on the three displays, the radiologists perceived that the consumer-grade color displays produced a brighter image (increased gray scale) with a green tint, which may not be appropriate for accurate diagnosis. However, when the same radiologists performed the readout of cases, they were able to give an accurate diagnosis, regardless of their image-quality impression related to the display used.

## Acknowledgments

We thank the radiologists Diego Aguirre, Bibiana Pinzón, Oscar Rivero, Nelson Bedoya, José Vega, and Erickson Moreno, who performed the 2736 readings.

## References

- Hwang SA, Seo JB, Choi BK, et al. Liquid-crystal display monitors and cathode-ray tube monitors: a comparison of observer performance in the detection of small solitary pulmonary nodules. *Korean J Radiol* 2003; 4:153–156
- Kamitani T, Yabuuchi H, Soeda H, et al. Detection of masses and microcalcifications of breast cancer on digital mammograms: comparison among hard-copy film, 3-megapixel liquid crystal display (LCD) monitors and 5-megapixel LCD monitors: an observer performance study. *Eur Radiol* 2007; 17:1365–1371
- Langer S, Fetterly K, Mandrekar J, et al. ROC study of four LCD displays under typical medical center lighting conditions. *J Digit Imaging* 2006; 19:30–40
- Geijer H, Geijer M, Forsberg L, Kheddache S, Sund P. Comparison of color LCD and medical-grade monochrome LCD displays in diagnostic radiology. *J Digit Imaging* 2007; 20:114–121
- Bender S, Lederle K, Weiß C, Schoenberg S, Weisser G. 8-bit or 11-bit monochrome displays: which image is preferred by the radiologist? *Eur Radiol* 2011; 21:1088–1096
- Kimpe T, Tuyschaever T. Increasing the number of gray shades in medical display systems: how much is enough? *J Digit Imaging* 2007; 20:422–432
- Hiwasa T, Morishita J, Hatanaka S, Ohki M, Toyofuku F, Higashida Y. Need for liquid-crystal display monitors having the capability of rendering higher than 8 bits in display-bit depth. *Radiol*

- Phys Technol* 2009; 2:104–111
8. Abboud S, Weiss F, Siegel E, Jeudy J. TB or not TB: interreader and intrareader variability in screening diagnosis on an iPad versus a traditional display. *J Am Coll Radiol* 2013; 10:42–44
  9. National Electrical Manufacturer's Association. Digital imaging and communications in medicine (DICOM). Washington, DC: National Electrical Manufacturer's Association, 2001
  10. Fetterly KA, Blume HR, Flynn MJ, Samei E. Introduction to grayscale calibration and related aspects of medical imaging grade liquid crystal displays. *J Digit Imaging* 2008; 21:193–207
  11. Barten P. Contrast sensitivity of the human eye and its effects on image quality. *Knegsel, Holland: HP Press*, 1999
  12. Salazar AJ, Camacho JC, Aguirre DA, Ocampo J, Diaz XA. Diagnostic accuracy of digitized chest x-rays using consumer-grade color displays for low-cost teleradiology services. *Telemed JE Health* [Epub 2014 Feb 7]
  13. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006; 27:861–874
  14. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29–36
  15. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148:839–843
  16. Pepe MS. The statistical evaluation of medical tests for classification and prediction. *New York: Oxford University Press*, 2004
  17. Obuchowski NA. Sample size tables for receiver operating characteristic studies. *AJR* 2000; 175:603–608
  18. Eglin TK, Feinstein AR. Context bias: a problem in diagnostic radiology. *JAMA* 1996; 276:1752–1755
  19. Rhea JT, DeLuca SA, Greene RE. Determining the size of pneumothorax in the upright patient. *Radiology* 1982; 144:733–736
  20. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971; 76:378–382
  21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159–174
  22. Salazar AJ, Camacho JC, Aguirre DA. Comparison between different cost devices for digital capture of x-ray films with computed tomography (CT) correlation. *Telemed JE Health* 2011; 17:275–282
  23. Salazar AJ, Camacho JC, Aguirre DA. Agreement and reading-time assessment of differently priced devices for digital capture of X-ray films. *J Telemed Telecare* 2011; 18:82–85
  24. Salazar AJ, Camacho JC, Aguirre DA. Comparison between different cost devices for digital capture of x-ray films: an image characteristics detection approach. *J Digit Imaging* 2012; 25:91–100
  25. Society of Motion Picture and Television Engineers. Specifications for medical diagnostic imaging test pattern for television monitors and hard-copy recording cameras. *SMPTE J* 1986; 95:693–695
  26. Gray JE. Use of the SMPTE test pattern in picture archiving and communication systems. *J Digit Imaging* 1992; 5:54–58
  27. Gray JE, Lisk KG, Haddick DH, Harshbarger JH, Oosterhof A, Schwenker R. Test pattern for video displays and hard-copy cameras. *Radiology* 1985; 154:519–527
  28. Forsberg DA. Quality assurance in teleradiology. *Telemed J* 1995; 1:107–114
  29. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method. *Invest Radiol* 1992; 27:723–731
  30. Hillis SL, Berbaum KS. Monte Carlo validation of the Dorfman-Berbaum-Metz method using normalized pseudovalues and less data-based model simplification. *Acad Radiol* 2005; 12:1534–1541
  31. Roe CA, Metz CE. Dorfman-Berbaum-Metz method for statistical analysis of multireader, multimodality receiver operating characteristic data: validation with computer simulation. *Acad Radiol* 1997; 4:298–303
  32. Quenouille MH. Notes on bias in estimation. *Biometrika* 1956; 43:353–360
  33. Dorfman DD, Berbaum KS, Lenth RV, Chen YF, Donaghy BA. Monte Carlo validation of a multireader method for receiver operating characteristic discrete rating data: factorial experimental design. *Acad Radiol* 1998; 5:591–602
  34. Hillis SL, Berbaum KS. Power estimation for the Dorfman-Berbaum-Metz method. *Acad Radiol* 2004; 11:1260–1273
  35. Hillis SL. A comparison of denominator degrees of freedom methods for multiple observer ROC analysis. *Stat Med* 2007; 26:596–619
  36. Hillis SL, Berbaum KS, Metz CE. Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. *Acad Radiol* 2008; 15:647–661
  37. Hillis SL, Obuchowski NA, Schartz KM, Berbaum KS. A comparison of the Dorfman-Berbaum-Metz and Obuchowski-Rockette methods for receiver operating characteristic (ROC) data. *Stat Med* 2005; 24:1579–1607
  38. Metz CE, Pan X. "Proper" binormal ROC curves: theory and maximum-likelihood estimation. *J Math Psychol* 1999; 43:1–33
  39. Dorfman DD, Berbaum KS. A contaminated binormal model for ROC data. Part II. A formal model. *Acad Radiol* 2000; 7:427–437
  40. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol* 1989; 24:234–245
  41. Asai Y, Shintani Y, Yamaguchi M, Uemura M, Matsumoto M, Kanamori H. Evaluation of grey-scale standard display function as a calibration tool for diagnostic liquid crystal display monitors using psychophysical analysis. *Med Biol Eng Comput* 2005; 43:319–324
  42. Thompson SK, Willis CE, Krugh KT, Jeff Shepard S, McEnery KW. Implementing the DICOM grayscale standard display function for mixed hard- and soft-copy operations. *J Digit Imaging* 2002; 15:27–32
  43. Chakrabarti K, Kaczmarek RV, Thomas JA, Romanukha A. Effect of room illuminance on monitor black level luminance and monitor calibration. *J Digit Imaging* 2003; 16:350–355
  44. Jones DM. Utilization of DICOM GSDF to modify lookup tables for images acquired on film digitizers. *J Digit Imaging* 2006; 19:167–171
  45. Uemura M, Asai Y, Yamaguchi M, Fujita H, Shintani Y, Sanada S. Psychophysical evaluation of calibration curve for diagnostic LCD monitor. *Radiat Med* 2006; 24:653–658