

## An ROC Study of Chest Radiographs: 2K Versus 4K High-Resolution Soft-Copy Images

Su Young Kim, MD,<sup>1</sup> Yoon Joon Hwang,<sup>1</sup> Yoon Hee Han,<sup>1</sup> Jung Wook Seo,<sup>1</sup> Ji Young Lee,<sup>1</sup>  
Yong Hoon Kim,<sup>1</sup> Soon Joo Cha,<sup>1</sup> Gham Hur,<sup>1</sup> and Sung Soon Lee<sup>2</sup>

Computed radiography of chest with a 4K image array was recently introduced. We performed a multiobserver study to compare the diagnostic accuracy of 2K (standard) and 4K (high quality) chest radiographs displayed on a 5-mega-pixel monitor (2K monitor). One hundred cases of posteroanterior chest radiographs (a total of 200 images) were selected by two chest radiologists. Those radiographs included pneumothorax ( $n = 14$ ), nodules ( $n = 15$ ), interstitial disease ( $n = 10$ ), or neither abnormality ( $n = 61$ ). These were interpreted by four radiologists in two separate sessions. They recorded their confidence scale for the presence or absence of abnormality. Diagnostic accuracy was determined by receiver operating characteristic (ROC) analysis for each observer. ROC analysis showed no statistically significant difference between the 2K and 4K modes for the detection of any of the different abnormalities by individual readers. Our preliminary study suggests that 2K mode would be sufficient for the detection of abnormality on chest radiograph and there is no considerable validity to incline toward the 4K mode in current picture archiving and communication system environment using 2K monitor. However, we think that additional investigation using more subtle parenchymal or rib lesion should be followed.

**KEY WORDS:** Observer performance, ROC, computed radiography, PACS

### INTRODUCTION

Current digital technology provides various display formats for digitally acquired images. One of those various display formats, soft-copy display, allows digital imaging data to be managed and viewed electronically in picture archiving and communication system (PACS). In our hospital, all computed radiography (CR) images have

been obtained with 4K image arrays. The chest radiographs obtained with 2K image arrays are generalization for most departments in a PACS environment. Many studies comparing conventional film, laser-printed film, and high-resolution work station images of the chest in diagnostic accuracy have been performed.<sup>1-4</sup> Some experimental studies comparing the image quality of 2K (standard) chest CR images with 4K [high quality (HQ)] images have been reported.<sup>5,6</sup> One recent report revealed physical performance [(a) modulation transfer function—a convenient description of a system's resolution, (b) noise power spectra—the noise amplitude and texture, and (c) frequency-dependent detective quantum efficiency—efficiency of a system) of the 4K system was not significantly better than that of 2K.<sup>5</sup> To the best of our knowledge, there has been no study comparing 2K with 4K images displayed on high-resolution viewing stations in detecting chest abnormality. Considering cost effectiveness and access, it would be

<sup>1</sup>From the Department of Radiology, Ilsan Paik Hospital, Inje University School of Medicine, 2240, DaeHwa-Dong, Ilsan-GU, GoYang-Si, KyungGi-Do, 411-706, South Korea.

<sup>2</sup>From the Department of Internal Medicine, Ilsan Paik Hospital, Inje University School of Medicine, GoYang-Si, South Korea.

Correspondence to: Su Young Kim, MD, Department of Radiology, Ilsan Paik Hospital, Inje University School of Medicine, 2240, DaeHwa-Dong, Ilsan-GU, GoYang-Si, KyungGi-Do, 411-706, South Korea. tel: +82-31-9107390; fax: +82-31-9107369; e-mail: swiminkim@hanmail.net

Copyright © 2006 by SCAR (Society for Computer Applications in Radiology)

Online publication 27 December 2006

doi: 10.1007/s10278-006-1048-z

important to know whether the 4K images can actually achieve better diagnostic accuracy or not. To evaluate the diagnostic accuracy for each mode, we measured and compared the area of under receiver operating characteristic (ROC) curve for the detection of nodule, pneumothorax, and interstitial disease on 2K and 4K posteroanterior (PA) chest radiographs displayed on high-resolution viewing stations.

## MATERIALS AND METHODS

Our institutional review board approved the study protocol. Informed consent for the acquisition of additional images was obtained from 140 patients between January and December 2003. Chest radiographs with or without disease were collected over a period of 10 months in 2004. The initial set included approximately 140 cases, 280 radiographs, from which, 100 cases, 200 radiographs, were ultimately used in this study. Forty patients were subsequently excluded for one or more of the following reasons: poor or unacceptable image quality (21 cases), lesion too obvious (13 cases, e.g., densely calcified nodule or massive pneumothorax), and overabundance of a particular lesion (6 cases, e.g., too many images with nodule). Three classes of abnormalities—nodule, pneumothorax, and interstitial disease—were evaluated. The radiographs showed nodules ( $n=15$ ), pneumothorax ( $n=14$ ), interstitial disease ( $n=10$ ), or no abnormality ( $n=61$ ). Only two cases contained more than one abnormality (nodule and pneumothorax and nodule and interstitial disease). Before this ROC study, we performed paired comparison of 2K and 4K images displayed side-by-side on high-resolution viewing stations (2K monitors). A similar study with 2K and 4K laser-printed PA chest images was reported by Good et al.<sup>7</sup> Although all observers felt that the difference between 2K and 4K laser-printed images was not noticeable or the difference was extremely small, they could select 4k images as sharper or better images in significant cases. Our result was similar. Considering these results, we tried to select the cases with smaller or more subtle lesions.

Abnormal findings were verified independently by means of histologic (biopsy) reports or reports from other imaging modalities such as computed tomography (CT). Fifty cases of negative radiographs were confirmed by a low-dose cancer screening CT. The remainders of negative radiographs were verified primarily by documents such as negative radiologic and physical examination results obtained at least 10 months after the image was used in the study.

Both 2K and 4K chest PA radiographs were obtained from each patient in one sitting. At first, the 4K image was obtained and after switching the mode to 2K, the 2K image was acquired. However, it took about 5 min to switch the mode of system from 4K to 2K, so two consecutive images from one patient were not exactly the same. All radiographs were obtained with FCR-5501/FCR-5501-HQ CR (Fuji, Tokyo, Japan). The FCR-5501 is an integrated chest system. It contains two CR plates and two plates are used orderly, but we used one CR plate for two images in one patient by manual selection. Image parameters were as

follows: 130 Kvp, 1.25-mm nominal focus, 183-cm film-focus distance, 10:1 oscillating grid, and phototimed exposure. The default mode of imaging processing was used, including dynamic range compression, gradation enhancement, and edge enhancement. The image parameters were the same in both 2k and 4k modes. The specification of acquisition modes of FCR-5501-HQ system is given in Table 1.

Two chest radiologists who did not participate in the subsequent ROC study determined the presence or absence of the abnormality, as well as the visibility of the abnormality (from subtle to obvious) with unanimous agreement. Very easily detected cases and poor images were eliminated.

The viewing station employed in the study utilizes one 21-in grayscale monitor with portrait type (DATA RAY, Westminster, CO, USA). Their maximum resolution is  $2,048 \times 2,560$  pixels (so-called 2K monitor). The resolution of the graphic card is  $2,048 \times 2,560$  pixels. They have a maximum brightness level of  $120 \text{ cd/m}^2$ , with a screen refresh rate of 71 Hz. Brightness, contrast, and resolution of the monitor was tested routinely to ensure that the display quality did not degrade throughout the study. The ambient room light was dimmed.

Four board-certified radiologists read the two sets (2K and 4K) of 100 normal and abnormal chest radiographs displayed on the work station. All readers had a minimum of 1 year of experience in interpreting chest radiographs displayed on high-resolution viewing stations. Two modes of images were separately reviewed (eg, a set of 4K first, and then 2K), but the 100 images were displayed in a different order for each session. Each reader was provided the option to adjust brightness and contrast, zoom, and zoom reset for every image. In our viewing system, the 4K image is returned to its real, larger size on the 2K monitor when the image was double clicked and 2K images can be zoomed only. The observers were not forced to zoom the images, but they could determine whether the given image was 2K or 4K by using the zoom options. To minimize the possible learning effect, an 8-week interval existed between two interpreting sessions.

The time allowed for interpretation of each image was not restricted and the reading times spent per image were recorded. The clinical information was not given and the reader did not know the proportion of cases containing abnormalities. For each ROC study, the reader was informed that the images might contain pneumothorax, interstitial disease, or nodule. Correct diagnoses of the images were not revealed to the observers until they had reviewed both sets of images. During interpretation, each reader reported diagnosis and confidence level by filling out the ROC form. For each category (nodule, pneumothorax, and interstitial disease), the following five-point confidence level rating scale was used, where 0 is definitely absent, 1 is probably absent, 2 is possibly present, 3 is probably present and 4 is definitely present.

**Table 1. Specification of Acquisition Modes of FCR-5501 and FCR-5501-HQ System**

Image Size (cm <sup>2</sup> )	Mode	Digital Matrix	Pixel Size (μm)	Limiting Frequency (cycles/mm)	Image Size (Mb)
35 × 43	HQ	3,520 × 4,280	100	5.0	30.1
35 × 43	ST	1,760 × 2,140	200	2.5	7.5

**Table 2. Area Under ROC Curve (Az) and Level of Significance for 2K and 4K Image Displayed on High-resolution Viewing Station**

Reader No.	A(z) (SD)		Significance (P Value)	95% Confidence Intervals for Area Difference
	2K	4K		
<b>Pneumothorax</b>				
1	0.961 (0.037)	0.962 (0.037)	0.964	(-0.042, 0.044)
2	0.962 (0.037)	0.964 (0.036)	0.962	(-0.079, 0.083)
3	0.927 (0.050)	0.960 (0.038)	0.245	(-0.023, 0.090)
4	0.964 (0.036)	0.964 (0.036)	1.000	(-0.080, 0.080)
Average	0.954 (0.018)	0.963 (0.002)		
<b>Nodule</b>				
1	0.895 (0.057)	0.911 (0.026)	0.541	(-0.036, 0.068)
2	0.919 (0.051)	0.915 (0.052)	0.792	(-0.024, 0.032)
3	0.856 (0.066)	0.864 (0.064)	0.788	(-0.048, 0.063)
4	0.878 (0.063)	0.887 (0.061)	0.715	(-0.039, 0.057)
Average	0.887 (0.267)	0.894 (0.024)		
<b>Interstitial disease</b>				
1	0.867 (0.076)	0.865 (0.077)	0.939	(-0.062, 0.067)
2	0.862 (0.077)	0.864 (0.077)	0.746	(-0.006, 0.009)
3	0.860 (0.078)	0.865 (0.077)	0.904	(-0.077, 0.087)
4	0.882 (0.072)	0.880 (0.073)	0.932	(-0.005, 0.060)
Average	0.867 (0.010)	0.870 (0.011)		

Numbers in parentheses are SDs. Average value of significance and 95% confidence intervals were not calculated.

Statistical evaluation was done by MedCalc version 6.00 containing ROC program (available at <http://www.medcalc.be/manual/roccurves.php>, accessed December 2004). The performance was measured by the area under the ROC curve (Az) and standard deviation was constructed separately for each mode (2K or 4K), for each disease, and for each reader. The program assumed a *binormal* model and produced *P* values to assess the significance of the difference between the two areas for each reader. The paired *t* test was also performed for the comparison of the average ROC area over all readers by disease category.

## RESULTS

The results of the ROC study for the four observers and three disease categories are summarized in Table 2, including areas under ROC curves (which can be used as an index for diagnostic accuracy), standard deviations, and *P* values. We calculated average Az. Table 3 shows the results of the paired *t* statistics and *P* values by disease category.

**Table 3. Paired *t* Statistics and *P* Value by Disease Category**

Disease	<i>t</i> Value	<i>P</i> Value
Pneumothorax	1.12	0.34
Nodule	1.75	0.18
Interstitial disease	1.51	0.23

In the detection of pneumothorax, One showed slight improvement. The other three were essentially of equal performance. For the detection of nodule, although not statistically significant (*P*=0.18), the average results for three observers slightly favored the 4K images. One observer performed slightly better in the 2K set. In the detection of interstitial disease, two of four observers showed slightly better performance in the 2K set, the other two observers performed at a higher level in the 4K set. There was also no statistical significance (*P*=0.23). Only one observer showed better performance in 4K set for the detection of any abnormality.

**Table 4. Average Reading Times for 2K and 4K Image**

Reader	2K (Time in Seconds)	4K (Time in Seconds)	Time Difference (Seconds)
1	61.5	63.7	2.2
2	58.9	60.5	1.6
3	60.5	61.9	1.4
4	57.4	58.7	1.3
Average	59.6 (1.80)	61.2 (2.12)	1.62
			<i>t</i> = 8.06 <sup>a</sup>
			<i>P</i> = 0.004

Standard deviations are in parentheses.

<sup>a</sup>Paired *t* test.

Tables 2 and 3 show the comparable performance level of four radiologists with both 2K and 4K images. No significant statistical differences were measured for any observer, for any abnormality or in the overall performance of the group (*t* test).

The average reading time spent per image is given in Table 4. The overall mean time required for displaying, reviewing, interpreting, and rating the cases was longer when 4k images were reviewed. Time for 2K is 59.6 s and time for 4K is 61.2 s. The time difference in the overall mean time is 1.62 s. There was a statistically significant difference ( $P=0.004$ ) in overall mean reading time between 2K and 4K. We can presume that the increased time in 4K images may be related to the longer display time (the average time required for displaying is 2.2 s for 2K images and 3.9 s for 4K images, the time difference in displaying is 1.7 s), but we are not convinced that the readers did not look longer and harder with the knowledge that they were looking at a 4K image. This does introduce a bias.

## DISCUSSION

Chest radiography with digital storage phosphor technology, known as CR, is usually performed using a 200- $\mu\text{m}$  pixel size producing an image array of 2K for a standard 14-by-17-in radiograph. Recently, a CR system for chest radiographs with a pixel size of 100  $\mu\text{m}$  in a 4K image array was introduced. It might be assumed that increasing the sampling rate (or reducing pixel size) could lead to achieving a higher resolution with an adverse effect on noise. However, a previous study by Fujita et al.<sup>6</sup> has reported that reducing the pixel size does not proportionally improve the resolution, suggesting that the intrinsic presampled transfer characteristics of the system may not be sufficient to substantiate the resolution dictated by the sampling rate. The study solving the question of whether the 4K system provides an improved image quality was done by Flynn and Samei.<sup>5</sup> The results are indicated as follows: (1) the 4K system is slightly better in noise characteristics, (2) 4K offers slightly better signal-to-noise characteristics, (3) the resolution response of 4K is inferior to that of 2K, and (4) overall the physical performance of the 4K system is not significantly better than that of 2K. The study by Miro et al.<sup>8</sup> showed equal

observer performance in the detection of various chest abnormalities on both 2K and 4K hard-copy film of computed radiographs. Another recent study by Tagashi et al.<sup>9</sup> has reported that diagnostic accuracy in the detection of subtle interstitial abnormalities on 4K full-sized film of digital storage phosphor chest radiographs was equivalent to that on 2K. Our study using 2K and 4K soft-copy images showed similar results in the detection of chest abnormalities. However, additional investigations using more subtle parenchymal or rib lesions on chest radiograph should be evaluated.

We emphasize here two serious criticisms of this study. The observers were obviously biased because they knew from the work station whether they were dealing with a 2K or 4K image. The other limitation is the low number of samples, which can lead to weak conclusions. The overall performance of our observers for the detection of pneumothorax, nodule, and interstitial disease was nearly equivalent in two modes and in respect of cost effectiveness and easier access; we may conclude that the 2K mode would be sufficient for the detection of abnormalities on chest radiograph and there is no considerable validity to incline toward the 4K mode in current PACS environment using a 2K monitor.

## ACKNOWLEDGEMENT

This article is supported by a 2004 scientific research grant of Inje University.

## REFERENCES

1. Slasky BS, Gur D, Good WF, et al: Receiver operating characteristic analysis of chest image interpretation with conventional, Laser-printed, and high-resolution workstation images. *Radiology* 174:775–780, 1990
2. Hayrapetian A, Aberle DR, Huang HK, et al: Comparison of 2048-line digital display formats in conventional radiographs: an ROC study. *AJR Am J Roentgenol* 152:1113–1118, 1989
3. Cox GG, Cook LT, McMillan JH, et al: Chest radiography: Comparison of high resolution digital displays with conventional and digital film. *Radiology* 176:771–776, 1990
4. Fajardo LL, Hillman BJ, Pond GD, et al: Detection of pneumothorax: Comparison of digital and conventional chest imaging. *AJR Am J Roentgenol* 152:475–480, 1989
5. Flynn MJ, Samei E: Experimental comparison of noise and resolution for 2k and 4k storage phosphor radiography systems. *Med Phys* 26(8):1612–1623, 1999

6. Fujita H, Morishita J, Ueda K, Tsai DY, Ohtsuka A, Fujikawa T: Resolution properties of a computed radiographic system. *SPIE Med Imaging* 1090:263–275, 1989
7. Good WF, Gur D, Feist JH, et al: Subjective and objective assessment of image quality—a comparison. *J Digit Imaging* 7(2):77–78, 1994
8. Miro SPM, Leung AN, Rubin GD, et al: Digital storage phosphor chest radiography; An ROC study of the effect of 2K versus 4K matrix size on observer performance. *Radiology* 218:527–532, 2001
9. Tagashi U, Takeshi J, Noriyuki T, et al: Full size digital storage phosphor chest radiography; effect of 2K versus 4K matrix size on observer performance in detection of subtle interstitial abnormalities. *Radiat Med* 23(3):170–174, 2005