Fracture Interpretation Using Electronic Presentation: A Comparison

Laura W. Bancroft, Thomas H. Berquist, Richard L. Morin, Jerald H. Pietan,
John M. Knudsen, and Hugh J. Williams, Jr

The purposes of this study were to determine whether (1) fractures are interpreted differently after digitization and electronic presentation; (2) there are differences in accuracy between screen radiographs and electronic presentation; (3) differences in interpretation are a function of monitor resolution; and (4) differences in interpretation between radiographs and electronic images relate to radiological subspecialty. Forty cases with fractures of varying degrees of subtlety and 35 cases without fractures were interpreted. Radiographs were digitized with 2 different systems and displayed on 3 monitors of different spatial resolution. Four radiologists, with varying experience, were asked to decide whether fractures were present, absent, or they were uncertain. Accuracy of interpretation increased with improved electronic image presentation and monitor resolution. The sensitivity, specificity, and accuracy of fracture detection on System A were 63%, 98%, and 78%, respectively. The results were 72%, 98%, and 84% with System B. System C results were 81%, 97%, and 88% with Lumiscan 75, and 82%, 96%, and 88% with Lumiscan 150. Sensitivity, specificity, and accuracy results of the original radiograph interpretation were 89%, 95%, and 92%. Results were significantly different for System A. No significant differences were found for the other systems compared with film radiographs. System A did not have adequate monitors for interpretation of subtle fractures. Systems B and C were capable of displaying even subtle fractures. Our initial results indicate that interpretation with high-quality 1K × 1K monitors is substantially similar to radiograph interpretation.

Copyright © 2000 by W.B. Saunders Company

KEY WORDS: diagnostic radiology, observer performance, fractures, images, interpretation, display, 40.41, picture archiving and communication systems.

AS THE PRACTICE of radiology progresses toward a completely "filmless" department, studies have been undertaken for the purpose of evaluating the interpretative accuracy of cases using various digitizers, monitors, and display systems (Table 1). The purpose of this study was to determine whether conventional radiographic images of skeletal structures, specifically fractures, are interpreted differently after digitization and subsequent electronic presentation providing grayscale reversal, window-level adjustment, and magnification. Furthermore, this study assessed differences in diagnostic accuracy for conventional screen

film radiographs versus electronic presentation. Finally, differences in image interpretation as a function of monitor resolution and radiology subspecialties were evaluated.

MATERIALS AND METHODS

Forty cases with clinically proven fractures and 35 cases without fractures were selected for digitization. Twenty-eight cases were subtle (Fig 1) and 12 were more obvious (Fig 2). The cases were selected from both pediatric and adult hospital radiology libraries, involved all anatomic regions, and were prospectively chosen to be particularly challenging.

The radiographs were first digitized with the Lumiscan 75 (Lumisys, Sunnyvale, CA), which has a helium-neon laser spot size of 200 μ m, 2,048 pixels/line, and a spatial resolution of 5 pixels/mm (Table 2). There was a fixed pixel size for Lumiscan 75 acquisitions; therefore, not all potential pixels were used. These images were transmitted to each of 3 systems. System A consisted of a 640 × 480-pixel Apple 1710 (Apple Computer Corp, Cupertino, CA) 16" × 17" monitor using proprietary 5:1 compression/decompression. System B consisted of a 1,200 × 1,600-pixel Data Ray (Cemax/Icon, Westminster, CO) 12.5" × 16" monitor without compression. System C consisted of a 1,280 × 1,024-pixel Magic View 1002 (Siemens Medical System, Iselin, NJ) 19" × 19" monitor using 2:1 lossless compression (Table 3), via our Sienet (Siemens Medical System, Iselin, NJ) picture archiving and communication system (PACS).

Finally, the radiographs were digitized with the Lumiscan 150 (Lumisys, Sunnyvale, CA) and transmitted to the Magic View 1002 viewing station. The Lumiscan 150 has a helium-neon laser spot size of 50 μ m, 4,096 pixels/line, and a spatial resolution of 20 pixels/mm (Table 2). In contrast, this digitizer does not have a fixed pixel size. Both the 75 and 150 digitizers have a grayscale capture of 10 bits. A linear look-up table was used to map the 10 bits of digitized information to 8 bits of grayscale for each of the 4 monitors, yielding 256 shades of gray.

All readers were instructed to use each of the system functions required for the study. Contrast and brightness, grayscale reversal, and magnification were available on all 3 systems. Grayscale mapping on Systems B and C, and inverse polarity (reversal) and twofold zoom (magnification) on all systems were mouse-driven.

From the Department of Radiology, Mayo Clinic Jacksonville, Jacksonville, FL.

Address reprint requests to Laura W. Bancroft, MD, Department of Radiology, Mayo Clinic Jacksonville, 4500 San Pablo Rd, Jacksonville, FL 32224.

*Copyright © 2000 by W.B. Saunders Company
0897-1889/00/1301-0003\$10.00/0*

Table 1. Interpretation Systems

System A (Apple 1710) and Lumiscan 75	Interpretation 1
System B (Data Ray) and Lumiscan 75	Interpretation 2
System C (Siemens Magic View 1002) and Lumiscan 75	Interpretation 3
System C (Siemens Magic View 1002) and Lumiscan 150	Interpretation 4
Radiograph	Interpretation 5

Four board-certified radiologists were asked to interpret each case, utilizing 5 different systems (Table 1). One radiologist was a subspecialized musculoskeletal radiologist, 1 had some training in musculoskeletal radiology, and the other 2 were general radiologists with interest in other subspecialties. The cases were presented without any clinical history, in randomized and various orders, and with at least 1 month elapsing between readings on different modalities. The cases were read sequentially, from the lowest to the highest resolution systems, to the original radiographs. Interpretations were made in the normal reading environment.

A case sheet was completed by the radiologist after interpretation and dictation of an examination. The radiologists indicated the diagnosis of "fracture," "no fracture," or "uncertain." They indicated whether they had used grayscale reversal, window and level adjustments, and/or magnification. The case interpretations were timed from either the moment the radiograph was removed from the jacket or the time the image was selected on the monitor to the end of the full dictation.

Sensitivities, specificities, and interpretation times were evaluated. A consensus panel (consisting of the readers) reviewed every misdiagnosed or "uncertain" case after the study. Sources of interpretative error using each modality were sought. Missed fracture cases were analyzed retrospectively on each monitor,

Fig 1. Subtle, nondisplaced right ischial fracture involving the inferior cortex (arrow). This was not detected by any of the 4 readers by either electronic or conventional interpretation.

Fig 2. Obvious, displaced left intertrochanteric fracture (arrow). This was universally detected by all readers on all modalities, and represents the most obvious fracture of the 75 cases.

after selective manipulation of window and levels, grayscale reversal, and magnification.

Of the 75 patients, 28 had subtle fractures, 12 had more obvious fractures, and 35 had no fractures. The distribution of fractures is shown in Table 4.

RESULTS

Entries designated "uncertain" were considered missed diagnoses, whether or not the dictations

Table 2. Digitizer Specifications

Property	Lumiscan 75	Lumiscan 150
Pixels/line	2,048	4,096
Line width	8" to 14"	8" to 14"
Maximum line resolution	5 pixels/mm	20 pixels/mm
Spot size	200 μ m	50 μ m
No. of lines	1,250	1,250
Scan rate	115 lines/s	75 lines/s
Density resolution	0.001 OD	0.001 OD

Table 3. Monitor Specifications

Property	System A: Apple 1710	System B: Data Ray	System C: Magic View
Pixels/line (horizontal)	640	1,200	1,280
No. of lines (vertical)	480	1,600	1,024
Horizontal width	16"	12.5"	19"
Vertical height	17"	16"	19"
Maximum brightness			
Foot-lamberts	25	—	175
Contrast ratio	—	—	1:500

suggested the correct diagnosis. The average number of uncertain diagnoses was 12 for System A and 8 for Systems B and C and the original film radiographs.

Sensitivity and specificity values were determined from the remaining entries, when "fracture" or "no fracture" were chosen. Accuracy rates were derived from the summation of true-positives and true-negatives divided by the total number of cases (75). The sensitivity, specificity, and accuracy rates for System A were 63%, 98%, and 78%, respectively, for all four radiologists (Tables 5 and 6). The results were 72%, 98%, and 84%, respectively, with System B. The results with System C averaged 82%, 97%, and 88%, without a statistically significant difference between the Lumiscan 75 and 150 scanners. Sensitivity, specificity, and accuracy of the interpretations of the original radiographs were 89%, 95%, and 92%, respectively.

The sensitivity, specificity, and accuracy results for the electronic systems were compared with those of the film radiographs. A summary of the statistical analysis using the *z* approximation is

Table 4. Distribution of Fractures

Location	No. of Cases
Skull	1
Clavicle	2
Humerus	1
Ulna	2
Carpus	2
Metacarpal/finger	10
Ribs	2
Lumbar spine	2
Pelvis	4
Femur	4
Patella	1
Tibia	2
Fibula	2
Metatarsal/toe	2

provided in Table 7. There were no significant differences between System C results and the radiograph results. Significant differences were observed with Systems A and B.

Consistent errors were made among the four radiologists. Subtle fractures involving the lumbar spine transverse processes were universally missed on Systems A and B (Fig 3). The fractures were correctly diagnosed 50% of the time on System C and 75% on the radiographs. Errors were attributed to the presence of overlying bowel gas and the cumbersome method of magnifying and adjusting window/level for each portion of the electronic image. Subtle phalangeal fractures were missed 39% of the time with System A, 33% with System B, 29% with System C, and 14% of the time when reading the original screen-film radiographs (Fig 4).

Window and level adjustment was used by the 4 radiologists an average of 89% of the time. Magnification and grayscale reversal options were used 68% and 40% of the time. Usage varied widely among readers. As expected, the greater the utilization of the various manipulation tools, the longer the time needed to interpret cases. The average time required per case was 1 minute 33 seconds on System A, 1 minute 30 seconds on System B, 1 minute 24 seconds on System C, and 1 minute 1 second using the radiograph (Table 8). For Systems A, B, and C (respectively), the interpretation times were 52%, 48%, and 38% longer than with screen-film studies. The discrepant times for System A interpretations were attributed to certain readers "admitting defeat" on the lower-resolution System A and progressing to the next case, whereas other readers deliberated over subtle cases. The difference in interpretation times among System C readings was attributed to familiarity with the system. Certain readers routinely use this system in daily practice, whereas others use it only occasionally.

Comments were invited for each case. The majority of readers disliked interpreting facial, rib, lumbar spine, sacral, and coccygeal cases from the electronic monitors. This was related to the cumbersome method of magnifying and adjusting windows and levels for each portion of the electronic image. The majority of readers also mentioned the increased time required to interpret electronic images as opposed to conventional ones.

Table 5. Diagnostic Accuracy

System A: Apple 1710													
	Time	TP	FP	TN	FN	Sens	Spec	Acc	F/NF (n)	Unc (n)	W/L	Mag	Inv
Rd 1	0:01.10	22	1	29	12	65%	97%	80%	64	11	100%	83%	29%
Rd 2	0:02.10	16	0	28	23	41%	100%	66%	67	8	89%	29%	96%
Rd 3	0:01.08	22	0	23	12	65%	100%	79%	57	18	84%	72%	0%
Rd 4	0:01.45	29	1	29	7	81%	97%	88%	66	9	99%	99%	4%
Avg	0:01.33	22	1	27	14	63%	98%	78%	64	12	93%	71%	32%
System B: Data Ray													
	Time	TP	FP	TN	FN	Sens	Spec	Acc	F/NF (n)	Unc (n)	W/L	Mag	Inv
Rd 1	0:01.00	31	1	27	8	79%	96%	87%	67	8	97%	64%	33%
Rd 2	0:01.39	19	1	31	17	53%	97%	74%	68	7	89%	24%	77%
Rd 3	0:01.15	27	0	28	10	73%	100%	85%	65	10	97%	83%	85%
Rd 4	0:02.04	32	1	31	6	84%	97%	90%	70	5	100%	67%	0%
Avg	0:01.30	27	1	29	10	72%	98%	84%	68	8	96%	59%	49%
System C: Magic View Lumiscan 75													
	Time	TP	FP	TN	FN	Sens	Spec	Acc	F/NF (n)	Unc (n)	W/L	Mag	Inv
Rd 1	0:00.49	33	1	28	5	87%	97%	91%	67	8	59%	51%	40%
Rd 2	0:01.34	23	0	31	15	61%	100%	78%	69	6	85%	64%	63%
Rd 3	0:01.13	31	1	27	8	79%	96%	87%	67	8	92%	77%	91%
Rd 4	0:01.45	37	2	28	1	97%	93%	96%	68	7	100%	86%	0%
Avg	0:01.21	31	1	29	7	81%	97%	88%	68	7	84%	70%	48%
Lumiscan 150													
	Time	TP	FP	TN	FN	Sens	Spec	Acc	F/NF (n)	Unc (n)	W/L	Mag	Inv
Rd 1	0:00.48	29	1	28	9	76%	97%	85%	67	8	39%	61%	47%
Rd 2	0:01.34	26	0	29	14	65%	100%	80%	69	6	93%	28%	12%
Rd 3	0:01.15	30	1	23	2	94%	96%	95%	56	19	99%	93%	61%
Rd 4	0:02.11	37	3	29	3	93%	91%	92%	72	3	99%	99%	0%
Avg	0:01.27	31	1	27	7	82%	96%	88%	66	9	82%	70%	30%
Original Radiograph													
	Time	TP	FP	TN	FN	Sens	Spec	Acc	F/NF (n)	Unc (n)			
Rd 1	0:00.40	32	1	30	2	94%	97%	95%	65	10			
Rd 2	0:01.19	28	0	31	10	74%	100%	86%	69	6			
Rd 3	0:00.57	36	3	25	2	95%	89%	92%	66	9			
Rd 4	0:01.09	38	2	28	2	95%	93%	94%	70	5			
Avg	0:01.01	34	2	29	4	89%	95%	92%	68	8			

Abbreviations: TP, true positive; FP, false positive; TN, true negative; FN, false negative; Sens, sensitivity; Spec, specificity; Acc, accuracy; F/NF (n), number of "fracture"/"no fracture" diagnoses; Unc (n), number of "uncertain" diagnoses; Rd, reader; Avg, average among the 4 readers.

DISCUSSION

The results of other studies encourage caution regarding the comparison of radiographic and electronic interpretations of fractures.¹ A study by the group at Mallinckrodt demonstrated that overall performance with their digital system was inferior to that with original radiographs in the evaluation of fractures.² Although our study detected a difference in interpretation of digitized electronic images

Table 6. Summary of Diagnostic Accuracy

	Sensitivity	Specificity	Accuracy
System A: Apple 1710	63%	98%	78%
System B: Data Ray	72%	98%	84%
System C: Magic View—Average	82%	97%	88%
and Lumiscan 75	81%	97%	88%
and Lumiscan 150	82%	96%	88%
Radiograph	89%	95%	92%

Table 7. Statistical Analysis of Results

	System A	System B	System C
Sensitivity	$P < .005$	$P < .01$	NS
Specificity	NS	NS	NS
Accuracy	$P < .05$	NS	NS

NOTE. Each value in the table was tested individually against the result for the original film radiographs.

Abbreviation: NS, no significant difference.

from screen-film images, our results indicate that interpretation with high-quality $1K \times 1K$ monitors is substantially similar to film radiograph interpretation. The overall fracture detection accuracies among the 4 readers were 78%, 84%, and 88% for Systems A, B, and C, respectively, compared with 92% for conventional radiographs. The reasons for the differences are multifactorial. They relate to the spatial resolution of different monitors and digitizers, unfamiliarity with viewing fractures on electronic systems as opposed to years of experience with radiographs, and the need for longer interpretation times because of the use of tools. It is difficult to evaluate how much effect the 5:1 lossy compression used in System A had on this study.

Other studies in which digitized images were interpreted have demonstrated that nondisplaced fractures are difficult to detect with pixel sizes greater than 0.16 mm.³ The output of the Lumiscan 75 digitizer at 0.2 mm and Lumiscan 150 digitizer at 0.05 mm showed no significant differences that affected the interpretation of subtle fractures. On retrospective review, however, the information enabling a correct diagnosis was present on both Systems B and C. Thus, diagnostic errors are attributable to interpretive error rather than a technical deficiency of the monitors. With the aid of window/level adjustments, grayscale reversal, and magnification on Systems B and C, all fractures in this study were visible. Therefore, the high rates of observer error and interobserver variability involved with interpretation of radiological studies far outweigh any shortcomings of the high-resolution monitor systems.

Utilization of any or all the imaging tools at the expense of time may have been a factor in interpretation differences. Although each electronic image appeared identically on each reader's monitor, the individual reader adjusted the image according to his own preference and speed. Other studies

Fig 3. Minimally displaced fracture of the left transverse process of the second lumbar vertebra. (A) Overlying bowel gas makes evaluation of subtle fractures difficult for both conventional and electronic interpretation. (B) The magnified image better demonstrates the subtle transverse process fracture (arrow). The wide range of optical densities of different portions of the image requires constant adjustment of windows and levels by the radiologist. This fracture escaped detection by all readers on the Apple and Data Ray systems, 2 readers on the Magic View systems, and 1 reader on screen-film.

Fig 4. Subtle, nondisplaced fracture of the first distal phalanx (arrow). This fracture eluded all readers using the Apple and Data Ray systems. One reader confidently identified it on the Magic View systems using both the Lumiscan 75 and 150 digitizers. Half the readers made a correct diagnosis from the original radiograph.

have described an interpretation time with on-screen images of 1.54 times longer than that with radiographs.⁴ This study concluded an increase of 38% to 52% over standard screen-film radio-

Table 8. Interpretation Time Increase (Electronic Versus Film)

	Reader 1	Reader 2	Reader 3	Reader 4	Average
System A	75%	65%	19%	52%	52%
System B	50%	25%	32%	78%	48%
System C—Average	22%	19%	21%	43%	38%
and Lumiscan 75	23%	19%	28%	52%	33%
and Lumiscan 150	20%	19%	14%	33%	43%

graphs interpretation times, depending on the system. In daily practice with electronic imaging, the cost/benefit ratio of such manipulations can be judged individually to balance accuracy with efficiency.

Despite at least 1 month elapsing between case reading sessions, readers relayed that certain cases were so memorable that they would have required at least 1 year between various modality readings to lose familiarity. Finally, the differences in image interpretation between radiographs and electronic images were not felt to be related to radiological subspecialty. All radiologists were able to detect fractures.

ACKNOWLEDGMENT

The authors thank Dr Martin Blum for his selection of the cases.

REFERENCES

1. Scott WW, Bluemke DA, Mysko WK, et al: Interpretation of emergency department radiographs by radiologists and emergency medicine physicians: Teleradiology workstations versus radiograph readings. *Radiology* 195:223-229, 1995
2. Wilson AJ, Hodge JC: Digitized radiographs in skeletal trauma: A performance comparison between a digital workstation and the original film images. *Radiology* 196:565-568, 1995
3. Murphy MD, Bramble JM, Cook LT, et al: Nondisplaced fractures: Spatial resolution requirements for detection with digital skeletal imaging. *Radiology* 174:865-870, 1990
4. Scott WW, Rosenbaum JE, Ackerman SJ, et al: Subtle orthopaedic fractures: Teleradiology workstation versus film interpretation. *Radiology* 187:811-815, 1993