Teleradiology for Remote Diagnosis: A Prospective Multi-Year Evaluation

M. Elon Gale, Miriam E. Vincent, and Alan H. Robbins

Teleradiology has been used for nearly 3 years at our institution to provide urgent radiologic interpretations for two outpatient clinics and an affiliated hospital. The purpose of this study was to evaluate the clinical reliability of the existing system. Teleradiology images were interpreted using 1600 × 1200 pixel display stations. The original films from the same cases were subsequently interpreted, usually by another radiologist. The initial and final interpretations were compared. Discrepancies were rated and adjudicated by another senior radiologist. These data were compared to peer review interobserver discrepancy rates. Among the 2688 teleradiology examinations evaluated, there were major discrepancies in 31 (1.15%). In three instances teleradiology rather than film interpretation was considered correct. Abnormalities missed on teleradiology were apparent in all but two at adjudication. Among the 628 peer-review cases, there were 6 (0.96%) major discrepancies. Major teleradiology discrepancy rates are statistically similar to film-based peer review discrepancy rates. Teleradiology is suitable for providing radiologic services to remote medical facilities.

Copyright © 1997 by W.B. Saunders Company

KEY WORDS: teleradiology, peer review.

EARLY TELERADIOLOGY systems based on video cameras have been largely replaced by digitally acquired or digitally scanned radiologic images that are transmitted at ever-increasing speeds over continually improving communication lines.^{1,2} These advances in technology have made it possible for our department to provide radiology services to two affiliated outpatient clinics and to two associated hospitals.

A number of studies have been performed to evaluate digital display systems and to compare them with standard radiography.³⁻⁵ The purpose of the present study was to evaluate the clinical reliability of a teleradiology system in practical circumstances across a range of diagnostic entities. The study evaluated the discrepancy rates in two parallel radiologic interpretation and second review

systems, teleradiology and peer-review. We hypothesized that interpretive error rates when performing primary readings from soft-copy images and interpretive error rates when performing primary readings from film would be statistically similar, and that from this similarity it could be concluded that primary soft-copy interpretation could be used with confidence as an alternate or substitute for primary film interpretation.

MATERIALS AND METHODS

Radiographic images were digitally scanned at the remote sites with a Lumisys 200 laser scanner (2500 × 2000 lines) (Sunnyvale, CA) and transmitted to our medical center using a Kodak-Vortech (Rochester, NY) teleradiology system that used a MacIntosh computer platform. Images were transmitted across T1 communication lines. Typical transmission time for a single chest radiograph was approximately 3.5 to 4.0 minutes, with a screen paint time of approximately 15 to 20 seconds. Viewing consoles consisted of two 1600 × 1200 pixel display monitors with mouse-activated image manipulation including flip, rotate, magnification, window, level, etc. Data collection occurred over 32 months from one outpatient clinic and the affiliated hospital and over 23 months from the other outpatient clinic.

Selection of cases to be sent via teleradiology was determined by the requesting physician or nurse practitioner based on clinical need prior to review by the requestor. Most transmitted cases represented a stat or urgent need for interpretation. Routine cases were not sent except when used for system testing. The transmitted images contained plain, noncontrast radiographic studies, typically chest, bone, and abdomen films. The requisition for each case was sent by facsimile at the time of image transmission. An interpretation was then made using the teleradiology system. The summarized results were handwritten on the faxed requisition and faxed back to the sending site. Significant results were also telephoned to the referring clinician. In most cases, the interpretations were made by staff

From the Department of Radiology, Department of Veterans Affairs Medical Center, Boston, MA.

Address reprint requests to M. Elon Gale, MD, Department of Radiology (114), Department of Veterans Affairs Medical Center, 150 S Huntington Ave, Boston, MA 02130.

*Copyright © 1997 by W.B. Saunders Company
0897-1889/97/1002-0003\$3.00/0*

radiologists alone, or in a few cases, by both a resident and a staff member.

Within 24 hours all original films that had been sent via teleradiology, as well as all routine films, were brought by courier for final interpretation. In most instances the final interpretation was rendered by a different radiologist. A few radiologists previewed the initial interpretation before rendering the final interpretation, although most did not. At the time of final interpretation, the results of the final interpretation were compared with the interpretation given on the faxed requisition for the teleradiology images.

The comparison was rated on a three-tiered scale: No significant difference, minor discrepancy, major discrepancy. Cases that were included in the first category were those in which the findings were essentially the same on both the handwritten fax and the final report. Cases in the second group were those in which there were small differences in interpretation and in which the discrepancy would have no impact on clinical management, eg, old rib fractures, spinal osteophytes, etc. that may well have been noted, but not recorded in the short handwritten faxed report. Cases in the third group were those in which there was a substantive difference that might have an impact on the clinical management of the patient, eg, pneumonia, acute fracture, congestive heart failure, etc. All cases that were rated as having a major discrepancy were reviewed by one or two other senior staff members for adjudication.

In order to determine a normative and comparable standard, a subset of the ongoing departmental peer review data was evaluated. This subset included data that were restricted to the same 32-month teleradiology data collection period. Reviews for each staff member were performed on a random selection of 10 to 15 cases per quarter. Each peer reviewed case was evaluated on the same three-tiered scale as the teleradiology review: No significant difference, minor discrepancy, major discrepancy. Because peer review cases were those in which a final interpretation had already been rendered, the language of the report as well as its contents were evaluated. Only peer review data for the 10 radiologists who provided teleradiology interpretations was included. Specific quarterly peer reviews were excluded when those reviews consisted of subspecialty case material, such as angiography or neuroradiology, so that only similar types of general radiology examinations would be included in the comparison with teleradiology.

To assess the prevalence of abnormalities in the transmitted cases, a random selection of teleradiology reports was evaluated by one of the investigators (M.E.G.) and rated according the following four categories: New major abnormality, known major abnormality, minor abnormality, and no abnormality. Major abnormalities were determined to be new or known based on the clinical history on the accompanying requisition.

RESULTS

During the entire study period, there were a total of 2688 transmitted cases, 2563 (95.4%) having no discrepancy, 94 (3.5%) having minor discrepancies, and 31 (1.15%) showing a major discrepancy. Among the 628 in-house, peer reviewed cases, there were 525 (83.6%) with no discrepancy, 97 (15.4%) with minor discrepancies, and 6 (0.95%)

with a major discrepancy. To compare the relative discrepancy rates, the data in the no discrepancy and major discrepancy categories were submitted to a χ^2 analysis (Table 1). The differences in major discrepancy rates between teleradiology and film were found to be nonsignificant ($\chi^2 = 0.1814$ $P < .80$; $df = 1$). A power analysis of the χ^2 established that the power was equal to 0.94. This indicates with a very high probability that the absence of significance was attributable to a real nondifference in these rates, ie, that the rates of major discrepancy were equivalent in these two conditions.

Of the transmitted cases, 661 were from the outpatient clinics and 2027 were from the associated hospital. The number of minor discrepancies was 43 (6.5%) and 51 (2.5%) respectively. The number of major discrepancies was 10 (1.5%) and 21 (1.0%) respectively. Table 2 lists all the specific major discrepancies identified. In 3 cases, a LLL infiltrate, a retrosternal nodule, and a parenchymal nodule, the teleradiology interpretation was adjudicated to be correct and the film interpretation incorrect. Twenty eight represented errors of interpretation on the teleradiology images. Twenty six of the 28 missed teleradiology abnormalities were readily identified at adjudication when proper windowing and magnification were used. In one case of a fractured elbow and one case of minimal congestive heart failure (CHF), the abnormalities were not identified on review of teleradiology images. Overall, 18 of the discrepancies involved pulmonary disease, usually under reporting of CHF or pneumonia, 9 involved discrepancies in skeletal diseases, 3 represented misinterpretation of abdominal radiography, and 1 was unknown.

Peer-review data included cases of similar type to the teleradiology cases interpreted by the same group of physicians. Table 3 lists the specific major discrepancies identified. Because the total number of major discrepancies in both the teleradiology and the peer review components of the study was small, no attempt was made to segregate data by individual interpreters.

Table 1. χ^2 Statistical Analysis

	No Discrepancy (%)	Major Discrepancy (%)	Total
Teleradiology	2563 (95.4)	31 (1.15)	2594
Peer review	525 (83.6)	6 (0.96)	531

Table 2. Major Discrepancies in Teleradiology/Film Interpretation

Abnormalities	TM	TO	FM	SA
Minimal CHF	X			
Minimal CHF	X			X
Minimal CHF	X			X
CHF present but worse	X			X
LLL infiltrate			X	X
LLL infiltrate	X			X
RLL infiltrate	X			X
RLL infiltrate	X			X
RLL infiltrate	X			X
RML infiltrate	X			X
RML infiltrate	X			X
Infiltrate		X		X
Air-fluid level in bulla	X			X
Faint lung density	X			X
Possible lung nodule			X	X
Retrosternal nodule			X	X
Hilar adenopathy, interstitial disease	X			X
Pleural thickening called infiltrate	X			X
Trochanteric fracture	X			X
Toe fracture	X			X
Aseptic necrosis of hip	X			X
Metatarsal fracture	X			X
Nasal fracture	X			X
Calcific bursitis of shoulder	X			X
Small paraspinal mass	X			X
Elbow fracture	X			
AAA on spine film	X			X
Possible cecal volvulus	X			X
Ischemic bowel	X			X
Gastric outlet obstruction	X			X
Unknown (unintelligible)				

Abbreviations: TM, teleradiology miss; TO, teleradiology overcall; FM, film miss; SA, seen at adjudication.

The analysis of a random selection of the 315 cases transmitted for teleradiology interpretations showed 152 (48.3%) with no significant abnormality, 52 (16.5%) with minor abnormalities, 39 (12.4%) with known major abnormalities, and 72 (22.9%) with new major abnormalities.

Table 3. Major Discrepancies in Peer Review Interpretation

Abnormalities	FM	SA
LLL infiltrate	X	X
Lung nodule	X	X
Retrosternal adenopathy	X	X
Foot fracture	X	X
Finger fracture	X	X
Ureteric calculus	X	X

Abbreviations: FM, film miss; SA, seen at adjudication.

DISCUSSION

It is clearly evident from analysis of the data that the major discrepancy rate on teleradiologic interpretations at 1600 × 1200 (1.5K) pixel size is functionally equivalent to that found with interobserver discrepancies on standard peer-reviewed radiographs with the same radiologists and the same type of radiology examinations. This suggests that, despite some reports that indicate that teleradiology is inadequate for interpretation in certain specific circumstances, teleradiology overall has a similar results outcome as standard film-based interpretation when performed with resolutions equivalent to or better than that used in this study. Our discrepancy rates are most similar to those found in the study performed by Goldberg et al, which showed a 2.6% discrepancy rate between teleradiology and standard film interpretation.⁶ The similarity is striking particularly because higher resolution 2500 × 2000 pixel display stations were used in that study. As indicated by the significant percentage and distribution of minor and major abnormalities found in our sample of teleradiology cases, the discrepancy rates were not caused by a bias of a high percentage of easily interpreted normal examinations, but rather were indicative of the applicability of teleradiology across a broad range of abnormalities.

In both the teleradiology and peer-review components of this study, the second film reader potentially benefitted from knowing the opinion or report of the first reader. As a result, in nearly all cases, the second interpretation was the more “accurate.” It was much less common for the adjudication process to find either type of second reader in error. This methodologic difference is also likely responsible for the relatively low major discrepancy rates compared with double-blind analyses of interobserver variation.

Review of the pattern of teleradiology discrepancies showed several features. First, the major discrepancies in teleradiology interpretation were slightly more common with radiologists who had major discrepancies on their peer review. Major teleradiology discrepancies did, however, occur with all readers. Second, and more important, during the adjudication process, significant abnormalities missed on teleradiology and seen on radiography were visible in retrospect on the teleradiology display in nearly every case. One case of Kerley B lines and one elbow fracture were not

clearly seen at adjudication. It was apparent that routine manipulation of window and level controls was essential, particularly with pulmonary disease cases. Lower lobe infiltrates that were difficult to appreciate without manipulation were readily apparent after minimal manipulation. In bone cases, fractures were more conspicuous when magnification was used.

Radiology residents spend much of their early training assimilating information about what constitutes normalcy, how film-based interpretations are made, and how technical factors are involved in the perception of the both. Experienced radiologists must expect and accept that understanding how to interpret images from a digital display may require careful retraining to learn the nuances that optimize perception.

In our study, there was a substantially higher minor discrepancy rate for the peer review data compared with the teleradiology data. This was probably related to the fact that in peer review the text of the report was evaluated as well as the

content. Differences in reporting emphasis or description were included in those classified as minor discrepancy. However, substantive rather than stylistic differences were required to categorize a peer review discrepancy as major. Although a slightly lower major discrepancy rate for peer review compared with teleradiology was observed, this was statistically not significant.

Our multiyear experience with screen-based interpretation has given us sufficient confidence to proceed toward full implementation of a PACS environment. Our PACS system installation commencing after the completion of this present study incorporates the wide area network teleradiology functionality to the two outpatient clinics and affiliated hospitals used as a basis for this study. The new system uses workstations with 2500×2000 pixel displays. It is our intention to replicate the current clinical study using this higher resolution system over the coming years to determine the impact of higher resolution on discrepancy rates.

REFERENCES

1. Pagé G, Grégoire A, Galand C, et al: Teleradiology in northern Quebec. *Radiology* 140:361-366, 1981
2. Gayler BW, Giltin JN, Rappaort W, et al: Teleradiology: An evaluation of a microcomputer-based system. *Radiology* 140:355-360, 1981
3. Scott WW Jr, Rosenbaum JE, Ackerman SJ, et al: Subtle orthopedic fractures: Teleradiology workstation versus film interpretation. *Radiology* 187:811-815, 1993
4. Franken EA, Berbaum KS, Marley SM, et al: Evaluation of a digital workstation for interpreting neonatal examinations. A receiver operating characteristic study. *Investigative Radiology* 27:732-737, 1992
5. Slovis TL, Guzzardo-Dobson PR: The clinical usefulness of teleradiology of neonates: expanded services without expanded staff. *Pediatr Radiol* 21:333-335, 1991
6. Goldberg MA, Rosenthal DI, Chew FS, et al: New high resolution teleradiology system: Prospective study of diagnostic accuracy in 685 transmitted clinical cases. *Radiology* 186:429-434, 1993