

Evaluation of Teleradiology for Interpretation of Intravenous Urograms

Ethan J. Halpern, Jeffrey H. Newhouse, E. Stephen Amis, Jr, Herman W. Lubetsky, Robert M. Jaffe, Peter D. Esser, Philip O. Alderson

The diagnostic yield of a commercial teleradiology/picture archiving and communication system (ATT-Philips Comm View T/PACS) was evaluated for 100 urograms. A single image from each examination was digitized ($2048 \times 1684 \times 12$ -bit pixels) and transmitted from a satellite hospital over a T-1 line using the T/PACS system. The video display of each digitized image was reviewed independently by four radiologists. The same four radiologists reviewed the original film images at a different time without knowledge of their T/PACS interpretation. There was no statistically significant difference in the sensitivity for clinical findings between T/PACS (86%) and film (89%). The false positive rate, however, was significantly higher with T/PACS than with film (44 versus 32 false positive findings per 100 films). We conclude that T/PACS of the type studied here demonstrates sufficient sensitivity for the detection of clinically important urographic findings in the emergency setting. A final reading of the original films is still necessary, however, to assure appropriate specificity.

Copyright © 1992 by W.B. Saunders Company

A WIDE SELECTION of commercial teleradiology picture archiving and communication systems (T/PACS) are available. They differ considerably in image quality, performance, and price. Specific features that affect diagnostic performance may differ from one system to the next. These include (1) image acquisition (video camera versus charge coupled device (CCD) camera versus laser scanner); (2) image transmission (analog versus digital); (3) data storage (uncompressed versus compressed); (4) display speed; (5) display resolution (512, 1024, 2048, etc); (6) display contrast (8 bits versus 12 bits); (7) display gray scale mapping (linear versus log, etc); and (8) post-processing (eg, smoothing versus edge enhancement). A recent review¹ of digital imaging workstations considers many of these issues.

Clinical studies of various T/PACS systems often differ in their conclusions regarding clinical reliability. With respect to chest films, one published study² found no statistical difference in diagnostic accuracy between conventional film and video digital display. A more recent study³ observed significant differences in the detection of interstitial disease and pneumotho-

rax relative to conventional radiographs. With respect to excretory urography, one study⁴ of a computed radiography system (not a PACS system, but nonetheless a digital system having 1760×2140 pixel resolution) showed no difference between digital and screen-film urograms in diagnostic sensitivity, specificity or receiver-operating characteristic (ROC) curves. A second study⁵ found that only two thirds of renal and ureteral calculi seen on the radiographs were discerned on the video images. Results are often reported for prototype and one-of-a-kind systems.

Many studies are out of date almost as soon as they are published because T/PACS systems are constantly being upgraded. The net result is a bewildered audience of radiologists facing an array of unsubstantiated claims by T/PACS manufacturers. We do not propose to solve this dilemma. An objective comparison of different T/PACS systems is probably best accomplished with phantom studies.⁶ The present study documents the performance of a single, widely used commercial T/PACS system for a specific, well controlled task. More specifically, the study was designed to compare diagnostic accuracy for the interpretation of intravenous urograms through teleradiology as compared with the interpretation of the original films.

MATERIALS AND METHODS

A study sample of single 14×17 -inch nontomographic images from 100 urograms were previously selected as part of a study of a data compression algorithm.⁷ Half of the images were obtained from a consecutive clinical series of urograms, including both normal and abnormal studies. The other 50 images were selected from a log of interesting cases to reflect a broader spectrum of pathology. The 100 images consisted of 22 scout views, 75 views obtained between 15

From the Department of Radiology, Columbia Presbyterian Medical Center, New York, NY.

Address reprint requests to Ethan J. Halpern, MD, Thomas Jefferson University Department of Radiology, Division of Diagnostic Ultrasound, 132 S 10th St, Room 796, Philadelphia, PA 19107.

*Copyright © 1992 by W.B. Saunders Company
0897-1899/92/0502-0008\$03.00/0*

and 30 minutes after contrast administration, and three postvoiding views.

These images had been previously interpreted by a panel of four board-certified radiologists for the earlier study.⁷ In that study, each of the four radiologists, working independently, first reviewed all 100 films. The four panel members then developed a "consensus interpretation." Any finding that was independently identified by at least two of the four panel members and subsequently agreed to by a majority of the panel members was considered a true finding. This consensus interpretation identified 201 findings that served as the "gold standard" reading for the current study.

The spectrum of findings in our gold standard is summarized in Table 1. It is worthwhile to note that the gold standard did not consider clinical history, follow-up examinations, or surgical pathology. Because the goal of this study was to compare teleradiology with interpretation of original

films, the interpretation of the original films alone was accepted as the gold standard.

Three of the original four radiologists were available to review the T/PACS images for the current study. They had not seen the original images in more than 2 years. They independently reviewed only the T/PACS images for the present study. A fourth radiologist reviewed both the T/PACS images and the original films for the present study. A 6-month interval was allotted between these two readings to minimize recollection of specific details. The extended time intervals between the film and T/PACS review sessions ensured that the two sets of interpretations were essentially independent, and thus a valid sample for paired comparison of T/PACS and film interpretation.

Films were digitized and reviewed on the Comm View Digital Radiology system version 2.2 (AT&T-Philips, West Long Branch, NJ). Each film was digitized with the FD2000 laser scanner (Dupont, Wilmington, DE) to 2,048 × 1,684-pixel resolution with 12 data bits per pixel (corresponding to 4096 possible shades of gray). Images were transmitted approximately three miles over a dedicated T1 line. They were viewed on a 1,024-line monitor with both zoom and window/level capability. Each image was first viewed at low resolution (1,024²) so that the entire image could be seen at the same time. Following this, the individual quadrants of each image were re-evaluated at full resolution. Window/level settings were adjusted to the desired level by the individual observers. To optimize use of the system, each observer was given a short T/PACS training session before the study.

Each radiologist worked independently. We believe that this simulated the emergency setting reasonably well. Because the backgrounds of the four readers differed, we analyzed their responses both independently and as a group. Readers 1 and 2 are urologists; readers 3 and 4 are general radiologists. Of the four, reader 4 spends the largest proportion of his typical work time interpreting cathode ray tube (CRT)-based images.

A response form (Table 1) was completed to identify the findings for each image. Positive teleradiology findings that had been detected by consensus in the previous study⁷ were tabulated as true positives. All other findings were considered false positives. Sensitivity was calculated as the number of true positive findings divided by 201. Because many false-positive findings might be enumerated for a single image, the false-positive rate was calculated as the number of false-positive findings identified per 100 films.

To minimize the effect of interobserver variation on the statistical analysis, the interpretations of each radiologist for the T/PACS images versus films were treated as paired data. For each image ($n = 100$), two difference values were tabulated between T/PACS and film. Let D_{ijtp} represent the difference in the number of true-positive findings for case i ($i = 1$ to 100) with radiologist j ($j = 1$ to 4), and D_{ijfp} represent the difference in the number of false-positive findings for case i with radiologist j . For each radiologist, mean D_{jtp} and D_{jfp} values were calculated along with their respective standard deviations. Each of these values was tested individually with the Student's t test against the null hypothesis $D_{jtp} = 0$ or $D_{jfp} = 0$ for each radiologist ($j = 1$

Table 1. Spectrum of Findings—100 Urograms

Listing on Response Form	No. Positive
Renal mass(es)	8
Primarily cystic	
Primarily solid	
Renal outline irregularity	1
Filling defect	
Intrarenal collecting system	3
Ureter	
Bladder	
Dilatation	
Pyelocaliectasis	11
Ureterectasis	
Congenital abnormality	6
Stricture	
Infundibulum	
Ureter (including UPJ, UVJ)	
Mucosal irregularity	
Intrarenal collecting system	
Ureter	
Bladder	
Ureteral deviation	4
Bladder deviation	3
Mass impression of BPH	17
Contour change	6
Nonrenal mass (eg, adrenal)	12
Diverticulum	
Calyx	3
Ureter	
Bladder	
Calcifications	
Nephrocalcinosis	
Stone(s) in collecting system	10
Other (eg, in mass)	11
Iatrogenic objects	23
Bilateral renal function	77
Single functioning kidney	2
Normal KUB	4
Total Positive Findings	201

Abbreviations: UPJ, ureteropelvic junction; UVJ, ureterovesical junction; BPH, benign prostatic hypertrophy.

to 4). An alpha error of less than .05 was considered significant.

The combined readings of all four radiologists were tabulated by computing the values $D_{i:tp}$ and $D_{i:fp}$ as the sum of the respective $D_{ij:tp}$ and $D_{ij:fp}$ values for each case i (adding the values for $j = 1$ to 4). Mean values, D_{tp} and D_{fp} , and standard deviations were calculated for the set of values $D_{i:tp}$ ($i = 1$ to 100) and $D_{i:fp}$ ($i = 1$ to 100), respectively. These mean values were also tested with the Student's t test against the null hypotheses $D_{tp} = 0$ and $D_{fp} = 0$.

RESULTS

Figures 1 and 2 demonstrate sample images from the study. The findings of mild caliectasis and bladder calculi are seen equally well with teleradiology and the original films. However, the T/PACS images were noticeably less bright than films viewed on the viewbox. This subjective finding was confirmed by the longer exposure times (four to six times longer) required to photograph images off the teleradiology system as opposed to the time required to photograph the corresponding films off a viewbox.

Tables 2 and 3 present the sensitivity and false-positive rate data for the independent interpretations of each of the four radiologists. Each radiologist reading individually detected fewer findings on film than were detected by the previous consensus interpretation (Table 2), and detected even fewer findings by teleradiology. However, these differences were small and significant only for reader number 1. The overall sensitivities (89% for film versus 86% for T/PACS) were not significantly different. Of particular interest, the mean sensitivity for calcifications ($n = 21$) was 74% with film versus 77% with T/PACS. The mean sensitivity for dilatation of the intrarenal collecting system (pyelectasis or caliectasis: $n = 11$) was 92% with film versus 88% with T/PACS.

False-positive rates were significantly lower with film than with T/PACS for the first three readers. Interestingly, reader 4, who showed almost no loss of sensitivity with T/PACS also showed no loss in specificity. The mean false-positive rate for film (32 findings per 100 im-

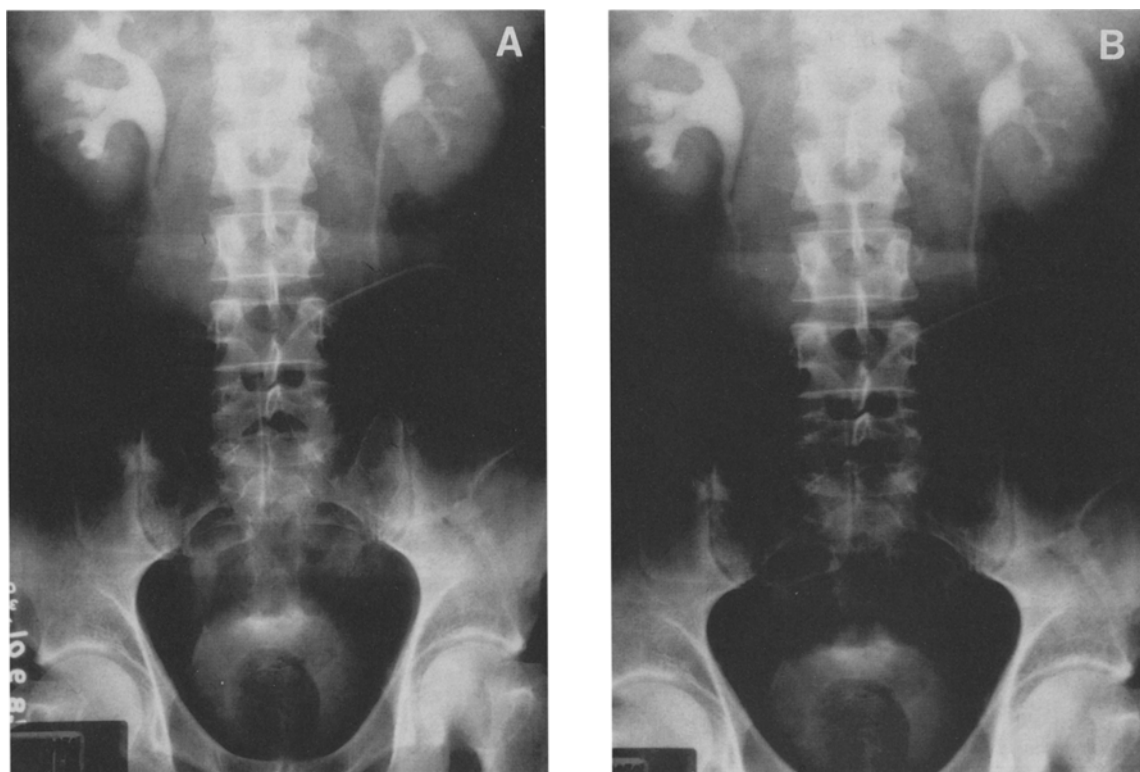


Fig 1. Single image from an intravenous urogram showing bilateral renal function with unilateral mild caliectasis. A compression device and Foley catheter are present. (A) Original film; (B) image displayed on the T/PACS system.

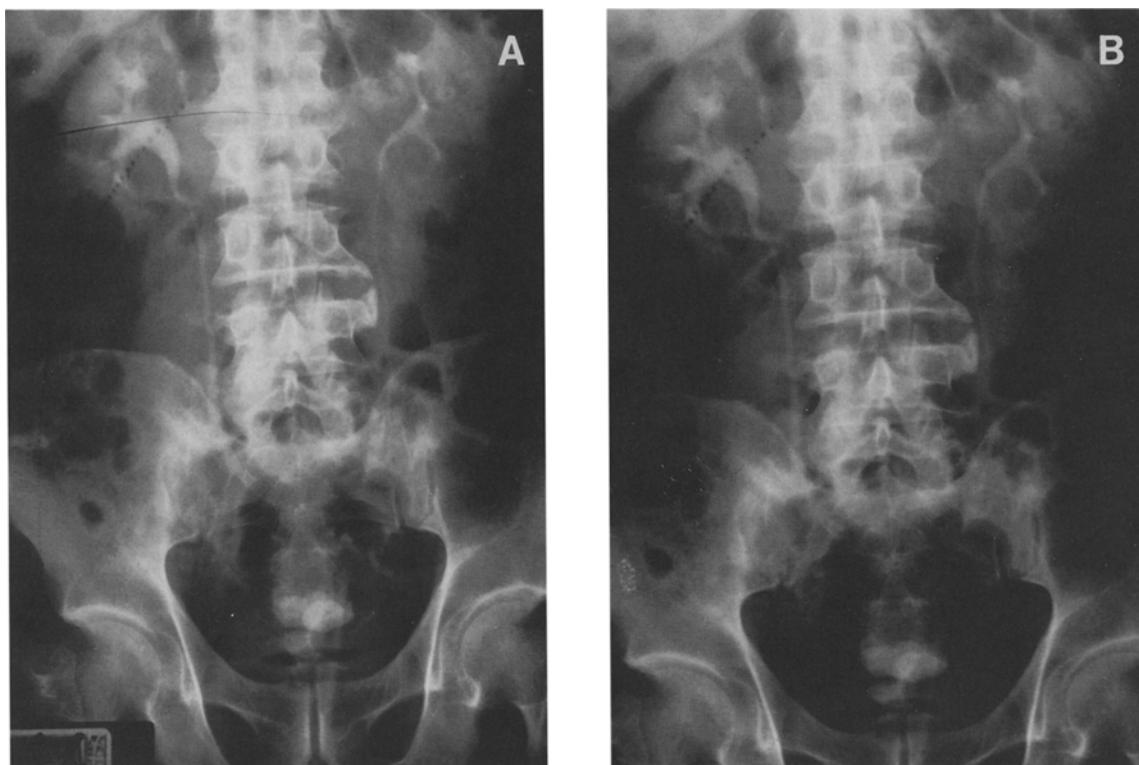


Fig 2. Single image from an intravenous urogram showing bilateral renal function. Several bladder calculi are present. (A) Original film; (B) image displayed on the T/PACS system.

ages) was significantly less than the rate for T/PACS (44 per 100). False positive findings on the T/PACS were diverse, including caliectasis ($n = 37$), calcifications ($n = 23$), prostatic enlargement ($n = 13$), bladder mucosal irregularity ($n = 11$), and duplication of the collecting system ($n = 3$). The n value here represents the total number of false-positive interpretations by all four radiologists in each category.

All four readers noted that the time required to interpret images with the T/PACS system was "substantially" greater than that required for interpretation of the original films. Unfortu-

nately, the time required for interpretation was not recorded in this study.

DISCUSSION

The use of teleradiology and T/PACS has been investigated for many types of radiographic studies, including, but not limited to, chest radiography,^{2,3} excretory urography,⁵⁻⁶ calcifications in mammography,⁸ periosteal reaction in bones,⁹ and mucosal abnormalities in the gastrointestinal tract.¹⁰ There is a general consensus that fine detail work such as the interpre-

Table 2. Diagnostic Sensitivity

Reviewer	Film	T/PACS
1	92.5%	89.1%*
2	92.0%	88.1%
3	86.6%	82.6%
4	84.6%	84.1%
Average	89%	86%

NOTE. $n = 201$ findings.

*Statistically significant ($P < .05$) difference between T/PACS and film reading.

Table 3. False-Positive Rate

Reviewer Number	Film	T/PACS
1	12	23*
2	11	24*
3	51	80*
4	52	49
Average	32	44*

Note. $n = 100$ films.

*Statistically significant ($P < .05$) difference between T/PACS and film reading.

tation of microcalcifications or periosteal reaction requires better resolution and contrast than many other types of studies. However, there are no broadly accepted standards of digital image resolution and contrast requirements for teleradiology and T/PACS.

The data concerning teleradiologic urography have been conflicting, with one study⁴ finding equivalence between digital radiography and film while a second study⁵ found teleradiology to be unsatisfactory for the detection of renal and ureteral calculi. The first study⁴ used a computed radiography system (model 201; Toshiba, Tustin, CA) with photostimulable plates. Ten bits of digitized data were stored per pixel (corresponding to 1,024 gray-scale shades). Films (14 × 17 in) were created from the digital data at a resolution of 1,760 × 2,140 pixels. The second study⁵ used a 1,024 × 1,024-matrix digitizer (TEL-X-100 Scanner, Raytel, San Jose, CA).

The present study used a matrix laser digitizer (FD2000 scanner, Dupont, Wilmington, DE). Our evaluation of this 2,048 × 1,684 resolution system was designed to test a widely used commercial teleradiology system (ATT Philips Comm View, West Long Branch, NJ) for use with urography. Our overall results suggest that such a system is adequate for detecting abnormalities on intravenous urograms, including calculi, in the emergency setting. Several aspects of the study, however, warrant further discussion.

As mentioned above, readers 1 and 2 practice primarily urology while readers 3 and 4 are general radiologists. Reader 4 has the greatest experience with CRT-based modalities. Reader 3 works primarily with plain film radiography. As seen in Tables 2 and 3, both the sensitivity and specificity of image interpretation were greatest for readers 1 and 2. Reader 1 was the only one in the group to show a significant difference with respect to both sensitivity and specificity for T/PACS versus those of conventional film. Reader 4, the youngest in the group, was the only one to show no difference with respect to either sensitivity or specificity. Reader 3 showed the greatest increase in false-positive findings with T/PACS. These differences among the readers suggest two separate learning curves,

one for urography and the other for image interpretation on a CRT. Additional empirical studies might better define the complex relationship of these learning curves to T/PACS image interpretation.

The number of false-positive findings tabulated for the readers in Table 3 seems somewhat high. This is at least partially a consequence of the interobserver variation that was present among the radiologists with respect to the more subtle findings in our series (ie, mild pyelocaliectasis or prostatic enlargement). Such findings were included in the gold standard only if they were identified by the independent interpretations of at least two of the four radiologists in the original study panel and accepted by the majority of that panel.⁷ Findings that were identified by one of the original study panel members, but not included in the gold standard, would be tabulated as false positives if they were identified again in the current independent interpretations.

The false positive rates of readers 3 and 4 seem particularly high in comparison with those of readers 1 and 2. It is possible that readers 1 and 2, who tend to have a high rate of concordance on their interpretations, had a slightly greater influence in deciding the gold standard. None of the original panel members, however, believed that this had occurred. In fact, the panel members all thought that they had participated equally in defining the gold standard. Thus, differences in false-positive rates between the various observers are most likely due to the varied backgrounds of the different radiologists. Similarly varied backgrounds are likely to be found among a sample of practicing radiologists.

As indicated above, the sensitivity/specificity data in the current study show a fair degree of interobserver variation among the radiologists. Nonetheless, the finding of a statistically significant difference between image interpretations on T/PACS and conventional urograms for any individual reader is valid because the interpretations of each reader were treated as paired data (eg, the difference in the number of false-positive findings between T/PACS and conventional urograms for each reader is a valid measure of T/PACS specificity relative to film).

Mean sensitivity values did not differ significantly between T/PACS and conventional film. Differences between the consensus findings and individual interpretations on film could represent, in part, intraobserver variability or individual uncertainty. In any case, T/PACS overall was comparable in allowing the findings to be seen. The mean false-positive rate, however, was significantly higher for the T/PACS interpretations. Furthermore, all four readers agreed that image interpretation required "substantially more time" with T/PACS as compared with film. The increased time was mostly spent adjusting the window/level settings and the degree of zoom. It is quite possible that the additional window/level and zoom features provided with our T/PACS system allow readers to see image details in a way to which they are not accustomed. This may result in added uncertainty and overinterpretation of findings. It is also possible that with additional experience, readers will learn to use the system more

quickly and to ignore these "false-positive" findings.

Interpretation of urograms with the T/PACS system tested in this study yielded a sensitivity for pertinent clinical findings comparable with that found with the original urogram films. Sensitivity for calcifications and pyelocaliectasis, the abnormalities most likely to occur in the emergency setting, were approximately equal with T/PACS and the original films. We therefore feel comfortable with the use of this T/PACS system in the emergency setting for interpretation of urograms. This does not imply that a system with lower resolution/contrast abilities will be equally suitable. A final interpretation of the original conventional urogram images still seems necessary, however, to ensure adequate specificity in these studies. Further studies will be required to demonstrate whether diagnostic specificity with T/PACS increases along a learning curve with additional experience.

REFERENCES

1. Arenson RL, Chakraborty DP, Seshadri SB, et al: The digital imaging workstation. *Radiology* 176:303-315, 1990
2. Hayrapetian A, Aberle DR, Huang HK, et al: Comparison of 2048-line digital display formats and conventional radiographs: an ROC study. *AJR* 152:1113-1118, 1989
3. 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
4. Fajardo LL, Hillman BJ, Hunter TB, et al: Excretory urography using computed radiography. *Radiology* 162:345-351, 1987
5. DiSantis DJ, Cramer MS, Scatarige JC: Excretory urography in the emergency department: utility of teleradiology. *Radiology* 164:363-364, 1987
6. Halpern EJ, Esser PD, Nickoloff EL, et al: A quality control phantom for digitization of radiographs. *J Digital Imaging* 3:1:42-48, 1990
7. Halpern EJ, Newhouse JH, Amis ES, et al: Evaluation of a quadtree-based compression algorithm with digitized urograms. *Radiology* 171:259-263, 1989
8. Chan H, Vyborny CJ, MacMahon H, et al: Digital mammography ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications. *Invest Radiol* 22:581-589, 1987
9. Murphey MD: Digital skeletal radiography: spatial resolution requirements for detection of subperiosteal resorption. *AJR* 152:541-546, 1989
10. Kastan DJ, Ackerman LV, Feczko PJ: Digital gastrointestinal imaging: the effect of pixel size on detection of subtle mucosal abnormalities. *Radiology* 167:853-856, 1987