Clinical Assessment of Phosphor-Plate Computed Radiography: Equipment, Strategy, and Methods

Bruce J. Hillman and Laurie L. Fajardo

Initial experiences with a medical innovation project promise future applicability. Scientific assessment of the innovation usually points out deficiencies that must be overcome before optimal, widespread use of the technology is feasible. This presentation details the equipment, strategy, and methodological principles used at the University of Arizona in assessing the clinical use of computed radiography based on using phosphor plates as the X-ray receptor. The focus of this research has been to evaluate the applicability of computed radiography to clinical indications which might test potential weaknesses or virtues of the technology. We present examples of experiments that illustrate the general principles described.
© 1989 by W.B. Saunders Company.

COMPUTED RADIOGRAPHY (CR) is an innovation in medical imaging that has several technological incarnations. Perhaps the most popular among those technologies currently undergoing testing in clinical practice is phosphor-plate CR, which uses conventional radiographic equipment to expose an image on a photoelectronic imaging plate rather than on a film-screen (FS) combination.

Early experiences with CR indicated potential applicability of the technology to some important imaging indications, particularly the evaluation of chest roentgenograms for pulmonary nodules. However, features of CR—notably its high contrast and low spatial resolution compared with roentgenograms and its potential for interactive postprocessing—suggested to us that radiologists' using CR successfully likely would depend on the nature of the diagnostic task.

To account for this, we evaluated the applicability of CR to clinical indications that would test

unique advantages or disadvantages of CR compared with conventional film/screen radiography. The results of these experiments would be representative of other similar diagnostic tasks and would guide the further technologic development of CR toward its implementation for comprehensive imaging. This presentation describes the CR equipment that we have applied to clinical imaging and the strategy and methodologic principles that have guided our assessments.

GENERATION OF COMPUTED RADIOGRAPHY IMAGES

Our clinical assessments have used the TCR 201 computed radiography system (Toshiba Corporation, Tokyo, Japan). This system typified current standards and designs for computed radiography and is comprised of several components: imaging plates, an image plate reader, image processor/arithmetic unit, image storage devices, cathode ray tube (CRT) image display, and an image recorder for producing hard-copy films.

Imaging Plates

Computed radiography uses imaging plates such as the x-ray receptor, that are exposed by conventional radiographic equipment. The plates are composed of europium-doped barium-fluorohalide crystals deposited on a polymer base. These crystals are "photostimulable." The crystals are raised to a higher energy state by exposure to x-rays and retain this energy. The stored energy is released when the plate is laser-scanned in the image plate reader. Imaging plates have a linear dynamic range of $10^6 : 1$, approximately 100,000 times that of conventional FS combinations. Spatial resolution for the final images is on the order of 2.5 to 5.0 line pairs per millimeter, depending on the size and type (standard or high resolution) of imaging plate used. Reading the imaging plate erases it and allows it to be reused.

Image Formation

Reading of the imaging plate involves an initial scan with a low-power helium-neon laser

to determine overall exposure and the dynamic range of tissues to be visualized. This initial scan allows for either automatic or user controlled adjustments of the system's sensitivity. Adjustments may be made to correct under- or overexposure of the image and to compensate for inaccuracies and nonuniformities in the x-ray source and detection system. The imaging plate is then rescanned with a high intensity heliumneon laser beam, that releases the stored energy from the imaging plate crystals as blue phosphoresence. The energy is conducted via a "light pipe" to a photomultiplier tube for conversion into an electrical signal. These electronic signals are digitized into a matrix that ranges from 1,760 × 2,140 for a 14 × 17–inch imaging plate to 2,000 × 2,510 for an 8 × 10–inch plate, with 10 bits or 1,024 gray levels. The 10 bits are fully used regardless of the radiation exposure used to form the image. This means that the specific densities of interest (eg, soft tissues, fat planes, calculi, iodinated contrast media) have their relative specific absorption characteristics expanded or contracted to fit the γ response curve of the digital system. The γ response curve can be automatically selected by the system's computer or generated manually by the user, as described below.

The digital information is then transferred to the image processor/arithmetic unit. After processing, the image can be displayed on a CRT console or stored on magnetic tape (short-term storage) or on optical disk (long-term storage). If hard-copy radiographs of the digital image are desired, the data are passed through a digital-to-analogue converter. An acousto-optical modulator[1] uses the resulting analog signals to modulate the intensity of a laser beam that "paints" each line of the image onto a single emulsion film.

Image Enhancement and Postprocessing

A great potential advantage of CR is the ability to tailor CR images to optimize visualization of desired information. Postprocessing of digital image data can be performed automatically or manually by using two image processing functions, gradation and frequency processing. Automatic computer processing is accomplished by preprogramming tables that specify parameters for a particular type of examination. With

manual processing, the radiologist can experiment with various combinations of processing parameters to enhance specific features of the image.

Gradation processing allows image optimization by changing the relationship between optical density and input (in a manner analogous to choosing the best image from many different films) each with its own characteristic H and D curve and resultant contrast relationships. In addition, an input-density, or γ curve, can be shifted along the input signal axis. This changes the signal input-to-density relationship, making the image lighter or darker (Fig 1). It is also possible to rotate the data around a preselected point on the γ curve, which primarily affects the

Fig 1. Effect of gradation processing on a CR excretory urogram image that is the same as that in Fig 4A. The only difference is that the H and D curve is shifted to the right (dotted curve), making the image lighter. QL, output image signal (horizontal axis); D, output density (vertical axis).

contrast of the image; a steeper angle provides for a shorter contrast scale, whereas a slowly rising curve allows for broader contrast range (Fig 2). These manipulations have potential use for evaluating particularly dark or light areas of an image for suspected abnormalities. The net effect of these operations is equivalent to providing the user with an infinite number of different output films, each with its own speed and contrast characteristics (Fig 3).

Frequency processing is a method of selectively enhancing or diminishing specific spatial frequency components in an image. Much of the information that the eye uses to observe an image is contained in the edge interfaces between different tissues. These edges contain a large compo-

Fig 3. Effect of gradation processing on a CR excretory urogram image that is the same as that in Fig 4A. In this case, the gradation scale has been reversed (dotted curve) to produce a "negative" image.

Fig 2. Effect of gradation processing on a CR excretory urogram image that is the same as that in Fig 4A. The only difference is that the H and D curve is steeper, representing a shorter contrast scale (dotted curve).

nent of high-frequency information. Thus, the user might choose to enhance edges by selectively amplifying high frequencies and suppressing the low frequencies which may tend to obscure the signal. Such an operation is called frequency processing or spatial filtering. The frequency response of a digital image can be enhanced over a range selected by the user; the degree of enhancement is a function of the signal strength at a particular point on the image. For observing fine detail, such as pulmonary interstitial lines or bony ridges, high-frequency processing is used. For evaluating large or smooth patterns, such as kidneys, low-frequency processing is applicable.

High-pass spatial filtering, or edge enhancement, attenuates smooth transitions and en-

hances edges. A blurred subtraction image is produced from the original image data and subtracted from it. This results in the emphasis of edges and other abrupt contrast changes. For situations such as chest imaging, some degree of edge enhancement is useful, especially when the image is displayed on film. Too much edge enhancement, however, can reduce tissue contrast and increase the background random noise level to a point where diagnostic information is lost. The converse of edge enhancement is called smoothing, or low-pass filtering. In this situation, high frequency information is suppressed and low frequency data amplified. Smoothing blurs abrupt transitions in an image and reduces the overall random noise by soothing the variations in the background intensity levels.

The TCR 201 conventionally produces a pair of processed, final images for each exposure. One is processed with an algorithm that produces an image which simultates a conventional FS im-age. The second image is frequency modified, using either low- or high-frequency enhancement, depending upon the type of study (Fig 4).

PHILOSOPHICAL FOUNDATIONS FOR THE SCIENTIFIC ASSESSMENT OF COMPUTED RADIOGRAPHY

The assessment of CR has followed the well described pattern that typifies the evaluation of almost any medical imaging innovation.[2] Initially, when a technology first becomes available to a few academic centers, a small series of anecdotes and reports appear on how the technology is useful in diagnosis for a variety of indications. These types of endeavors, although necessary to the early understanding of a technology, are neither scientific nor reliable, because they are usually retrospective, lack appropriate controls, cannot report results in generally accepted metrics, and often do not compare results with a reliable standard. There is a pronounced ten-
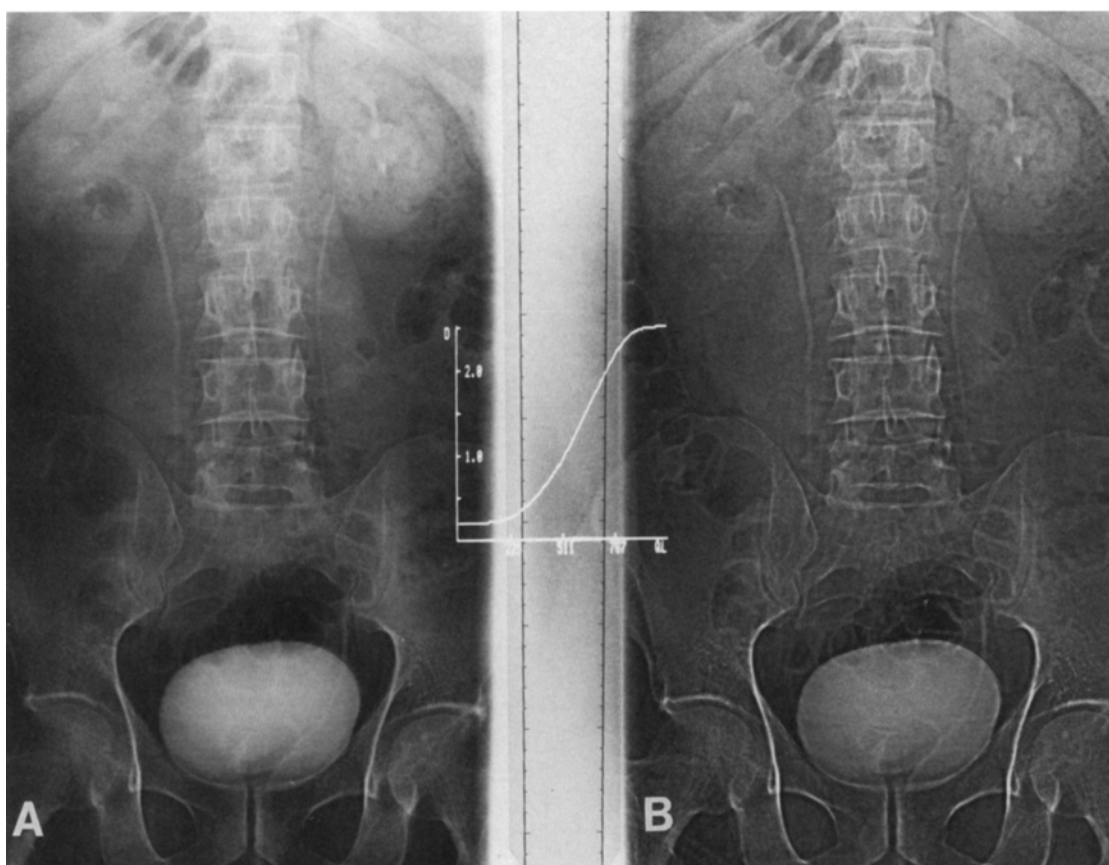

Fig 4. Typical CR pair of images produced from a single urographic exposure (A) Processed by applying an algorithm that simulates a conventional radiograph; (B) frequency modified to enhance edges.

dency for such reports to be biased in favor of the new technology.

Scientific assessment of a new technology follows a hierarchy which is relevant to the use of the innovation for clinical practice.[3] First, is CR as technically efficacious as the current technology which it might replace, FS radiography? The question asked in this type of assessment is whether radiologists perform as well interpreting CR images as they do with FS. Comparisons of performances with the two modalities usually are couched in such standard measurements of diagnostic performance as sensitivity, specificity, and receiver operating characteristic (ROC) curve areas. This level of assessment typifies most of what has been done to evaluate CR scientifically to date.

Assurances that CR provides technical efficacy comparable to FS are a necessary condition for its clinical acceptance; however, some unique benefit attributable to CR must be demonstrated to warrant its use. Does CR provide a benefit in improving referring clincans' diagnoses or therapies? Will patients' health outcomes be improved or will imaging be less expensive with CR than is currently the case with FS? The experiments described in this presentation largely focus on evaluating the technical efficacy of CR; often the results show how using CR might improve patient outcomes and reduce imaging costs.

SELECTION OF INDICATIONS AND METHODOLOGIC CONSIDERATIONS

Any clinical indication selected for evaluating the technical efficacy of CR had to be representative of a genre of clinical tasks that is prevalent and clinically important. Elements of an imaging task selected for evaluation should test some aspect of the CR system that, from observations, might be expected to represent either a significant virtue or deficit of CR. While methodologic specifics varied somewhat among experiments, certain commonalities exist that typify our assessment of the technical efficacy of CR.

Investigations of CR that will be useful for other settings depend on selecting case material that is either representative or important. To contend that a selection of cases is representative usually necessitates that selection be either consecutive or random and that the sample be large enough to include examples of all versions of the test findings that might be encountered in clinical practice. However, often we could only include a relatively small number of cases in an experiment, either because of time constraints imposed by the responsibilities of the experimental readers or the rarity of the finding of interest. We recognized that in such instances we could not make a case for the representativeness of the sample. Rather, we insisted that included cases represent findings that would be "important" to radiological practice and thus test the clinical applicability of the CR system. We also insisted that findings represent a range of difficulty in detection, so that we would be able to elicit differences in the performances of radiologists interpreting CR and FS images if any such differences existed. We always included normal cases to test the rates at which CR and FS engendered false positive diagnoses.

Selecting experimental readers representative of all radiologists was also problematic. We avoided using junior radiology residents, because they may not be similar to more advanced residents or practicing radiologists, which we have viewed as equivalent. The results of our experiments uniformly endorse this assumption. Still, there is no way of assuring that the performances of our subjects are representative of radiologists who might use CR in clinical practice.

We have adopted numerous controls to promote unbiased and accurate paired comparisons of radiologists' performances with CR and FS. Both CR and FS images epitomized the technologies at the time of the experiments. We have avoided recall bias (a radiologist's seeing a finding on one modality that ordinarily would have been missed by recalling seeing it on the other) either by allowing each radiologist to see only one image of the CR-FS pair, or by spacing multiple reading sessions and ensuring that paired images from the same patient did not appear in the same session. When indicated, we included case material showing abnormalities other than the one for which we were testing to discourage directed searches that might bias the result. We required readers to indicate the location of a detected finding to insure that a combination of false positive and false negative decisions were not scored as a true positive response. Because of the

impossibility of obtaining definitive pathological diagnoses for most important imaging tasks, the performances of test readers were assessed using the diagnoses of a panel of experts as a standard. These panels were composed of a group of expert readers, who had all of the images for each patient available for simultaneous review to help them arrive at a consensus diagnosis.

Finally, to enhance generalizability, we assessed individual and group performances with CR and FS using standard measurements. Our approach was to require test readers to respond to the presence or absence of findings in terms of a certainty scale; indecision was not permitted. Sensitivity and specificity were calculated by designating each response as either positive or negative, based on being in either the first or second half of the certainty scale, and comparing the response with the correct diagnosis. These data, along with the certainty estimations, were used to develop ROC curves that served as the ultimate standard to compare radiologists' performances for an imaging task using CR and FS technologies. We used various paired parametric statistical analyses to test the significance of differences between CR and FS. When appropriate, we developed 95% confidence intervals to show the clinical pertinence of the magnitude of the differences.[1]

EXAMPLES OF ASSESSMENTS OF THE TECHNICAL EFFICACY OF COMPUTED RADIOGRAPHY

The following subsections are not intended to represent an exhaustive review of the assessments of CR performed in our department. Rather, they summarize several assessments which have been published or submitted for publication that are representative of the general principles for assessment that have been discussed above.

CR for Excretory Urography

Excretory urography depicts the subtle contrast differences manifested by normal and abnormal tissues in the urinary system. For most important urinary tract abnormalities, spatial resolution is less of a constraint on accurate diagnosis. Because this combination of circumstances matched the profile of physical characteristics offered by CR, excretory urographic diagnosis of urinary tract abnormalities seemed a good subject for a first, gross evaluation of the potential efficacy of CR in clinical practice.

One hundred consecutive patients presenting for excretory urography were enrolled in the experiment; only children and pregnant women were excluded. We exposed paired CR and FS scout radiographs prior to intravenously administering contrast material. Paired images exposed following injection included immediate nephrotomograms, 10-minute zonograms with ureteral compression, and 15-minute abdominal films. Exposure factors were noted for each CR and FS image.

Three experienced radiologists reviewed the CR and FS images, noting their decisions and confidence levels (1 to 6 scale of certainty; 1, definitely normal, 6, definitely abnormal) concerning the presence or absence of clinically significant abnormalities on standardized response forms. Readers also offered specific diagnoses for detected abnormalities and subjectively assessed the quality of the CR and FS images.

The study included 40 normal cases and 60 cases with abnormalities that might affect patient care. There were no significant differences between CR and FS in sensitivity, specificity, predictive values, or ROC curve area. Computed radiography engendered fewer omissions of specific diagnoses than FS but a greater number of incorrect ones. Receiver operating characteristic analysis indicated that readers' performances with the two technologies were equivalent. Image quality was also similar. The x-ray exposure required to produce CR images was 53% of that needed to produce FS urograms. Subsequent studies have sustained that this is an important improved patient outcome associated with CR imaging.[3] A further investigation of the same series of urograms indicates that radiologists respond to differences in CR image quality with respect to the accuracy and certainty of their diagnoses in a fashion similar to that of FS.[4]

CR for the Detection of Renal Calculi in Patients Undergoing Extracorporeal Shock Wave Lithotripsy

Extracorporeal shock wave lithotripsy (ESWL) is a new therapy developed to treat kidney stones noninvasively. In tests to date, ESWL appears generally safe; however, its effectiveness in com-

pletely curing patients of renal calculi has been questioned because of the methods of imaging evaluation that have been applied.[5,6] For the most part, patients have been evaluated before and after ESWL by plain abdominal radiographs that are relatively insensitive to the small fragments of stones left by ESWL and that may result in recurrent stone formation. Because CR offers excellent contrast sensitivity, it was thought that CR might improve the evaluation of ESWL patients.

In each of 50 cases, we exposed a plain abdominal radiograph (KUB) and CR and FS renal tomograms both within 1 week before and within 24 hours after ESWL. Six radiologists individually interpreted the resulting 300 imaging studies, noting the presence or absence of calculi in each of four locations for each kidney and indicating their confidence in their decisions. The result was 2,400 decisions for each radiologist (300 examinations × 4 locations × 2 kidneys) that formed the basis for categorizing responses and measuring individual readers' performances interpreting studies exposed before and after ESWL.

An expert panel of two radiologists and a urologist determined that prior to ESWL, stones were present in 58 of 100 kidneys with fairly uniform distribution with respect to size and location within the kidneys. After ESWL, stones were still present in 54 kidneys. Readers were significantly more sensitive detecting calculi by FS tomography than they were by KUB for both pre- and post-ESWL studies. All six radiologists were more sensitive interpreting CR tomograms than FS tomograms; this result was statistically significant for studies performed both before and after ESWL. However, CR interpretations also were significantly less specific in both settings. Significant differences also were evident with respect to ROC curve area. In the pre-ESWL setting, both CR and FS tomograms were interpreted significantly more accurately than were KUBs; there was no difference between CR and FS performances. Post-ESWL, the areas under individual radiologists' CR tomogram curves were significantly greater than those under the FS tomogram curves, which still represented a considerable improvement in detection over that with KUBs. Assessment of 95% confidence inter-

vals for ROC areas indicates that these differences are likely to be important in clinical practice.[7]

Accurate imaging evaluation of patients before ESWL is important for planning ESWL treatment. Following ESWL, accurate imaging is essential in determining whether ancillary treatment is necessary and for assessing prognosis. This study indicates that CR may provide radiologists with incremental improvement in their ability to provide this information, to the extent that patient outcomes might be improved.

CR for the Detection of Pneumothorax

A major concern of recommending the implementation of CR for comprehensive imaging is that limitations in spatial resolution may hinder the detection of high frequency abnormalities. A good example of such a clinically important abnormality is pneumothorax, which, in addition, is also subtle with respect to contrast. To test the ability of radiologists to detect pneumothoraces on CR, we accumulated 50 cases of matched CR and FS anteroposterior chest roentgenograms, 25 of which showed pneumothoraces and the other 25 showed different chest abnormalities. Eight radiologists independently reviewed the cases without knowledge about what was the focus of the study, indicating the presence of up to five abnormalities on each roentgenogram and their level of certainty concerning the presence of each lesion. The analysis was restricted to the detection of pneumothoraces.

There was no difference in specificity between CR and FS. Four of the radiologists were much more sensitive in detecting pneumothoraces on FS chest roentgenograms than they were on CR images, while the other four were equally sensitive with both technologies. A similar result was obtained for individuals' ROC curve areas. Interestingly, it was the three most sensitive radiologists by FS that performed the worst interpreting CR images. Success in detecting pneumothoraces with either modality was unrelated to the size of the pneumothorax.[8]

The bimodal result of this experiment suggests that factors other than spatial resolution alone may be responsible for the poorer performance of radiologists detecting pneumothoraces on CR chest roentgenograms. One possibility is that

peculiarities of current CR technology, such as smaller image size or lack of experience with the edge-enhanced image, might have differentially affected radiologists' performances. Nonetheless, the poor performances of the most accurate radiologists with CR in detecting pneumothoraces is cause for concern.

CONCLUSIONS

The scientific evaluation of digital radiography is only barely under way. However, a considerable amount has been learned about the current clinical capabilities of the technology that can guide its future development. The purpose of this presentation has been to detail our experience in assessing CR, setting forth both the strategy and the methodologies that we have pursued, and to provide examples of individual assessments guided by these principles.

We have not even begun to approach investigating the value of several potentially important qualities of CR. Notably, greater use of interactive postprocessing, histogram analysis, and interpreting images directly from the CRT screen are features that might improve diagnosis and merit evaluation. These qualities lend themselves to evaluation by the same principles outlined above and are on our agenda for investigation.

A common criticism of scientific assessment of any innovation is that the technology is advancing quickly, such that new iterations of the technology might invalidate earlier assessments.[9] Sometimes this is the case; more often, extent assessments remain qualitatively valid. Regardless of which is true for CR, assessment at this time provides continuing surveillance of the appropriateness of CR for clinical application and provides the most reliable information to instruct the improvement of the technology.

REFERENCES

1. Hanley JA: The place of statistical methods in radiology (and in the bigger picture). Invest Radiol 24:10-16, 1989

2. McKinley JB: From 'promising report' to 'standard procedure': Seven stages in the career of a medical innovation. Milbank Q 59:374-411, 1981

3. Fajardo LL, Hillman BJ, Hunter TB, et al: Excretory urography using a computed radiography system. Radiology 162:345-353, 1987

4. Fajardo LL, Hillman BJ: Image quality, diagnostic certainty and accuracy: Comparison of conventional and digital urograms. Urol Radiol 10:72-74, 1988

5. Drach GW, Dreisler S, Fair WR, et al: Report of the United States cooperative study of extracorporeal shock wave lithotripsy. J Urol 135:1127-1133, 1986

6. Mulley Jr AG, Carlson KJ, Dretler SP: Extracorporeal shock wave lithotripsy: Slam-bang effects, silent side-effects? Am J Roentgenol 150:316-318, 1988

7. Sacks EM, Fajardo LL, Hillman BJ, et al: Prospective comparison of plain abdominal radiography with conventional and digital tomography in assessing renal ESWL patients. Unpublished.

8. Fajardo LL, Hillman BJ, Pond GD, et al: Detection of pneumothorax by digital and conventional chest imaging. Am J Roentgenol 152:475-480, 1989

9. Hillman BJ: The value of imaging technology to patient's health. Am J Roentgenol 150:1191-1982, 1988