
Hany Kashani, MD, Jianan Grace Gang, BASc, Nicholas A. Shkumat, MSc, Carlos A. Varon, BEng, John Yorkston, PhD, Richard Van Metter, PhD, Narinder S. Paul, MD, Jeffrey H. Siewerdsen, PhD

Rationale and Objectives. The aim of this study was to assess the performance of a newly developed dual-energy (DE) chest radiographic system in comparison to digital radiographic (DR) imaging in the detection and characterization of lung nodules.

Materials and Methods. An experimental prototype was developed for high-performance DE chest imaging, with total dose equivalent to a single posterior-anterior DR image. Projections at low and high peak kilovoltage were used to decompose DE soft tissue and bone images. A cohort of 55 patients (31 men, 24 women; mean age, 65.6 years) was drawn from an ongoing trial involving patients referred for percutaneous computed tomography–guided biopsy of suspicious lung nodules. DE and DR images were acquired of each patient prior to biopsy. Image quality was assessed by means of human observer tests involving five radiologists independently rating the detection and characterization of lung nodules on a nine-point scale. Results were analyzed in terms of the fraction of cases at or above a given rating, and statistical significance was evaluated using Wilcoxon’s signed-rank test. Performance was analyzed for all cases pooled as well as by stratification of nodule size, density, lung region, and chest thickness.

Results. The studies demonstrated a significant performance advantage for DE imaging compared to DR imaging (P < .001) in the detection and characterization of lung nodules. DE imaging improved the detection of both small and large nodules and exhibited the most significant improvement in regions of the upper lobes, where overlying anatomic noise (ribs and clavicles) are believed to reduce nodule conspicuity on DR imaging.

Conclusions. DE imaging outperformed DR imaging overall, particularly in the detection of small, solid nodules. DE imaging also performed better in regions dominated by anatomic noise, such as the lung apices. The potential for improved nodule detection and characterization at radiation doses equivalent to DR imaging is encouraging and could augment the broader use of DE imaging. Future studies will extend the initial cohort and rating scale tests to a larger cohort evaluated by receiver-operating characteristic analysis and will evaluate DE imaging in comparison and as an adjuvant to low-dose computed tomography.

Key Words. Dual-energy imaging; flat-panel detector; diagnostic radiology; pulmonary nodule; chest radiography; lung cancer.

Because the prognosis for advanced-stage lung cancer is extremely poor, early detection is a major priority (1,2). X-ray projection imaging (screen film, computed radiography, or digital radiography) is still the most common tool in chest radiology, in part because of its low dose, low cost, and high work flow efficiency, but it is known to exhibit poor sensitivity in the detection of small, subtle pulmonary nodules (3,4). A major limiting factor is anatomic clutter superimposed within the projection, which can impede the detection and characterization of subtle lung nodules (5,6). To mitigate this limitation, dual-energy (DE) imaging...
acquires two projections of the patient at different x-ray energies and selectively decomposes the image into soft tissue and bone components. The former presents soft tissue structures in a context that is largely free from the main source of anatomic clutter (the ribs and clavicles), thus improving the conspicuity of subtle nodules (7). Furthermore, because the presence of calcification is an important indicator of benignancy, DE imaging could help characterize benign lesions with a higher level of specificity (3,8).

Previous studies have investigated single- and dual-exposure DE imaging in comparison to conventional chest radiography (9–12), with each demonstrating increased detectability of lung nodules in DE images. A variety of DE imaging systems are commercially available and are finding broad clinical utilization, including single-shot computed radiographic systems (FCR XU-D1; Fujifilm, Tokyo, Japan) and double-shot flat-panel detector systems (XQ/I Revolution, GE Healthcare, Milwaukee, WI) featuring a 200-ms delay between low- and high-energy exposures. Kelcz et al (9) established the advantage of DE images in terms of both the detection of nodules and the improved visibility of nodule calcification. More recently, investigators have described the use of a flat-panel detector in DE imaging to improve the detectability of lung nodules even further because of higher detective quantum efficiency compared to screen-film and optimized imaging techniques (eg, the optimal low- and high-kVp pair and dose allocation). Finally, the ability of the prototype to acquire DE images at a total radiation dose equivalent to that used for the DR images suggests a potential “new normal” in x-ray radiographic examination, wherein a composite radiograph roughly equivalent to a DR image may always be decomposed, but material-specific “soft tissue” and “bone-only” decompositions offer improved conspicuity and characterization of subtle abnormalities.

MATERIALS AND METHODS

Volunteer Patient Cohort

Patients were accrued under informed consent in a prospective, nonrandomized trial with approval from the institutional research ethics board and with Health Canada Investigational Testing Authorization. Total accrual for the trial is approximately 200 patients, each drawn from the patient population referred for percutaneous computed tomography (CT)-guided biopsy of suspicious lung nodules. Patients were accrued into five arms differing in DE imaging technique to evaluate different DE imaging parameters, including differential x-ray spectra, cardiac gating, dose allocation, and total dose. All images used in the study reported herein were taken from group 1 (using the “optimal” DE imaging technique, described later), with 55 patients from the 90 patients accrued at the time of this study, including 31 men and 24 women. The mean age was 65.6 years (range, 26–90). The following data were gathered for each patient in the trial: standard-of-care image data (diagnostic CT, ultra-low-dose CT acquired just prior to lung biopsy, and a postbiopsy computed radiographic image), experimental protocol image data (a DR image and a DE image, each described later), and biopsy data. Percutaneous CT-guided transthoracic biopsies were performed immediately following the DE and DR imaging examinations to provide definitive diagnoses of the lung lesions. Biopsies were performed either by fine-needle aspiration and cytologic examination or by core biopsy and histologic examination (18).

DE Imaging System

The DE imaging system was developed in collaboration with Carestream Health Inc. (Rochester, NY). The basic platform for the prototype was a modified Kodak RVG 5100 DR chest stand (Carestream Health Inc., Rochester, NY)
An acquisition workstation controls generator technique setting, filter selection, detector acquisition parameters, and data transfer. Modifications include a high-performance flat-panel detector (Pixium-4600; Trixell, Moirans, France) with a thallium-activated cesium iodide scintillator and 143-μm pixel pitch. A system for cardiac-gated image acquisition was implemented, using a fingertip pulse oximeter to trigger x-ray exposure coincident with diastole, thus reducing cardiac motion artifacts. DE images were acquired according to optimal techniques identified in previous work (13,14). For example, for the “average” patient size (24-cm chest thickness), kVp<sub>low</sub> = 60 kVp (2.5-mm aluminum total filtration), and kVp<sub>high</sub> = 120 kVp (4.5-mm aluminum + 0.6-mm silver total filtration). The high-energy filter was selected to “harden” the beam, reduce spectral overlap between the low- and high-energy projections, and thereby improve contrast in the resulting DE image. Such was the subject of considerable investigation in previous works (13,14).

Added filtration for the DR image was typical of that in conventional digital chest radiography. Three different filters were therefore required for the low-energy, high-energy, and DR exposures and were implemented using a computer-controlled, multiposition filter wheel within the collimator. The dose (ie, imparted energy [19]) was computed separately controlled, multiposition filter wheel within the collimator. DR exposures and were implemented using a computer-

Figure 1. Illustration of dual-energy (DE) image processing, registration (Reg), decomposition (Decomp), and display. Processing steps include offset, gain, and defect correction; deformable registration; weighted log-subtraction decomposition; and the transformation of pixel values to log-exposure space for display. The photograph at the right shows the experimental setup for observer studies, with images displayed on the left and the rating scale on the right. Bone, bone decomposition; Comp, composite equivalent radiograph; HE, high-energy projection; LE, low-energy projection; Soft, soft-tissue decomposition.

(13). An acquisition workstation controls generator technique setting, filter selection, detector acquisition parameters, and data transfer. Modifications include a high-performance flat-panel detector (Pixium-4600; Trixell, Moirans, France) with a thallium-activated cesium iodide scintillator and 143-μm pixel pitch. A system for cardiac-gated image acquisition was implemented, using a fingertip pulse oximeter to trigger x-ray exposure coincident with diastole, thus reducing cardiac motion artifacts. DE images were acquired according to optimal techniques identified in previous work (13,14). For example, for the “average” patient size (24-cm chest thickness), kVp<sub>low</sub> = 60 kVp (2.5-mm aluminum total filtration), and kVp<sub>high</sub> = 120 kVp (4.5-mm aluminum + 0.6-mm silver total filtration). The high-energy filter was selected to “harden” the beam, reduce spectral overlap between the low- and high-energy projections, and thereby improve contrast in the resulting DE image. Such was the subject of considerable investigation in previous works (13,14).

Added filtration for the DR image was typical of that in conventional digital chest radiography. Three different filters were therefore required for the low-energy, high-energy, and DR exposures and were implemented using a computer-controlled, multiposition filter wheel within the collimator. The dose (ie, imparted energy [19]) was computed separately controlled, multiposition filter wheel within the collimator. DR exposures and were implemented using a computer-

Figure 1. Illustration of dual-energy (DE) image processing, registration (Reg), decomposition (Decomp), and display. Processing steps include offset, gain, and defect correction; deformable registration; weighted log-subtraction decomposition; and the transformation of pixel values to log-exposure space for display. The photograph at the right shows the experimental setup for observer studies, with images displayed on the left and the rating scale on the right. Bone, bone decomposition; Comp, composite equivalent radiograph; HE, high-energy projection; LE, low-energy projection; Soft, soft-tissue decomposition.

(13). An acquisition workstation controls generator technique setting, filter selection, detector acquisition parameters, and data transfer. Modifications include a high-performance flat-panel detector (Pixium-4600; Trixell, Moirans, France) with a thallium-activated cesium iodide scintillator and 143-μm pixel pitch. A system for cardiac-gated image acquisition was implemented, using a fingertip pulse oximeter to trigger x-ray exposure coincident with diastole, thus reducing cardiac motion artifacts. DE images were acquired according to optimal techniques identified in previous work (13,14). For example, for the “average” patient size (24-cm chest thickness), kVp<sub>low</sub> = 60 kVp (2.5-mm aluminum total filtration), and kVp<sub>high</sub> = 120 kVp (4.5-mm aluminum + 0.6-mm silver total filtration). The high-energy filter was selected to “harden” the beam, reduce spectral overlap between the low- and high-energy projections, and thereby improve contrast in the resulting DE image. Such was the subject of considerable investigation in previous works (13,14).

Added filtration for the DR image was typical of that in conventional digital chest radiography. Three different filters were therefore required for the low-energy, high-energy, and DR exposures and were implemented using a computer-controlled, multiposition filter wheel within the collimator. The dose (ie, imparted energy [19]) was computed separately controlled, multiposition filter wheel within the collimator. DR exposures and were implemented using a computer-

(13) An acquisition workstation controls generator technique setting, filter selection, detector acquisition parameters, and data transfer. Modifications include a high-performance flat-panel detector (Pixium-4600; Trixell, Moirans, France) with a thallium-activated cesium iodide scintillator and 143-μm pixel pitch. A system for cardiac-gated image acquisition was implemented, using a fingertip pulse oximeter to trigger x-ray exposure coincident with diastole, thus reducing cardiac motion artifacts. DE images were acquired according to optimal techniques identified in previous work (13,14). For example, for the “average” patient size (24-cm chest thickness), kVp<sub>low</sub> = 60 kVp (2.5-mm aluminum total filtration), and kVp<sub>high</sub> = 120 kVp (4.5-mm aluminum + 0.6-mm silver total filtration). The high-energy filter was selected to “harden” the beam, reduce spectral overlap between the low- and high-energy projections, and thereby improve contrast in the resulting DE image. Such was the subject of considerable investigation in previous works (13,14).

Added filtration for the DR image was typical of that in conventional digital chest radiography. Three different filters were therefore required for the low-energy, high-energy, and DR exposures and were implemented using a computer-controlled, multiposition filter wheel within the collimator. The dose (ie, imparted energy [19]) was computed separately controlled, multiposition filter wheel within the collimator. DR exposures and were implemented using a computer-

### Figure 1

Illustration of dual-energy (DE) image processing, registration (Reg), decomposition (Decomp), and display. Processing steps include offset, gain, and defect correction; deformable registration; weighted log-subtraction decomposition; and the transformation of pixel values to log-exposure space for display. The photograph at the right shows the experimental setup for observer studies, with images displayed on the left and the rating scale on the right. Bone, bone decomposition; Comp, composite equivalent radiograph; HE, high-energy projection; LE, low-energy projection; Soft, soft-tissue decomposition.

For the DE image, the total dose was ε<sub>DE</sub> = ε<sub>low</sub> + ε<sub>high</sub>, and the “dose allocation” was such that the fraction of dose from the low-kVp projection was approximately 0.33 of the total (ie, ε<sub>low</sub>/ε<sub>DE</sub> = 0.33), previously determined to be optimal for soft tissue image quality (13). The DE image dose was equivalent to that of a conventional PA DR image (ie, ε<sub>DE</sub> = ε<sub>DR</sub>). Such was accomplished by (1) fixing ε<sub>DR</sub> to that of current clinical technique (1 mR to the detector, etc, as described previously) and (2) adjusting mAshigh such that the allocation was approximately 0.33 while ensuring that ε<sub>DE</sub> = ε<sub>DR</sub>.

DE images were processed and decomposed as illustrated in Figure 1. Offset and gain corrections were based on 25 averaged dark fields and 25 averaged flood fields (acquired at about 50% sensor saturation), respectively. Prior to DE image decomposition, the high-energy image was automatically registered to the low-energy image by means of deformable registration on the basis of mutual-information maximization and a morphologic pyramid (20). The average vector magnitude displacement of high-energy projections (averaged over all 55 cases) was approximately 3.1 mm, corrected using the deformable registration technique. A complete description and quantitative characterization of the registration technique is in progress. The low-energy image and the (registered) high-energy image were decomposed into soft tissue and bone images by log-weighted subtraction:

\[
\ln(t_{\text{soft}}) = \ln(t^H) - w_1 \ln(t^L),
\]

(1a)

\[
\ln(t_{\text{bone}}) = \ln(t^H) + w_b \ln(t^L),
\]

(1b)

and

\[
\ln(t_{\text{comp}}) = \ln(t^H) + w_c \ln(t^L),
\]

(1c)

where the tissue cancellation parameters (w<sub>a</sub>, w<sub>b</sub>, and w<sub>c</sub>) were selected according to previous work (15) that identified...
optimal settings (dependent on kVp selection) and were qualitatively validated during image decomposition by an expert observer (who was not among the observers in the performance evaluation described later). For this initial study, values of $w_s = 0.20$, $w_b = 0.57$, and $w_c = 0.9$ were identified as optimal and were judged to give acceptable tissue cancellation across all 55 patients. Future work will include more advanced means of patient-specific, automatically selected, spatially varying parameter selection and will consider the incorporation of $w_s$ and $w_b$ as parameters that may be freely varied by radiologists in a manner analogous to window level and width.

The composite image (Eq. 1c) is intended to be nearly identical to a conventional radiograph, and future clinical implementations could present this image in combination with the soft tissue and bone images, potentially boosting the performance of DE imaging further. For the purposes of the present study, only the soft tissue and bone images were included in the DE image set in the observer tests.

Soft tissue and bone images were transformed to log-exposure space, and window level and width settings were adjusted qualitatively by the same expert observer to yield display-ready images for each case. Example DE images are shown in Figure 2a and b. For the observer studies described here, the DE images were presented as a “two-slice volume.” Observers were able to scroll easily (using a mouse wheel) between the soft tissue and bone images, a feature that allowed readers to quickly evaluate corresponding locations in the images (as opposed to, for example, side-by-side display). The composite image was not used, because the objective of the present study was to evaluate the diagnostic performance associated with the DE decompositions (rather than the combined information of composite and decomposed images).

DR image processing was consistent with that in clinical DR systems based on the same imaging platform (Kodak RVG-5100 DR chest stand; Carestream Health Inc.), with tissue equalization, edge enhancement, and other postprocessing parameters defined in the associated WIISE and Eclipse image-processing software (Carestream Health Inc.), with parameters therein proprietary to the manufacturer. DR images were acquired on the same imaging prototype immediately (within a few seconds) following DE image acquisition. The DR image acquisition technique was 120 kVp (1 mm aluminum + 0.2 mm copper), with mAs interpolated from the technique chart in Shkumat et al (13) such that exposure to the detector was 1 mR (eg, approximately 1.6 mAs for an average [24-cm-thick] patient). For each patient, the total dose for the DE image and DR image were equivalent (eg, 0.11 mGy for 24-cm chest thickness). Offset and gain corrections, transformation to log-exposure space, and window level and width adjustment for the DR images were the same as for the DE images. Both DE and DR images were therefore acquired using optimal techniques and at an equivalent dose, allowing a reasonably fair comparison between the two modalities. As mentioned previously, image processing and display were similarly consistent between the two. Such included image-processing filters and tone scaling that may have favored the DR images, because such techniques have been thoroughly investigated over the past decade for DR images but are yet to be fully optimized for DE images (the subject of ongoing work). The specific image-processing filters and tone-scaling techniques were proprietary to the manufacturer (Kodak Health Imaging Systems, now Carestream Health Inc.). They are standard to commercially available DR image-reading stations, but additional technical details are proprietary and not available.
Table 1
Satisfaction Rating Scale and Descriptions

<table>
<thead>
<tr>
<th>Score</th>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Very satisfied</td>
<td>The abnormality is <strong>perfectly obvious</strong> and easily characterized.</td>
</tr>
<tr>
<td>8</td>
<td>Satisfied</td>
<td>The abnormality is <strong>visible</strong> and can be well characterized.</td>
</tr>
<tr>
<td>7</td>
<td>Neither satisfied nor dissatisfied</td>
<td>The abnormality is <strong>reasonably well seen</strong> and characterized.</td>
</tr>
<tr>
<td>6</td>
<td>Dissatisfied</td>
<td>The abnormality is visible, but detection and characterization of subtle features are <strong>a bit challenging</strong>.</td>
</tr>
<tr>
<td>5</td>
<td>Very dissatisfied</td>
<td>The abnormality <strong>could be overlooked</strong> or mischaracterized.</td>
</tr>
</tbody>
</table>

This table was displayed to observers on a second monitor during all tests, as shown in Figure 1.

An example DR image for the same patient as in Figure 2a and b is shown in Figure 2c.

Observer Study

Because the initial data set was not well suited to ROC analysis (specifically because of a lack of true-negative cases in the patient cohort), a simple test based on diagnostic satisfaction was adopted as an initial evaluation of diagnostic performance. Although such tests do not provide measures of diagnostic sensitivity and specificity, they have been shown to provide objective differentiation of performance associated with the perception of subtle image quality factors (21). Such tests were deemed sufficient to establish whether there was a performance advantage for DE or DR imaging on the basis of the initial data set. Future evaluation on the basis of the completed trial will include sensitivity and specificity (ROC analysis). In the event that the full patient cohort similarly lacks true-negative cases, we hope to overcome this limitation by performing an ROC study using images of the left or right half of the chest (because patients typically exhibit nodules in one lung or the other, but not both). Each reader scored each image on a nine-point satisfaction scale, as shown in Table 1. Scales with fewer (five) and more (up to 100) levels of rating were considered in preliminary studies, but nine levels were found to be tolerated well by observers and consistent with reasonably fine levels of image quality discrimination, as described by Van Metter and Foos (21).

Five expert observers (three radiology fellows and two radiology staff members, each a specialist in thoracic imaging) were independently presented with the images. The reading order was randomized, with one image presented at a time (either a DR image or a DE image [two-slice volume, described previously]). The observer was asked to rate each image in terms of his or her ability to detect and characterize abnormalities according to the nine-point satisfaction scale shown in Table 1. The scale was visible to observers at all times on a second monitor (illustrated in Fig 1).

The following methods were used to minimize observer bias in the image evaluation. The study was conducted in a clinical radiology reporting room with subdued lighting on diagnostic-quality, monochrome liquid crystal display monitors (AM-QX21-A9300; National Display, San Jose, CA). The monitors were adjusted to meet the Digital Imaging and Communications in Medicine grayscale standard. To reduce inter-reader variability and simplify image controls, window width and level and magnification were fixed. In this initial study, given that all cases were positive (and often conspicuous), such display controls would likely not have affected the results. Future tests using the completed trial data will allow individual readers to freely vary window width and level to better accommodate variability in reader preferences, particularly with respect to a broader spectrum of cases.

Preceding each observer test, a training session involving eight DE and seven DR images was conducted to familiarize the observers with the software and standardize their understanding of the rating scale. The training images were drawn from the pool of cases available at the time of the study and did not overlap with those used as test images. Given the differences between the two modalities (DE and DR imaging) and the fact that the observers were specialist thoracic radiologists, it was obvious whether a given image was a DE image or a DR image. Although this represents a potential source of bias, readers were asked to respond strictly with respect to their satisfaction as described by the rating scale, irrespective of the image modality. We used this approach together with randomization of the case reading order to minimize observer bias for or against a given modality. To examine intrareader consistency during the actual test, the first seven images displayed in each test were displayed again at the end of the test (without informing the observers that the images were repeated), with differences examined in terms of Wilcoxon’s signed-rank P value, as described later. The first seven images were rejected from the study, except for purposes of assessing intrareader variability; therefore, the analysis pertains to each DE and DR image shown once (no repeats). Repeat readings were found to be highly reproducible for expert observers in preliminary studies and would not add to the statistical power of the study.

Statistical Analysis

Fraction of Responses at or Above a Given Rating

The rating scale responses constitute ordinal data. The fraction of observer responses (F) at or above a given rating
(R) was plotted against rating scale, giving curves that ranged from zero to one on the vertical axis plotted against the ordinal rating scale (ranging from one to nine) on the horizontal axis, essentially a cumulative histogram of responses. Because each observer rated the images on the basis of a nine-point rating scale, we were able to calculate the fraction of responses at or above a given value of R. For example, the percentage rated R = 1 or higher was always 100%, the percentage rated R = 2 or higher was a bit less, and so forth. The plots summarize the results in a simple form, with “higher curves” corresponding to superior performance. The analysis is similar to that of Van Metter and Foos (21), who showed that the technique is capable of differentiating fairly subtle differences in image quality. Moreover, computing fractions at or above a given rating is an appropriate way to handle qualitative, ordinal data (as opposed to computing the mean value of R).

Error bars on such plots reflect a two-sided 95% confidence interval computed according to a binomial distribution (22), as described in the following. The corresponding error bars are asymmetric and appropriately bounded between zero and one. Each image was scored as either (1) at or above a certain rating or (2) below that rating, giving two mutually exclusive outcomes such that F (the fraction at or above a given rating) followed a binomial distribution. The upper bound of the confidence intervals was calculated as

\[ \sum_{k=0}^{N_d} \binom{N}{k} p_U^k (1 - p_U)^{N-k} = \frac{\alpha}{2} \]  

(2a)

and the lower bound as

\[ \sum_{k=0}^{N_d-1} \binom{N}{k} p_L^k (1 - p_L)^{N-k} = 1 - \frac{\alpha}{2}, \]  

(2b)

where \( N \) is the total sample size, \( N_d \) is the number of samples with the outcome of interest, \( p \) is the proportion of samples with the outcome of interest, \( p_U \) is the upper bound of \( p \), \( p_L \) is the lower bound of \( p \), and \( \alpha \) (commonly fixed at 5%) is the type 1 error (ie, the probability of rejecting the null hypothesis when in fact the null hypothesis is correct). The results were similarly interpreted in terms of the P value (ie, the probability of obtaining the measured results if the null hypothesis were correct, also commonly taken as a statistically significant difference for \( P < .05 \)). The data were complemented by 95% confidence intervals, as in Equations 2a and 2b, and plotted as error bars to indicate the limits within which the true difference between groups was expected to lie. The upper and lower bounds were calculated using the Newton-Raphson method in Matlab (The MathWorks, Natick, MA) to a precision of \( 1 \times 10^{-6} \).

### Statistical Significance: P Value From Wilcoxon’s Signed-Rank Test

The statistical significance in differences observed between DE and DR scores was evaluated in terms of the \( P \) value at a 95% level of significance from Wilcoxon’s signed-rank test, a nonparametric test suitable to paired ordinal data, assuming that all observations within a given modality are independent (23). The \( P \) value was calculated using the Matlab function “signrank,” accounting for both the sign and the magnitude of the difference in ratings. A one-sided hypothesis test was used (with the alternative hypothesis that DR scores would be significantly higher than DE scores), and the \( P \) values were corrected by a factor of .5 from that of the two-sided test.

To examine intrareader consistency, Wilcoxon’s signed-rank test was also used in relation to the first and last seven images in the reading study, repeated at the beginning and end of the test as described previously. In this case, the alternative hypothesis was that the two sets of scores would not be equivalent; therefore, a two-sided \( P \) value was calculated directly using the “signrank” function. Across five observers, all \( P \) values assessing intrareader consistency were greater than .05 (specifically, \( P = .125, .0625, .5313, \) and 1 for observers A to E, respectively), suggesting that there was no significant difference in observer readings at the beginning and end of the test (ie, that training appears
to have been sufficient, and there was no evidence of observer fatigue).

**Stratification of the Results**

Performance was analyzed for all cases pooled, as well as by post hoc stratification of the data according to lesion size, lesion density, chest thickness, gender, and location of the lesion. Lesion size was characterized as the greatest linear dimension as measured on CT, and the results were stratified as nodule (<3 cm) or mass (≥3 cm), consistent with typical clinical terminology. Lesion density was measured using the attenuation measurement tool on a picture archiving and communication system workstation (Fusion E-film 2.1; Merge Healthcare, Milwaukee, WI), and the results were stratified as solid (≥20 Hounsfield units [HU]) or nonsolid (<20 HU). Chest thickness was characterized as the anterior-posterior distance measured from the xiphoid process to the T9 thoracic vertebra taken from the axial computed tomographic image at this level, and the results were stratified as average (≤26 cm) or thick (>26 cm). Lesion location was determined according to the anatomic position with respect to lung and lobe (or mediastinum), and the results were stratified as right upper, left upper, right middle, left middle, right lower, left lower, and mediastinum. The numbers of cases overall and within each stratum are summarized in Table 2.

**RESULTS**

**All Cases Pooled**

The results for the 275 total ratings (five radiologists and 55 cases) for each of the two modalities (DE and DR imaging) are summarized in Figure 3. Individual case-by-case comparison of DE and DR imaging for each patient is shown in Figure 3a. In 41.5% of cases (114 of 275), the DE image was rated superior to the DR image by a difference of $R = 1$ or higher. In 38.9% of cases (107 of 275), the DE and DR images...
were rated equal. In 19.6% of cases (54 of 275), the DR image was rated superior. Further to this case-by-case examination, the proportion of cases for which one modality was superior (or equal) to the other as judged by three or more observers (of five) is plotted in Figure 3b. Such analysis yielded 55 responses (of 275) in which three or more observers agreed on the comparison between the two modalities. Of these, 36.4% (20 of 55) scored the DE image superior to the DR image, 36.4% (20 of 55) rated the DE and DR images equivalent, and 5.5% (3 of 55) rated the DR image superior to the DE image. In the remaining 12 cases (21.8%), a majority could not be reached regarding the superiority, equality, or inferiority between the two modalities. The fraction of images rated at or above a given rating score (Fig 3c) shows that DE images were rated consistently higher than DR images \((P < .001)\) in the detection and characterization of lung nodules.

**Stratification by Lesion Size**

The data were subsequently analyzed in terms of cases for which the lesion size was \(\leq 3\) and \(> 3\) cm (36 and 19 cases, respectively, as shown in Table 2). The rationale for the size demarcation is consistent with clinical convention, in which lesions \(< 30\) mm are termed nodules, whereas lesions \(\geq 30\) mm are termed masses. The results are shown in Figure 4. A statistically significant improvement in satisfaction was observed for DE imaging in each case \((P < .001\) for nodules \([\text{lesion size} < 3\text{ cm}]\), \(P = .0264\) for masses \([\text{lesion size} \geq 3\text{ cm}]\)). The advantage of DE imaging was more
pronounced for nodules, as seen from the distinctly separated curves in Figure 4a and the correspondingly smaller $P$ value.

**Stratification by Lesion Density**

Cases were alternatively stratified according to lesion density as solid ($\geq 20$ HU) and nonsolid (<20 HU), with 46 and 9 cases, respectively (Table 2, Fig 5). A statistically significant improvement in diagnostic performance was observed for DE imaging of solid lesions ($P < .001$). For nonsolid lesions, DE and DR scores were not significantly different overall ($P = .0968$). As mentioned previously, we classified nodules as solid for a computed tomographic attenuation $\geq 20$ HU, corresponding to typical clinical terminology and representing the prevalent phenotype for the most common histologic subtypes of lung cancer (adenocarcinoma, squamous-cell, and small-cell lung cancer).

Nonsolid lesions, on the other hand, such as bronchioloalveolar carcinoma (a subtype of adenocarcinoma), characteristically present as ground-glass opacities that typically take several years to grow. Such lesions are difficult to detect on chest radiography and are typically followed by annual (low-dose) chest CT. The superior detection of solid nodules observed for DE imaging in comparison to DR imaging is important in the detection of primary or metastatic disease of the thorax. The results showed no statistical difference in the detection of nonsolid nodules between DE and DR imaging ($P = .0968$), and the difficulty in detecting ground-glass opacities (for both DE and DR imaging) is an issue that will be investigated further using the completed trial data and modified DE imaging techniques (eg, noise reduction algorithms).

**Stratification by Chest Thickness**

Cases were further grouped on the basis of gender, with 31 men and 24 women. The results in Figure 7 suggest a significant improvement in diagnostic performance for DE imaging in each case ($P < .001$ for men, $P = .013$ for women). This is likely consistent with the trend for improved performance overall for DE imaging (Figure 3). The smaller level of improvement suggested for the subcohort of women (although still a statistically significant improvement) is possibly related to that observed for larger chest thickness (in this case breast tissue), which was correlated with a smaller improvement in diagnostic performance.

**Stratification by Region**

Cases stratified according to seven regions of the chest are listed in Table 2. Results grouped according to seven regions of the chest (the left and right apical, left and right middle, left and right lower, and mediastinal regions) are shown in
A significant improvement in diagnostic performance was observed for DE imaging in the left and right apical, left middle, and left lower regions. The results for the left and right apical regions are consistent with the hypothesis that DE imaging improves diagnostic quality by removing anatomic noise, in this case the clavicles and first and second ribs, which pose complex anatomic clutter and can significantly diminish conspicuity. The significant improvement observed for DE imaging in the left lower region was somewhat surprising, given that this region is challenged by a preponderance of soft tissue structures (the heart) and is most susceptible to cardiac motion artifacts.

**DISCUSSION**

DR imaging of the chest represents a cost-effective, widely available, low-dose modality used for a broad spectrum of applications, ranging from bedside exams to the initial examination and diagnosis of lung disease. Still, it is known to suffer in sensitivity for the detection of subtle lesions, limited primarily by a lack of conspicuity caused by superimposed anatomic structures. DE imaging, which reduces anatomic clutter by selectively removing material-specific components from the image, showed a significant improvement in the satisfaction associated with the detection and characterization of pulmonary nodules. The results were based on an initial evaluation in 55 patients drawn from an ongoing trial, providing initial investigation of diagnostic performance and supporting the hypothesis that DE imaging boosts lesion conspicuity. That this performance improvement is achieved without an increase in radiation dose is particularly encouraging (ie, the total dose for the DE image is equivalent to that for single PA DR image). We attribute the boost in diagnostic satisfaction primarily to the nature of DE imaging itself (ie, reduction of anatomic noise) but also to a system design in the experimental prototype on the basis of first principles of performance metrology (contrast, noise, noise-equivalent quanta, and task-dependent detectability) and system optimization (eg, differential added filtration, selection of kVp pair, and optimal allocation of dose between low-energy and high-energy projections) (13–17). Furthermore, the availability of a composite radiograph might improve the overall performance of a DE imaging system. Because the composite image presents information comparable to that of a conventional radiograph, the DE imaging system could present radiologists with the familiar context of the (composite) radiograph, within which they could examine tissue-specific characteristics via the soft tissue and/or bone images.

The differences in satisfaction between DE and DR imaging are shown in Figures 3 to 8, with qualitative differences illustrated in the patient images in Figures 9 to 11. For example, the difference observed for small, solid nodules (<3 cm diameter, ≥20 HU density) in Figures 4 and 5 is illustrated qualitatively in Figure 9. The improved visualization of small, solid nodules by virtue of rib cancellation (particularly rib crossings) is clear. Figures 10 and 11 similarly illustrate the improved visibility in the lung apex, in which the clavicles and first and second ribs present confounding clutter that is significantly reduced in the DE soft tissue image. That the difference in performance was greatest for small, solid nodules located in the apex is particularly valuable, because these characteristics represent precisely the most challenging (ie, small), important (ie, solid, more likely malignant), and frequent (ie, location in lung apex) cases and the areas most in need of improvement in chest imaging.

The initial results are encouraging, although the study is not without its limitations. First, the number of cases was
low, particularly for certain stratifications (e.g., anatomic region), because these early data constitute an initial study from an ongoing trial. Furthermore, the stratifications within the data were post hoc, and although the overall study was prospectively designed to investigate the difference in performance between DE and DR imaging, the data pertaining to individual strata (i.e., nodule size, nodule density, chest thickness, gender, and nodule location) should be considered.

Figure 8. Diagnostic satisfaction in dual-energy (DE) and digital radiographic (DR) image readings stratified by lesion location. The curves show the fraction of responses at or above a given rating score for lesion located in (a) right upper, (b) left upper, (c) right middle, (d) left middle, (e) right lower, (f) left lower, and (g) mediastinal regions. The error bars reflect 95% confidence intervals.

Figure 9. Illustration of (a) dual-energy soft tissue, (b) dual-energy bone, and (c) digital radiographic images for a case exhibiting a small, solid nodule (0.8 cm, 58.4 Hounsfield units, benign lymphoid tissue), marked by the arrows. The bone image exhibits some residual soft tissue attributable to incomplete cancellation as well as motion artifact, which is more obvious in this case because of the cardiac motion artifact.
“hypothesis generating” in the sense of retrospective analysis. Furthermore, the strata were distinct and independent (eg, grouping small or solid or apical nodules, but not small, solid, and apical nodules), and joint grouping was not examined.

Second, the satisfaction tests on the basis of a nine-point scale represent a fairly coarse method. For example, one may justifiably wonder whether such a test comparing DR to screen-film images would be sufficiently sensitive to demonstrate a statistically significant difference. That a performance improvement was in fact observed for DE imaging in this study suggests a considerable improvement in conspicuity. Statistical significance was evaluated in terms of $P$ values obtained from Wilcoxon’s signed-rank tests. A possible limitation associated with clustering effects within ratings for the same patient across observers will be investigated with the benefit of more cases and could be addressed using a modified version of Wilcoxon’s signed-rank test for clustered data in future work (23). A more comprehensive characterization of diagnostic sensitivity and specificity for DE imaging would include evaluation of the ROC curve, to be investigated in future work.

A third limitation of the study involves the DE image processing and decomposition. Although the DR images were postprocessed according to techniques established for clinical DR imaging that were optimal, the DE image post-processing was fairly simple. The DE images used in the observer study used the simple image-processing and decomposition techniques illustrated in Figure 1, specifically, single-point offset and gain corrections, nonoptimized registration of low- and high-energy projections, and simple log-weighted subtraction according to a scalar tissue cancellation parameter. Future work will improve on each of these as well as optimal rendering, tissue equalization, and edge enhancement and presumably further increase the performance of DE imaging. In addition, noise reduction algorithms (eg, anticorrelated noise reduction, smoothing of the high-energy image) will be implemented in the decomposition of images for tests using the full patient cohort.

The clinical role of DE imaging of the chest is yet to be fully determined. The studies herein relate only to a PA exam. Whether clinical implementation would use a lateral DE or DR image is under investigation. Furthermore, the role of DE imaging with respect to volumetric imaging modalities, such as tomosynthesis and CT, is subject to future analysis of diagnostic sensitivity, specificity, work flow, cost-effectiveness, and clinical implementation. Also, the effect of using DE imaging as an adjunct to DR imaging is of clinical interest and will be evaluated in the future study.
Although DE imaging requires the assessment of a greater number of images compared to DR imaging, this will be compensated for by an increased level of diagnostic confidence in lesion detection, which in turn should translate into the earlier detection of disease. In addition, the characteristics of DE imaging offer promise in other areas of thoracic disease, such as the earlier detection of airspace disease (pneumonia in patients with fever of unknown origin), the improved demonstration of airway disease (bronchiectasis in patients with chronic productive cough), and the improved visualization of catheters, tubes, and pneumothoraces in patients in intensive care units (24).

DE imaging, with diagnostic performance exceeding that of DR imaging at equivalent radiation dose, could represent a new normal means of chest projection imaging, because an image equivalent to a DR image may always be decomposed, but more important, soft tissue images present subtle lesions more conspicuously by virtue of reduced anatomic clutter. Although the observer response (rating) pertained to the combined DE image set (ie, the soft tissue and bone images considered together, rather than each rated individually), it was clear that the soft tissue image was the more important in nodule detection, while the bone image presented complementary information regarding characterization (eg, calcification). For this patient cohort in particular (drawn from a clinical patient population referred for a lung nodule biopsy), there were few cases exhibiting calcified nodules; therefore, the bone images were likely used less than in a general screening population (in which the frequency of calcified nodules would presumably be greater). Furthermore, the bone image could provide diagnostic value regarding bone differentiation of rib metastasis from fracture and improved visualization of fine bony detail to exclude cortical invasion. Future studies will also examine the use of DE imaging as an adjuvant to low-dose CT in lung nodule evaluation, maintaining the high sensitivity associated with low-dose CT but possibly improving on specificity through better quantification of subtle nodule calcification using DE imaging.

ACKNOWLEDGMENTS

The authors acknowledge the expert assistance from the following radiologists: Dr. Sidney Sussman, Dr. Leon Zelovitzky, Dr. Demetris Patsios, Dr. Ute Wagnetz, Dr. Priya Gupta, and Dr. Andre Pereira, University Health Network, Toronto, ON, Canada.

REFERENCES