Small Lung Cancers: Improved Detection by Use of Bone Suppression Imaging—Comparison with Dual-Energy Subtraction Chest Radiography¹

Feng Li, MD, PhD
 Roger Engelmann, MS
 Lorenzo L. Pesce, PhD
 Kunio Doi, PhD
 Charles E. Metz, PhD
 Heber MacMahon, MB, BCh

Purpose:

To determine whether use of bone suppression (BS) imaging, used together with a standard radiograph, could improve radiologists' performance for detection of small lung cancers compared with use of standard chest radiographs alone and whether BS imaging would provide accuracy equivalent to that of dual-energy subtraction (DES) radiography.

Materials and Methods:

Institutional review board approval was obtained. The requirement for informed consent was waived. The study was HIPAA compliant. Standard and DES chest radiographs of 50 patients with 55 confirmed primary nodular cancers (mean diameter, 20 mm) as well as 30 patients without cancers were included in the observer study. A new BS imaging processing system that can suppress the conspicuity of bones was applied to the standard radiographs to create corresponding BS images. Ten observers, including six experienced radiologists and four radiology residents, indicated their confidence levels regarding the presence or absence of a lung cancer for each lung, first by using a standard image, then a BS image, and finally DES soft-tissue and bone images. Receiver operating characteristic (ROC) analysis was used to evaluate observer performance.

Results:

The average area under the ROC curve (AUC) for all observers was significantly improved from 0.807 to 0.867 with BS imaging and to 0.916 with DES (both $P < .001$). The average AUC for the six experienced radiologists was significantly improved from 0.846 with standard images to 0.894 with BS images ($P < .001$) and from 0.894 to 0.945 with DES images ($P = .001$).

Conclusion:

Use of BS imaging together with a standard radiograph can improve radiologists' accuracy for detection of small lung cancers on chest radiographs. Further improvements can be achieved by use of DES radiography but with the requirement for special equipment and a potential small increase in radiation dose.

©RSNA, 2011

¹From the Department of Radiology, MC-2026, University of Chicago, 5841 S Maryland Ave, Chicago, IL 60637. Received January 28, 2011; revision requested March 31; revision received June 1; accepted June 21; final version accepted July 20. Supported in part by the United States Public Health Service (grants CA61625, CA09119, and EB00341) and a research grant from Riverain Medical.

Address correspondence to F.L. (e-mail: feng@uchicago.edu).

©RSNA, 2011

Dual-energy subtraction (DES) techniques can distinguish bone from soft tissue, and DES chest radiography has been shown by several observer performance studies to improve the ability to detect and characterize lung nodules (1–5). However, the disadvantages of DES radiography include the requirement of specialized equipment and a small potential increase in radiation dose. Recently, a new image processing system for bone suppression (BS) imaging has been developed, which can suppress the conspicuity of bones by using a software-only approach. We found that the BS image was very similar in appearance to the soft-tissue image produced by using energy subtraction, although bone subtraction was less complete in the apical areas and local enhancement was more pronounced. The purpose of this observer study was to determine whether use of BS imaging, used together with a standard radiograph, could improve radiologists' performance for detection of small lung cancers compared with use of a standard chest radiograph alone and whether BS imaging would provide accuracy equivalent to that of DES radiography.

Advances in Knowledge

- Bone suppression (BS) imaging, used together with a standard chest radiograph, has the potential to improve radiologists' ability to detect small nodular lung cancers (a significant improvement in detection of approximately 0.06 in the area under the receiver operating characteristic curve [AUC] in this observer study) without any special equipment or increase in radiation dose.
- Further improvements can be achieved by use of dual-energy subtraction radiography (a further improvement of approximately 0.05 in the AUC in this study) but with the requirement for special equipment and a potential small increase in radiation.

Materials and Methods

Institutional review board approval was obtained, and the requirement for informed patient consent was waived. Our study was compliant with the Health Insurance Portability and Accountability Act. Funding for this research was provided by Riverain Medical (Miamisburg, Ohio), and one of the authors (H.M.) is a consultant to the company. However, the company did not have any control over the design of the experiment or case selection, which was at the discretion of the first author. In addition, the authors had full control over the data during all phases of the study.

Patient Database and Reference Standard

The source database of images used in this study consisted of a combination of two separate databases. The first database included 94 patients with 105 peripheral nodular cancers (with 11 patients having two primary cancers in different lungs) on standard and DES chest radiographs, with all cancers confirmed with computed tomographic (CT) results, and were obtained from the clinical workload at the University of Chicago Medical Center from January 2004 to September 2008. The second database, which has been used in our previously reported studies (5,6), was obtained in the same medical center and consisted of standard and DES chest radiographs from 19 patients with 22 peripheral nodular cancers. The location of each cancer on the standard and DES chest radiographs was identified by consensus of two radiologists (F.L. and H.M., with 15 and 30 years of experience, respectively, in interpretation of chest radiographs), with all cancers confirmed with CT results. Among the total 127 nodular cancers, 101 were confirmed by using surgery or biopsy as the reference standard, and 22 were diagnosed by using follow-up. The 127

cancers were graded for subtlety (from extremely subtle to extremely obvious with a 1–10 scale) on the standard posteroanterior (PA) chest radiographs by the same two radiologists independently, and the numeric ratings were averaged. The mean diameter (average of the length and width) of each nodular cancer on the standard PA images was measured by one radiologist (F.L.). These two radiologists did not participate in the observer study.

DES and BS Chest Radiographs for Observer Study

The exclusion criteria established by two radiologists (H.M. and F.L.) for selection of chest radiographs in the observer study included (a) obvious nodules (subtlety ratings of 5.5 or more), (b) the existence of potentially confusing underlying abnormalities such as diffuse lung disease, and (c) poor image quality. Prior to the observer study, and prior to seeing the results of BS in these cases, 50 patients (22 men, 28 women; mean age, 70 years; mean age of men, 74 years; mean age of women, 67 years; age range, 44–98 years) with 55 nodular cancers on DES chest radiographs were selected (as a subset from the 127 nodular cancers described previously) by the two radiologists for inclusion in the observer study. Cases were included on

Published online before print

10.1148/radiol.11110192 Content code: CH

Radiology 2011; 261:937–949

Abbreviations:

AUC = area under the ROC curve

BS = bone suppression

DES = dual-energy subtraction

PA = posteroanterior

ROC = receiver operating characteristic

Author contributions:

Guarantors of integrity of entire study, F.L., H.M.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, F.L., K.D.; clinical studies, F.L., R.E., K.D., H.M.; statistical analysis, F.L., L.L.P., C.E.M.; and manuscript editing, all authors

Potential conflicts of interest are listed at the end of this article.

Implication for Patient Care

- BS imaging can improve detection accuracy for small lung cancers.

the basis of the subtlety of the cancer and not because of any other factors that might affect the results on BS or DES images. The mean diameter of the 55 lung cancers was 20 mm (median, 21 mm; range, 9–30 mm; seven cancers smaller than 10 mm) on standard PA images. Among the 55 nodular cancers, 14 had not been detected or described in the original radiology reports, while 41 had been detected. Thirty cancers had subtlety ratings of 1 to 3 (very subtle), and 25 cancers had subtlety ratings of 3.5 to 5.0 (subtle to relatively obvious). We did not include lateral radiographs in this observer study, because among the 55 nodular lung cancers, 42 were not visible, even in retrospect, on the lateral view, and among the remaining 13 visible cancers, only two were more conspicuous on the lateral than on the PA view on the basis of the consensus of the same two radiologists. Single view PA or anteroposterior chest radiographs are commonly interpreted without lateral views.

Twenty-eight lung cancers were located in the right lung (19 in the right upper lobe, three in the right middle lobe, and six in the right lower lobe), and 27 were located in the left lung (22 in the left upper lobe and five in the left lower lobe). Among the 55 cancers, 49 (23 adenocarcinomas, seven squamous cell carcinomas, two large cell carcinomas, 16 unspecified non-small cell carcinomas, and one small cell carcinoma) were pathologically confirmed, and six were diagnosed by using image follow-up.

Thirty control subjects (15 men, 15 women; mean age, 60 years; mean age of men, 62 years; mean age of women, 58 years; age range, 37–79 years) without cancer were used, with examinations performed with the same DES radiography system for various indications. Radiographs from eight of these subjects without lung cancers included incidental abnormal findings (other than cancer) that were similar to findings in the selected cancer patients, including pleural abnormalities and very small calcified benign nodules. The findings for all 30 control subjects were confirmed at CT.

All radiographs in the 50 patients with cancer and the 30 patients without cancer used in the observer study were obtained with a single-exposure DES computed radiography system (Fuji Medical Systems, Stamford, Conn) by using 110 kVp and 2.5–16 mAs, according to patient size. A new BS imaging processing system (SoftView, version 2.0; Riverain Medical) was applied to the 80 standard PA radiographs to create corresponding BS images. SoftView, which has received Food and Drug Administration clearance for marketing in the United States, is a software solution for BS which employs image normalization, feature extraction, and regression networks to predict the bone image. By subtracting the predicted bone image, a soft-tissue image is formed. Because SoftView is a proprietary system, details of the method have not been publicly disclosed and were not available to the investigators in this study.

Observer Study

Three 2-megapixel monochrome liquid crystal display monitors (MDL 2004A; Totoku Electric, Tokyo, Japan) in our clinical reading area were used in the observer study. These monitors were calibrated to Digital Imaging and Communications in Medicine standards. The reading room is dimly lit with indirect lighting, and the monitors are calibrated regularly. Observers were advised that cases in the study might show no cancers, a unilateral cancer, or bilateral cancers. However, observers were not informed of the number of patients with or without cancers. Confidence bars were shown, one under each lung, for each image set. Observers indicated their confidence levels regarding the presence of a nodular cancer for each lung separately by clicking with the mouse to mark an appropriate point on the confidence bars. This was done first by reading standard PA radiographs only, then reading standard and BS images, and finally reading standard and DES images (soft-tissue and bone images). Each successive mark was displayed (each a different color) until all three reading conditions had been completed for a given case. The standard PA radio-

graphs were shown on the first liquid crystal display monitor, and the same confidence bars were used sequentially to obtain sets of ratings for each of the three display configurations for each patient. The additional BS images were shown on the second monitor (and ratings were obtained by using this configuration) but were not shown in the third display configuration (wherein only the standard and the DES images were displayed and an additional set of ratings was obtained). This sequential viewing test design was used in this observer study because we believe that this is how the images would be viewed clinically (ie, standard chest radiography followed by BS imaging or DES). However, we do not expect that both BS imaging and DES would likely be used in the same examination in practice. Windowing values were controlled manually by the observers, and a magnification tool was available. No clinical parameters (such as age or sex) were provided to the observers.

The 10 observers who participated in this observer study were divided into two groups corresponding to a limited reading time (group 1) or an unlimited reading time (group 2). Group 1 (five observers) consisted of three general radiologists with a mean experience of 19 (13, 22, and 23) years and two radiology residents (1st and 2nd year). The reading time for group 1 for each patient (two lungs) was limited to 12 seconds for each reading condition (ie, standard image, standard image plus BS image, and standard plus DES images). The maximum reading time was therefore 48 minutes for the total of 80 patients. This limited reading time used in group 1 was based on our estimate but may not reflect the actual reading time in clinical practice. The five observers in group 2 consisted of two chest radiologists who each had 7 years of experience, one general radiologist with 3 years of experience, and two radiology residents (both 4th year). The reading time in group 2 was not limited but was measured for each set of images. In group 2, the average reading time for four of five observers was 48 minutes (range, 39–58 minutes) in

total and 36 seconds for each patient, including 14 seconds for the standard image, 11 seconds for the BS image, and 11 seconds for the DES soft-tissue and bone images. Only one observer used a substantially longer time, 139 minutes in total and an average of 104 seconds for each patient. However, the accuracy of this observer for the three conditions was very similar to that of the other observers.

Observers were instructed regarding the nature of the task and appropriate use of the confidence rating scales. The confidence scales (0–1.00) regarding the presence of a nodular cancer were calculated as a percentage from the leftmost position (definitely absent, 0.01) to the rightmost (definitely present, 0.99), although the numbers were not shown on the confidence bars. In this test, the observers could change their confidence levels before moving on to the next image set; however, after seeing the BS or DES images, the observers could not go back and revise previous entries. We informed the observers regarding the size range of the nodules in the test. The observers were also instructed that scars or calcified granulomas might be present, but only suspected cancers should be scored. We provided a training session before the test with eight patients who were not used in the study, which included four representative examples with a nodular cancer and four without cancer. In these training examples, the correct diagnosis was shown after each case had been completed by use of standard radiographs, standard image plus BS image, and standard image plus DES radiographs. All of the 10 observers had previous experience with reading DES radiographs, but none of them had interpreted BS images before.

Data Analysis

The confidence level ratings from each observer in the two groups for 80 patients (all 160 lungs, 80 left lungs or 80 right lungs separately) were analyzed by use of receiver operating characteristic (ROC) methodology, and a quasi-maximum likelihood estimation of the binormal distribution was fitted to the

radiologists' confidence ratings (7). The statistical significance of the difference in the values of the area under the ROC curve (AUC) between observer readings with standard image versus standard plus BS image or DES images and between standard image plus BS image versus standard image plus DES images were tested by use of the Dorfman-Berbaum-Metz method (8,9), which included both reader variation and case sample variation (multireader multicase) by means of an analysis of variance approach (DBM MRMC, version 2.2; <http://rocweb.bsd.uchicago.edu/MetzROC>). The AUC was fit by use of the conventional binormal model. Because ROC area can be interpreted as sensitivity averaged over all specificities, larger AUCs indicate greater diagnostic accuracy (10). The AUCs were also compared for observers with more or less experience. The analysis also included a 95% confidence interval for the statistical significance of an estimated difference. Thus, in the context of no significant difference, a narrow 95% confidence interval for treatment differences would indicate that any true difference was small (11).

Second Observer Study

To address a concern regarding possible bias that might be produced by sequential reading of different versions of the same cases, a second observer study was performed. This second test was performed 1 year after the first by using a random subset of the same readers and cases.

In this second test, 80 pairs of identical standard PA radiographs with and without subtle nodules were presented sequentially in random order to the observers. These identical pairs of radiographs were randomly intermingled with additional pairs of cases in which the second image had been altered slightly to motivate the observers by rewarding them for careful scrutiny of the second image. Each motivator case consisted of two images, the first of which showed a very subtle nodule, and the second of which showed a more obvious version of the same nodule (by using Adobe Photoshop CS4 Extended Edition [Adobe Systems, San Jose, Calif] for

enhancement or suppression). There was a noticeable difference in the visibility of the nodule on the second image in the motivator cases, which served to hold the observers' attention and caused them to change their confidence rating for the presence of a cancer in these cases. This difference in nodule visibility was reflected in the observers' markedly improved detection accuracy when using the nodule-enhanced image compared with the nodule-suppressed image. However, the results for these motivator cases were excluded from the main analysis.

We randomly divided the original 100 patients, including the 80 patients with and without nodular cancers used in the main observer study and the 20 additional patients with motivator images, into two parts. Only part 1 with 50 cases, which included 39 of the same cases as the main study (26 patients with 30 cancers and 13 patients without cancers) and 11 motivator cases, was ultimately used in this supporting study.

Five of the same attending radiologists who participated in the original observer test (three general radiologists from group 1 and two radiologists, one general and one thoracic, from group 2) participated in this second supporting observer study. As in the original test, observers were asked to indicate their confidence level regarding the presence of a nodular cancer for each lung separately, first by reading standard PA radiographs alone, immediately followed by reading a combination of the standard image and a second image, which was either identical to the first or a manipulated version of the first image.

The observers were informed that some image pairs had been processed differently, and that others might appear very similar, but that they should pay equal attention to all cases. The density and contrast of the first and second version of each image, including the motivator cases with digitally altered nodules, were matched. A short pause between the display of the first and second image (2 seconds) was introduced, so that small changes (or absence of change) between the two images were less easily perceived.

Other reading and rating conditions were the same in this supporting study as in the main observer study. Before the test, six additional cases, including four with identical second images and two with enhanced motivator second images, were provided for training.

Results

Average AUCs were 0.807 with use of standard chest radiographs, 0.867 with standard radiographs plus BS images, and 0.916 with standard radiographs plus DES images for all lungs by 10 observers. The differences were statistically significant between standard image and standard image plus BS image ($P < .001$), between standard image and standard image plus DES images ($P < .001$), and between standard image plus BS image and standard image plus DES images ($P = .001$) (Table 1, Fig 1). The main analysis of variance including the three reading conditions provided a P value of less than .001. Average AUCs for seven small lung

cancers less than 10 mm were 0.736 with standard chest radiographs, 0.837 with standard plus BS image, and 0.857 with standard plus DES images. The differences were statistically significant between standard and standard plus BS image ($P = .02$) and between standard and standard plus DES images ($P = .004$) but were not statistically significant between standard plus BS image and standard plus DES images ($P = .37$). Corresponding average AUCs for 14 radiologist-missed lung cancers were 0.705, 0.813, and 0.860, respectively. The differences were statistically significant between standard and standard plus BS image ($P < .001$), between standard and standard plus DES images ($P < .001$), and between standard plus BS image and standard plus DES images ($P = .04$).

To address concerns regarding lack of independence of the right and left lung scores, which can affect ROC analysis, we also analyzed the right and left lung scores separately. By considering only data from the right lungs,

Figure 1

Figure 1: Graph shows ROC curves for detection of cancers on chest radiographs by 10 observers. The differences in AUCs were statistically significant between standard images alone and standard images plus BS images ($P < .001$), between standard images alone and standard images plus DES images ($P < .001$), and between standard images plus BS images and standard images plus DES images ($P = .001$).

Table 1

Detection of Cancers without and with BS or DES Chest Images in Two Groups of Observers with Limited and Unlimited Reading Time

				95% Confidence Interval for Treatment Differences*		
Parameter	Standard Image	Standard Image with BS	Standard Image with DES	Standard plus BS Image versus Standard Image	Standard plus DES versus Standard Image	Standard plus DES versus Standard plus BS Image
Group 1 (limited time)						
Observer 1	0.877	0.909	0.949	−0.009, 0.072	0.015, 0.129	−0.000, 0.082
Observer 2	0.843	0.894	0.959	−0.043, 0.146	0.061, 0.171	−0.029, 0.159
Observer 3	0.832	0.876	0.947	−0.011, 0.100	0.064, 0.166	0.018, 0.123
Observer 4†	0.699	0.779	0.775	0.013, 0.146	0.012, 0.141	−0.016, 0.009
Observer 5†	0.740	0.823	0.928	0.025, 0.141	0.117, 0.260	0.051, 0.161
Mean‡	0.798 (0.750, 0.847)	0.856 (0.801, 0.912)	0.912 (0.881, 0.943)	0.020, 0.096	0.079, 0.148	0.022, 0.089
Group 2 (unlimited time)						
Observer 6	0.852	0.924	0.939	−0.020, 0.164	0.035, 0.137	−0.070, 0.099
Observer 7	0.808	0.862	0.926	0.003, 0.104	0.061, 0.173	0.014, 0.112
Observer 8	0.861	0.895	0.949	−0.007, 0.076	0.027, 0.150	−0.001, 0.109
Observer 9†	0.794	0.871	0.927	−0.009, 0.165	0.062, 0.205	−0.010, 0.121
Observer 10†	0.767	0.835	0.862	0.010, 0.123	0.033, 0.156	−0.016, 0.071
Mean‡	0.817 (0.767, 0.867)	0.878 (0.831, 0.924)	0.921 (0.879, 0.962)	0.028, 0.094	0.068, 0.139	0.011, 0.075
All‡	0.807 (0.761, 0.854)	0.867 (0.819, 0.915)	0.916 (0.882, 0.951)	0.028, 0.090	0.078, 0.139	0.022, 0.076

Note.—Unless otherwise indicated, data are AUCs. The differences in AUCs were statistically significant for all 10 observers between standard images and standard images plus BS images ($F_{1,231} = 19.40$, $P < .001$), between standard images and standard images plus DES images ($F_{1,51} = 42.02$, $P < .001$), and between standard images plus BS images and standard images plus DES images ($F_{1,41} = 11.82$, $P = .001$).

* Treatment-by-case analysis of variance confidence intervals for each reader.

† Radiology resident.

‡ Data in parentheses are 95% confidence intervals.

average AUCs were 0.809 with standard chest radiographs, 0.840 with standard plus BS image, and 0.913 with standard plus DES images. The differences were statistically significant between standard and standard plus BS image ($P = .04$), between standard and standard plus DES images ($P < .001$), and between standard plus BS image and standard plus DES images ($P = .002$). The differences in AUCs (0.810, 0.905, and 0.937, respectively) estimated only from the left lungs with the above image combinations were also statistically significant ($P = .001$, $P < .001$, and $P = .02$, respectively) (Table 2).

For observer groups 1 and 2, the average AUCs were 0.798 and 0.817,

respectively, when using standard chest radiographs alone, 0.856 and 0.878, respectively, with standard plus BS image, and 0.912 and 0.921, respectively, with standard plus DES images. For group 1, the P value was .002 for the difference between standard and standard plus BS image, .002 between standard and standard plus DES images, and .03 between standard plus BS image and standard plus DES images. Between these same pairs of image sets, the P values for group 2 were .001, $< .001$, and .005, respectively. Thus, the differences achieved with each of the three image types were significant for both groups.

For all attending radiologists as a group and all radiology residents as a

group, respectively, the average AUCs were 0.846 and 0.750 with use of standard chest radiographs alone, 0.894 and 0.827 with standard plus BS image, and 0.945 and 0.873 with standard plus DES images. For each group of attending radiologists, the differences between each image set were statistically significant. For the attending radiologists, the P value was less than .001 between standard and standard plus BS image, less than .001 between standard and standard plus DES images, and .001 between standard plus BS image and standard plus DES images. Between these same pairs of image sets, the P values for the radiology residents were less than .001, .008, and .136, respectively (Fig 2).

Table 2

Detection of Cancers without and with BS or DES Chest Images in Right and Left Lungs for 10 Observers

Parameter	Standard Image	Standard Image with BS	Standard Image with DES	95% Confidence Interval for Treatment Differences*		
				Standard plus BS Image versus Standard Image	Standard plus DES versus Standard Image	Standard plus DES versus Standard plus BS Image
Right lung						
Observer 1	0.905	0.912	0.944	−0.016, 0.031	−0.040, 0.118	−0.047, 0.109
Observer 2	0.858	0.846	0.959	−0.069, 0.046	0.029, 0.173	0.043, 0.182
Observer 3	0.795	0.828	0.943	−0.015, 0.083	0.064, 0.232	0.036, 0.193
Observer 4†	0.658	0.721	0.718	−0.084, 0.211	−0.087, 0.206	−0.022, 0.014
Observer 5†	0.728	0.791	0.944	−0.040, 0.165	0.091, 0.340	0.067, 0.239
Observer 6	0.863	0.932	0.964	−0.018, 0.157	0.035, 0.167	−0.035, 0.098
Observer 7	0.814	0.868	0.914	−0.020, 0.127	0.030, 0.170	−0.013, 0.105
Observer 8	0.885	0.873	0.946	−0.035, 0.011	0.003, 0.118	0.014, 0.133
Observer 9†	0.837	0.825	0.939	−0.120, 0.096	0.021, 0.183	0.022, 0.207
Observer 10†	0.746	0.806	0.856	−0.017, 0.136	0.017, 0.202	−0.021, 0.121
Mean‡	0.809 (0.743, 0.875)	0.840 (0.774, 0.907)	0.913 (0.868, 0.957)	0.000, 0.062	0.066, 0.142	0.037, 0.108
Left lung						
Observer 1	0.849	0.906	0.974	−0.032, 0.147	0.026, 0.225	0.010, 0.127
Observer 2	0.827	0.957	0.974	0.043, 0.217	0.065, 0.228	−0.031, 0.063
Observer 3	0.863	0.932	0.953	−0.007, 0.143	0.018, 0.161	−0.014, 0.057
Observer 4†	0.726	0.871	0.871	0.029, 0.262	0.031, 0.258	−0.043, 0.041
Observer 5†	0.745	0.860	0.939	0.030, 0.199	0.079, 0.308	−0.009, 0.166
Observer 6	0.840	0.926	0.945	−0.034, 0.205	0.031, 0.179	−0.094, 0.132
Observer 7	0.829	0.872	0.948	−0.012, 0.099	0.039, 0.200	0.013, 0.139
Observer 8	0.859	0.937	0.963	−0.008, 0.163	0.011, 0.196	−0.012, 0.064
Observer 9†	0.772	0.927	0.934	0.055, 0.256	0.067, 0.257	−0.092, 0.105
Observer 10†	0.788	0.865	0.871	−0.009, 0.164	−0.007, 0.174	−0.037, 0.049
Mean‡	0.810 (0.744, 0.875)	0.905 (0.851, 0.959)	0.937 (0.892, 0.982)	0.046, 0.145	0.076, 0.178	0.004, 0.059

Note.—Unless otherwise indicated, data are AUCs. The differences in AUCs were statistically significant for left lungs between standard and standard plus BS images ($F_{1,23} = 5.00$, $P = .04$), between standard and standard plus DES images ($F_{1,26} = 25.30$, $P < .001$), and between standard plus BS image and standard plus DES images ($F_{1,33} = 11.28$, $P = .002$). The differences in AUCs were statistically significant for right lungs, with P value of .001 at $F_{1,143}$ of 15.45, with P value less than .001 at $F_{1,235}$ of 26.54, and with P value of .02 at $F_{1,33}$ of 5.69 between the above image orders.

* Treatment-by-case analysis of variance confidence intervals for each reader.

[†] Radiology resident.

[‡] Data in parentheses are 95% confidence intervals.

Sensitivity and specificity values were based on recorded confidence levels regarding the presence of cancer in a lung at three reading conditions for 10 observers (Table 3). For 55 lungs with a positive finding, the mean confidence level regarding the presence of a lung

cancer was increased from 0.51 with standard image alone, to 0.61 with standard plus BS image, and to 0.69 with standard plus DES. For 105 lungs with a negative finding, the mean confidence level was almost the same for the standard image alone (0.24) and standard

plus BS image (0.23) and decreased to 0.18 with standard plus DES.

Among very subtle lung cancers on the standard radiographs, 40% (12 of 30) on the BS images and 73% (22 of 30) on the soft-tissue DES images were relatively obvious (mean confidence level regarding the presence of a lung cancer by 10 observers increased 0.10 or more) (Figs 3, 4). However, using BS images also increased the mean confidence level more than 0.10 regarding the presence of cancer in four lungs with a negative finding (four [3.8%] of 105) compared with the DES images. For instance, a spurious nodular opacity caused by overlapping vessels and bones appeared more prominent on the BS image (probably because of the enhancement technique used) than on the standard radiograph or DES images (Fig 5).

In the supporting study using sequential identical images, the average reading time for the same 39 part 1 cases by the same five observers was 22 seconds (range, 14–35 seconds) for the first standard image and 15 seconds (range, 11–26 seconds) for the second identical image. The average AUCs were 0.830 with use of standard chest radiograph once and 0.837 with use of the same radiograph twice in 39 cases, which was

Figure 2

Figure 2: Graphs show ROC curves for detection of cancers on chest radiographs with (a) six attending radiologists and (b) four radiology residents. (a) For the six attending radiologists, the differences in AUCs were statistically significant between standard and standard plus BS images ($P < .001$), between standard and standard plus DES images ($P < .001$), and between the standard plus BS images and standard plus DES images ($P = .001$). (b) P values between the corresponding image sets were less than .001, .008, and .136, respectively, for the four less-experienced radiology residents.

Table 3

Recorded Confidence Level Regarding the Presence of Cancer in a Lung at Three Reading Conditions for 10 Observers

Parameter	Lungs with Positive Finding ($n = 55$)			Lungs with Negative Finding ($n = 105$)		
	Standard Image	Standard Image with BS Image	Standard Image with DES Image	Standard Image	Standard Image with BS Image	Standard Image with DES Image
Observer 1	0.59	0.65	0.74	0.20	0.19	0.14
Observer 2	0.52	0.63	0.74	0.19	0.19	0.12
Observer 3	0.53	0.64	0.74	0.20	0.19	0.15
Observer 4*	0.32	0.51	0.51	0.11	0.13	0.13
Observer 5*	0.53	0.56	0.65	0.48	0.44	0.36
Observer 6	0.49	0.62	0.63	0.17	0.20	0.16
Observer 7	0.57	0.69	0.80	0.18	0.20	0.17
Observer 8	0.59	0.66	0.77	0.19	0.19	0.14
Observer 9*	0.48	0.58	0.63	0.30	0.30	0.24
Observer 10*	0.49	0.59	0.66	0.34	0.29	0.24
Mean $\pm$ standard deviation	0.51 $\pm$ 0.19	0.61 $\pm$ 0.17	0.69 $\pm$ 0.18	0.24 $\pm$ 0.15	0.23 $\pm$ 0.15	0.18 $\pm$ 0.13

Note.—The scale for confidence level was 0 for definitely absent and 1 for definitely present. For 55 lungs with a positive finding, the mean confidence level regarding the presence of a lung cancer was increased from 0.51 with standard image alone, to 0.61 with standard plus BS image, and to 0.69 with standard plus DES, with P values of .003 and less than .001 between the first and second or third reading condition and of .034 between the second and third condition. For 105 lungs with a negative finding, the mean confidence level was almost same for standard image alone (0.24) and standard plus BS image (0.23) and decreased to 0.18 with standard plus DES, with no statistically significant difference.

* Radiology resident.

Figure 3

Figure 3: Chest PA radiographs in 79-year-old man with nodular cancer in left upper lobe that was overlooked at the time of initial interpretation. **(a)** Standard image shows the cancer (arrow) that is partly overlapped with ribs. **(b)** BS image clearly shows the cancer but also enhances other structures, especially in the apical areas. **(c)** DES soft-tissue image also clearly shows the cancer. **(d)** DES bone image shows some calcifications in paratracheal lymph nodes and costal cartilages (small arrows) but no calcification in the nodule.

not statistically significant (Table 4). In contrast, in the main study, the average AUCs were 0.849 with standard chest radiographs alone, 0.901 with standard plus BS image, and 0.939 with standard

plus DES in the same 39 cases. The *P* value for the difference between standard and standard plus BS image was .013 and less than .001 between standard and standard plus DES images (Table 5).

Discussion

Missed lung cancers (ie, cancers that were not detected or described at the time of interpretation) include those

Figure 4

Figure 4: Chest PA radiographs in 82-year-old woman with nodular cancer in the left upper lobe that was detected at the time of interpretation. **(a)** Standard image shows the cancer (arrow) partly obscured by ribs. **(b)** BS image shows the cancer more clearly. **(c)** DES soft-tissue image shows the cancer and achieves more complete BS in the apical areas. **(d)** DES bone image shows no evidence of calcification in the nodule. The average recorded confidence level regarding the presence of cancer in this lung by 10 observers was 0.28, 0.62, and 0.77 (0 = definitely absent, 1 = definitely present) with standard radiograph alone, standard plus BS image, and standard plus DES images, respectively.

lesions that are more difficult for radiologists to detect. A major cause of such oversight errors is obscuration of the cancer by superimposed bones such

as ribs and clavicles. Shah et al (12) reported that, in a missed cancer series with 40 cancers, 38 (95%) were partly obscured by overlying ribs, including

26 (65%) that were obscured by two or three bones and nine (22%) that were also obscured by a clavicle. In a previous study (13), we reported that,

Figure 5

Figure 5: Chest PA radiographs in 79-year-old man. **(a)** Standard image shows faint opacity at the left heart border because of superimposed ribs and soft tissue. **(b)** BS image suggests a possible nodule at this location (arrow). DES **(c)** soft-tissue and **(d)** bone images do not confirm a nodule. The average recorded confidence level regarding the presence of cancer in this lung by 10 observers was 0.41, 0.59, and 0.43 (0 = definitely absent, 1 = definitely present) with standard radiograph, standard plus BS image, and standard plus DES images, respectively.

of 59 primary lung cancers (which included 24 radiologist-missed cancers), more than 90% were partly obscured by overlying bones. In the present observer study, the 14 radiologist-missed

and 41 radiologist-detected nodular cancers were selected on the basis of the visual subtlety of the nodules on chest radiographs, independent of the spatial relationship with bones. The

incidence of obscuration of these cancers by bones was similar to that in our previous study (13).

Image-processing techniques for suppressing skeletal structures on standard

chest radiographs have been reported recently, and a postprocessing BS imaging algorithm can suppress the conspicuity of bones without altering the acquisition parameters (14,15). For this observer study, we used a new commercial image processing system for BS imaging (SoftView, version 2.0). The radiologists' performance for detection of relatively subtle nodular cancers was significantly improved when using BS images compared with standard images for all 10 of our observers.

In a previous observer study (5), we showed that radiologists' performance for detection of missed cancers was significantly improved by use of DES chest

radiographs. Szucs-Farkas et al (4) also reported that energy subtraction radiographs improved the detection of lung nodules and masses in daily routine reading. In the present study, the observers' performance was further improved by use of DES (AUC from 0.867 to 0.916), and the improvement was statistically significant for the six experienced observers (AUC from 0.894 to 0.945). However, DES radiographs have the requirement of specialized equipment, and some users have reported a small increase in average radiation dose.

The differences in observer performance between standard plus BS image

and standard plus DES soft-tissue images are partly attributable to the fact that this version of BS imaging, while diminishing visibility of ribs in the lung areas, also increased the prominence of normal vascular structures and small scars, apparently because of mid-frequency enhancement. However, our observers were using BS imaging for the first time in this test, although they all had extensive previous experience with DES. In addition, DES has the advantage of providing a bone image that reveals calcifications that can mimic nodules in some cases.

As with most observer tests, the results reflected the experience of the observers, so that the diagnostic accuracy of the experienced radiologists was superior to that of the less experienced radiology residents (16–18). Our results also indicated that the performance of senior radiology residents in detection of subtle nodular cancers on chest radiographs was higher than that of junior radiology residents.

Many factors, including the characteristics of the individual cases, the levels of skill of the observers, and the image properties, can influence the reading time for each image (19–21). Confusing cases with false-negative and false-positive results may receive prolonged visual attention by observers (19,20). In the present study, four of five observers in group 2 (unrestricted

Table 4

Detection of Nodular Cancers for Five Attending Radiologists in 39 Cases with Single Viewing (One Look) versus Sequential Viewing (Two Looks) of Identical Standard Images

Parameter	One Look	Two Looks	95% Confidence Interval for Two Looks versus One Look*
Observer 1	0.837	0.826	−0.057, 0.035
Observer 2	0.831	0.858	−0.008, 0.061
Observer 3	0.845	0.869	−0.013, 0.061
Observer 6	0.813	0.795	−0.074, 0.038
Observer 8	0.825	0.838	−0.033, 0.059
Mean†	0.830 (0.750, 0.910)	0.837 (0.756, 0.918)	−0.015, 0.029

Note.—Unless otherwise indicated, data are AUCs. No difference was statistically significant with $F_{1,7}$ of 0.45 in AUC between reading first and second same images. Observer numbers are the same as in the main observer study.

* Treatment-by-case analysis of variance confidence intervals for each reader.

† Data in parentheses are 95% confidence intervals.

Table 5

Detection of Nodular Cancers for Five Attending Radiologists in 39 Cases without and with BS or DES Chest Images

Parameter	Standard Image	Standard Image with BS	Standard Image with DES	95% Confidence Interval for Treatment Differences*		
				Standard plus BS Image versus Standard Image	Standard plus DES versus Standard Image	Standard plus DES versus Standard plus BS Image
Observer 1	0.878	0.912	0.947	−0.009, 0.077	0.001, 0.137	−0.022, 0.092
Observer 2	0.826	0.890	0.923	−0.020, 0.148	0.022, 0.172	−0.038, 0.104
Observer 3	0.848	0.885	0.950	−0.035, 0.110	0.027, 0.177	−0.017, 0.146
Observer 6	0.846	0.93	0.913	0.013, 0.156	−0.017, 0.151	−0.106, 0.072
Observer 8	0.849	0.885	0.958	−0.046, 0.120	0.007, 0.212	−0.001, 0.147
Mean†	0.849 (0.784, 0.914)	0.901 (0.843, 0.959)	0.938 (0.894, 0.983)	0.007, 0.096	0.044, 0.134	0.002, 0.074

Note.—Unless otherwise indicated, data are AUCs. The difference in AUC between standard and standard plus BS image was statistically significant ($F_{1,66} = 6.56$, $P = .013$), and the difference in AUC between standard and standard plus DES was statistically significant ($F_{1,59} = 27.24$, $P < .001$). However, the difference in AUC between standard plus BS image and standard plus DES was not significant ($F_{1,6} = 4.71$, $P = .075$). Observer numbers are the same as in the main observer study.

* Treatment-by-case analysis of variance confidence intervals for each reader.

† Data in parentheses are 95% confidence intervals.

time) used an average of 14 seconds for reading the standard image, an additional 11 seconds for reading the BS image, and 11 more seconds for reading the DES soft-tissue and bone images, compared with group 1, where the maximum reading time was limited to 12 seconds per image (standard or BS image) or images (paired DES soft-tissue and bone images). The average AUCs in detection of these nodular cancers by the two groups were very similar for each class of image. We did not find any substantial improvement in observer performance by allowing unlimited reading time in the present study. This is consistent with previously published studies that have examined the relationship between prolonged image scrutiny and observer accuracy (22). However, the additional 11 seconds used for reading the BS images in this study may not accurately reflect clinical practice insofar as the observers were obliged to complete their review and scoring of the standard image prior to viewing the BS image; whereas, in clinical practice, the images would be reviewed essentially simultaneously by toggling back and forth between them on a picture archiving and communication system workstation, as is done currently with DES radiographs.

Observer performance in detection of lung nodules on chest radiographs or detection of colon polyps on CT colonographic images has been shown to be improved by using computer-aided diagnostic schemes in studies that have used both independent and sequential designs (23,24). However, to address a concern regarding possible bias that might be produced by sequential reading of different versions of the same cases, we performed an additional observer test 1 year after this main study by using a random subset of the same readers and cases, in which two identical versions of each image were presented to the observers. The results from our supporting study indicated that observer performance was not substantially improved by reading identical images twice, even when the observers were motivated to scrutinize the images carefully. Therefore, we believe that this sequential

method is appropriate for evaluating image processing techniques, including computer-aided diagnostic schemes, in observer performance studies.

Although the results that consider the two lungs as independent were highly significant, they did not account for potential correlation between ratings from the two lungs of each patient because of similarities between the two lungs and/or satisfaction of search. To investigate this possibility, we performed separate analyses of ratings from the right and from the left lungs. That analysis was much less powerful because the number of ratings in each comparison was reduced by half, but it eliminated any effect of potential correlation between paired organs.

The confidence intervals of the differences computed from pooled ratings of left and right lungs were the most anticonservative, whereas those computed from only left or right lungs were the most conservative. Because both analyses indicated that all estimated differences in AUC differ significantly from zero, we concluded that those differences are real. To date, we are not aware of any software that implements a method capable of accounting quantitatively for possible correlation between each reader's ratings of left and right lungs in multireader multicase data. However, we attempted to identify the presence of any such correlation by using the nonparametric rank-based Spearman correlation coefficient conditional on truth, the Wilcoxon signed-rank test, and subjective assessments of scatter diagrams and found very little evidence supporting it (25).

Limitations of this study included the fact that it involved use of selected cases under controlled conditions. This order of sequential presentation was based on the assumption that the standard image would not be superior to standard plus BS image, and that standard plus DES would not be inferior to either standard images alone or standard images plus BS image. We believe these assumptions were reasonable on the basis of the nature of the various images and the fact that the standard images were included as a component

of every viewing condition. The magnitude and significance of the differences between the three image combinations together with the negative results of the sequential identical image presentation test serve to validate this approach.

Because BS imaging is optimized for nodular opacities, it is not intended to replace the standard radiograph in other applications, although it likely will be beneficial for visualization of a range of focal opacities. Neither do we envision that BS imaging would supplant DES in those situations where DES is used currently. However, for the large majority of digital chest radiographs that are obtained without energy subtraction, including bedside (portable) examinations, it seems likely that BS could be beneficial. In summary, the results of this observer study suggest that, while DES radiography provides the highest accuracy, use of BS software in combination with a standard radiograph can significantly improve detection of small nodular cancers compared with use of standard radiographs alone.

Acknowledgments: The authors thank Hiroyuki Abe, MD, Karin Dill, MD, Alexandra Funaki, DO, Hao Lo, MD, Jay Patel, MD, Valeria Patigailo, MD, Yonglin Pu, MD, Christopher Straus, MD, Dave Schacht, MD, and Steven M. Zangan, MD, for participating as observers and Mara MacMahon for rendering realistic enhanced and suppressed nodules used in the supporting observer test with Adobe Photoshop.

Disclosures of Potential Conflicts of Interest:

E.L. Financial activities related to the present article: institution receives grant money from Riverain Medical (although research not specific for this publication). Financial activities not related to the present article: institution receives money (patents and royalties) from R2 Technology (now Hologic), Deus Technologies, Riverain Medical, Mitsubishi Space Software, Median Technologies, General Electric, and Toshiba (CAD technologies developed at the University of Chicago have been licensed to these companies). Other relationships: none to disclose.
R.E. Financial activities related to the present article: institution receives grant money from Riverain Medical (although research not specific for this publication). Financial activities not related to the present article: institution receives money (patents and royalties) from R2 Technology (now Hologic), Deus Technologies, Riverain Medical, Mitsubishi Space Software, Median Technologies, General Electric, and Toshiba (CAD technologies developed at the University of Chicago have been licensed to these companies). Other relationships: none to disclose.
L.L.P. Financial activities related to the present

article: institution receives grant money from Riverain Medical (although research not specific for this publication). Financial activities not related to the present article: institution receives money (patents and royalties) from R2 Technology (now Hologic), Deus Technologies, Riverain Medical, Mitsubishi Space Software, Median Technologies, General Electric, and Toshiba (CAD technologies developed at the University of Chicago have been licensed to these companies). Other relationships: none to disclose. **K.D.** Financial activities related to the present article: none to disclose. Financial activities not related to the present article: author is shareholder in Hologic and consultant for Riverain Medical. Other relationships: none to disclose. **C.E.M.** Financial activities related to the present article: institution receives grant money from Riverain Medical (although research not specific for this publication). Financial activities not related to the present article: author is consultant for VuCOMP; author and institution have patents with and receive royalties from UChicagoTech. Other relationships: none to disclose. **H.M.** Financial activities related to the present article: institution receives grant money from Riverain Medical (although research not specific for this publication). Financial activities not related to the present article: author is consultant for Riverain Medical; author and institution have patents with and receive royalties from UChicagoTech; author is shareholder in Hologic. Other relationships: none to disclose.

References

- McAdams HP, Samei E, Dobbins J 3rd, Tourassi GD, Ravin CE. Recent advances in chest radiography. *Radiology* 2006;241(3):663–683.
- MacMahon H, Li F, Engelmann R, Roberts R, Armato S. Dual energy subtraction and temporal subtraction chest radiography. *J Thorac Imaging* 2008;23(2):77–85.
- Kelcz F, Zink FE, Peppler WW, Kruger DG, Ergun DL, Mistretta CA. Conventional chest radiography vs dual-energy computed radiography in the detection and characterization of pulmonary nodules. *AJR Am J Roentgenol* 1994;162(2):271–278.
- Szucs-Farkas Z, Patak MA, Yuksel-Hatz S, Ruder T, Vock P. Single-exposure dual-energy subtraction chest radiography: detection of pulmonary nodules and masses in clinical practice. *Eur Radiol* 2008;18(1):24–31.
- Li F, Engelmann R, Doi K, MacMahon H. Improved detection of small lung cancers with dual-energy subtraction chest radiography. *AJR Am J Roentgenol* 2008;190(4):886–891.
- Li F, Engelmann R, Metz CE, Doi K, MacMahon H. Lung cancers missed on chest radiographs: results obtained with a commercial computer-aided detection program. *Radiology* 2008;246(1):273–280.
- Metz CE, Herman BA, Shen JH. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Stat Med* 1998;17(9):1033–1053.
- Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method. *Invest Radiol* 1992;27(9):723–731.
- Hillis SL, Berbaum KS, Metz CE. Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. *Acad Radiol* 2008;15(5):647–661.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143(1):29–36.
- Metz CE. Quantification of failure to demonstrate statistical significance: the usefulness of confidence intervals. *Invest Radiol* 1993;28(1):59–63.
- Shah PK, Austin JHM, White CS, et al. Missed non-small cell lung cancer: radiographic findings of potentially resectable lesions evident only in retrospect. *Radiology* 2003;226(1):235–241.
- Li F, Engelmann R, Doi K, MacMahon H. Small nodular lung cancers on chest radiographs: obscuration by bones and mean size at the time of detection (abstr). In: Radiological Society of North America scientific assembly and annual meeting program. Oak Brook, Ill: Radiological Society of North America, 2007; 326.
- Loog M, van Ginneken B, Schilham AMR. Filter learning: application to suppression of bony structures from chest radiographs. *Med Image Anal* 2006;10(6):826–840.
- Suzuki K, Abe H, MacMahon H, Doi K. Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN). *IEEE Trans Med Imaging* 2006;25(4):406–416.
- MacMahon H, Engelmann R, Behlen FM, et al. Computer-aided diagnosis of pulmonary nodules: results of a large-scale observer test. *Radiology* 1999;213(3):723–726.
- Quekel LG, Kessels AG, Goei R, van Engelsehoven JM. Detection of lung cancer on the chest radiograph: a study on observer performance. *Eur J Radiol* 2001;39(2):111–116.
- Monnier-Cholley L, Carrat F, Cholley BP, Tubiana JM, Arrivé L. Detection of lung cancer on radiographs: receiver operating characteristic analyses of radiologists', pulmonologists', and anesthesiologists' performance. *Radiology* 2004;233(3):799–805.
- Kundel HL, Nodine CF, Krupinski EA. Searching for lung nodules: visual dwell indicates locations of false-positive and false-negative decisions. *Invest Radiol* 1989;24(6):472–478.
- Manning D, Barker-Mill SC, Donovan T, Crawford T. Time-dependent observer errors in pulmonary nodule detection. *Br J Radiol* 2006;79(940):342–346.
- Kakeda S, Kamada K, Hatakeyama Y, et al. Effect of temporal subtraction technique on interpretation time and diagnostic accuracy of chest radiography. *AJR Am J Roentgenol* 2006;187(5):1253–1259.
- Oestmann JW, Greene R, Kushner DC, Bourgouin PM, Linetsky L, Llewellyn HJ. Lung lesions: correlation between viewing time and detection. *Radiology* 1988;166(2):451–453.
- Kobayashi T, Xu XW, MacMahon H, Metz CE, Doi K. Effect of a computer-aided diagnosis scheme on radiologists' performance in detection of lung nodules on radiographs. *Radiology* 1996;199(3):843–848.
- Obuchowski NA, Mezziane M, Dachman AH, Lieber ML, Mazzone PJ. What's the control in studies measuring the effect of computer-aided detection (CAD) on observer performance? *Acad Radiol* 2010;17(6):761–767.
- Glantz SA. Primer of biostatistics. 5th ed. New York, NY: McGraw-Hill, 2002.