

Feng Li, MD, PhD
 Hidetaka Arimura, PhD
 Kenji Suzuki, PhD
 Junji Shiraishi, PhD
 Qiang Li, PhD
 Hiroyuki Abe, MD, PhD
 Roger Engelmann, MS
 Shusuke Sone, MD, PhD
 Heber MacMahon, MD
 Kunio Doi, PhD

Published online

10.1148/radiol.2372041555
 Radiology 2005; 237:684–690

Abbreviations:

A_z = area under ROC curve
 CAD = computer-aided detection
 GGO = ground-glass opacity
 LROC = localization ROC
 ROC = receiver operating characteristic

¹ From the Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology, MC-2026, University of Chicago, 5841 S Maryland Ave, Chicago, IL 60637 (F.L., K.S., J.S., Q.L., H.A., R.E., H.M., K.D.); Department of Health Sciences, Faculty of Medicine, Kyushu University, Fukuoka, Japan (H.A.); and J. A. Azumi General Hospital, Nagano, Japan (S.S.). From the 2003 RSNA Annual Meeting. Received September 8, 2004; revision requested October 29; revision received December 27; accepted January 21, 2005. Supported in part by U.S. Public Health Service grants CA62625 and CA98119. **Address correspondence** to F.L. (e-mail: feng@uchicago.edu).

See Materials and Methods for pertinent disclosures.

Author contributions:

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

© RSNA, 2005

Computer-aided Detection of Peripheral Lung Cancers Missed at CT: ROC Analyses without and with Localization¹

PURPOSE: To retrospectively evaluate whether a difference-image computer-aided detection (CAD) scheme can help radiologists detect peripheral lung cancers missed at low-dose computed tomography (CT).

MATERIALS AND METHODS: Institutional review board approval and informed patient and observer consent were obtained. Seventeen patients (eight men and nine women; mean age, 60 years) with a missed peripheral lung cancer and 10 control subjects (five men and five women; mean age, 63 years) without cancer at low-dose CT were included in an observer study. Fourteen radiologists were divided into two groups on the basis of different image display formats: Six radiologists (group 1) reviewed CT scans with a multiformat display, and eight radiologists (group 2) reviewed images with a “stacked” cine-mode display. The radiologists, first without and then with the CAD scheme, indicated their confidence level regarding the presence (or absence) of cancer and the most likely position of a lesion on each CT scan. Receiver operating characteristic (ROC) curves were calculated without and with localization to evaluate the observers’ performance.

RESULTS: With the CAD scheme, the average area under the ROC curve improved from 0.763 to 0.854 for all radiologists ($P = .002$), from 0.757 to 0.862 for group 1 ($P = .04$), and from 0.768 to 0.848 for group 2 ($P = .01$). The average sensitivity in the detection of 17 cancers improved from 52% (124 of 238 observations) to 68% (163 of 238 observations) for all radiologists ($P < .001$), from 49% (50 of 102 observations) to 71% (72 of 102 observations) for group 1 ($P = .02$), and from 54% (74 of 136 observations) to 67% (91 of 136 observations) for group 2 ($P = .006$). The localization ROC curve also improved.

CONCLUSION: Lung cancers missed at low-dose CT were very difficult to detect, even in an observer study. The use of CAD, however, can improve radiologists’ performance in the detection of these subtle cancers.

© RSNA, 2005

In the past decade, low-dose single-detector row computed tomography (CT) with 10-mm-thick sections has been used to screen asymptomatic smoking and nonsmoking populations for lung cancer (1–3). The results of these CT screening studies showed that early peripheral lung cancers usually appeared as solitary noncalcified lesions with or without areas of ground-glass opacity (GGO), and the detection of these lesions at an early stage was greatly improved with use of low-dose CT rather than chest radiography. Most CT scans obtained in screening programs, however, showed only minor benign abnormalities, including noncancerous abnormalities such as diffuse lung disease (emphysema and interstitial changes) and focal lung disease (active infections, scars, and calcified nodules); in addition, 5%–27% of patients had noncalcified nodules that were detected at baseline screening with use of low-dose CT and 10-mm-thick sections (1–3). When reading images obtained in a CT screening program, radiologists must search for suspicious

noncalcified lung nodules, differentiate these lesions from benign nodules and lung cancer, and, finally, recommend follow-up actions for the detected lesions. In the studies mentioned above, no additional clinical information was provided to radiologists reviewing the CT scans except for age, sex, and smoking status.

At baseline CT screening performed in a general population that included smokers and nonsmokers in Nagano, Japan (2), the fraction of lung cancers among the detected noncalcified lesions was 9% and the prevalence of cancers was only 0.48%. The corresponding data were 12% and 2.7%, respectively, for smokers in the U.S. Early Lung Cancer Action Project (3). In CT screening programs, however, 32%–39% of lung cancers (4,5) were missed in previous years, and the numbers of these missed cancers were not included in the determination of the prevalence of lung cancers in these studies. We previously reported (5) that 32 missed lung cancers were very difficult to detect in the Nagano series; in general, they were very subtle and appeared as small, faint nodules with GGO that overlapped normal structures or as opacities in a complex background of other diseases.

When an automated lung nodule-detection method (6) was used, 84% of these missed lung cancers in the Nagano series were marked by the computer; however, the false-positive rate was high (1.0 false-positive marks per section, 28 false-positive marks per study), and this is not acceptable to radiologists. Recently, we developed a computer-aided detection (CAD) scheme (7) that is based on a difference-image technique for enhancing lung cancers and suppressing most normal background structures, and the false-positive rate has improved to about 3.0 marks per study (sensitivity, 87%) with use of a multiple massive training artificial neural network (8). Thus, the purpose of our study was to retrospectively evaluate whether a difference-image CAD scheme can help radiologists detect peripheral lung cancers missed at low-dose CT.

MATERIALS AND METHODS

H.M. and K.D. are shareholders in R2 Technology, Sunnyvale, Calif. K.D. is a shareholder in Deus Technology, Rockville, Md. CAD technologies developed in the Kurt Rossmann Laboratories have been licensed to companies including R2 Technology, Deus Technologies, Riverain

Medical Group, Mitsubishi Space Software, Median Technologies, GE, and Toshiba.

Database

An annual low-dose CT screening program for lung cancer in Nagano, Japan, began in May 1996 and ended in March 1999. In the program, 17 892 examinations were performed in 7847 individuals (4288 men, 3559 women; mean age, 61 years; age range, 19–92 years). All individuals gave informed consent to undergo CT screening and for use of the data for research purposes. The database used in this study consisted of data from 38 low-dose CT examinations performed in 31 patients with missed peripheral lung cancers. All of the CT studies had been performed as part of the 3-year lung cancer screening program (5,6). Twenty-three cancers were missed because of detection errors, and 15 cancers were missed because of interpretation errors.

As described previously (5), the locations of missed lung cancers on sections obtained at 39 CT examinations (one examination was excluded from this study because of technical error) were determined in consensus by two radiologists (F.L. and S.S., with 20 and 42 years of experience, respectively). One radiologist (F.L.) measured the length and width of cancers on at least one section. Three radiologists (F.L., H.A., and H.M., with 20, 18, and 29 years of experience, respectively) first independently classified the low-dose CT scans with the 38 cancers into three patterns, and the final judgment was based on agreement by at least two radiologists. The mean diameter of the 38 lesions missed at low-dose CT was 12 mm (range, 6–26 mm). The following patterns were noted: 10 nodules had pure GGO (nonsolid), 16 had mixed GGO (part solid), and 12 had solid opacity.

The 31 missed cancers, which included 28 adenocarcinomas, two small cell carcinomas, and one squamous cell carcinoma, were confirmed with surgery. The CT examinations were performed with a mobile scanner (CT-W950SR; Hitachi Medical, Tokyo, Japan) with use of a low-dose protocol and a tube current of 25 or 50 mA, a scanning time of 2 seconds per rotation of the x-ray tube (tube rotation time, 2 seconds), a table speed of 10 mm/sec (pitch, 2), 10-mm collimation, and a 10-mm reconstruction interval. The mean number of sections per study was 30, and the pixel size was 0.586 or 0.684 mm for scans with a 512 × 512 image matrix size. The use of this database and the participa-

tion of radiologists in this observer performance study were approved by the University of Chicago Institutional Review Board. Informed consent for the observer performance study was obtained from all observers.

CAD Scheme

Our scheme was based on a difference-image technique (7,9,10) that enhances the lung nodules and suppresses most of the background normal structures. The difference image for each CT study was obtained by subtracting the nodule-suppressed image processed with a ring average filter from the nodule-enhanced image processed with a matched filter. By applying a multiple-gray-level threshold technique to the difference image, on which most nodules showed strong enhancement, the initial nodule candidates were identified. A number of false-positive findings were removed by using the two rule-based schemes on the localized image features related to morphologic characteristics and gray levels, and a false-positive rate of 15.8 per study was achieved (7). Most (81%) of the remaining false-positive findings were eliminated without removing any true-positive findings by using a multiple massive training artificial neural network trained to reduce various types of false-positive findings (8). The CAD scheme had a sensitivity of 87% (33 of 38 cancers) for 38 missed cancers, with an average of 3.0 false-positive findings per study (7,8).

Observer Study

Among 23 studies in which cancer was missed due to detection errors, 17 studies in 17 patients (eight men and nine women; mean age, 60 years; age range, 48–69 years) were performed the year before the cancers were found; the other six studies, including three that were performed in the same 17 patients 2 years before the cancers were found and three that revealed a coexisting benign nodule (diameter, 4–5 mm), were not used in this investigation. All 17 cancers were adenocarcinomas. At low-dose CT, six nodules had pure GGO, 10 nodules had mixed GGO, and one nodule had solid opacity. The mean diameter of the 17 missed cancers was 10 mm (range, 6–17 mm). Fifteen studies in which cancer was missed due to interpretation errors were also excluded from the observer study. In addition, we included studies obtained in 10 control subjects (five men and five women;

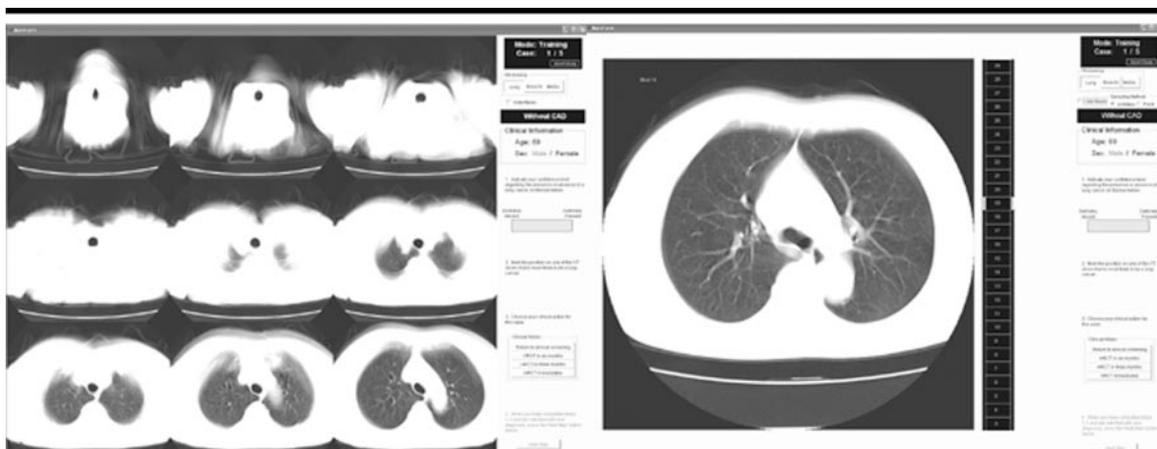


Figure 1. (a) Example of the multiformat display used by the six radiologists in group 1. From the top to the bottom of the entire lung for each patient, 27 consecutive transverse CT sections were displayed in a multiformat (3×3) mode on three high-spatial-resolution monitors. (b) Stacked cine-mode display used by the eight radiologists in group 2. Magnified and stacked transverse CT sections were displayed on one monitor.

TABLE 1
 A_z Values for 14 Radiologists in Detection of Missed Cancers without and with CAD Scheme

Group and Observer	A_z Value	
	Without CAD	With CAD
Group 1*		
1	0.856	0.962
2	0.792	0.825
3	0.613	0.824
4	0.723	0.865
5	0.851	0.876
6	0.706	0.818
Group 2†		
7	0.589	0.828
8	0.723	0.834
9	0.818	0.824
10	0.865	0.936
11	0.811	0.899
12	0.826	0.856
13	0.786	0.822
14	0.728	0.784
All observers	0.763	0.854

Note.—The difference in A_z values without and with the CAD scheme was statistically significant, with a P value of .002 for all radiologists, .04 for group 1, and .01 for group 2. No statistically significant difference in A_z values between the two viewing modes was found for observers without and with the CAD scheme.

* This group used a multiformat display. The mean A_z value for this group without CAD was 0.757; the mean value with CAD was 0.862.

† This group used a cine-mode display. The mean A_z value for this group without CAD was 0.768; the mean value with CAD was 0.848.

mean age, 63 years; age range, 49–69 years) without cancer who had participated in the same screening program

and whose ages and sexes closely matched those of the patient group; findings in these subjects were confirmed with 2-year follow-up. Some of the 27 studies revealed other abnormal findings such as scars, focal interstitial lung lesions, and small (<3 mm) benign nodules. The CAD scheme had a sensitivity of 82% (14 of 17 cancers), with 3.0 false-positive findings per study (range, zero to eight) for patients with missed cancers and 2.4 false-positive findings per study (range, zero to five) for the 10 control subjects (7,8).

Two image display formats were used in this investigation: a multiformat display and a “stacked” cine-mode display (Fig 1). For the multiformat display, from the top to the bottom of the entire lung for each patient, 27 consecutive sections with the original matrix size at low-dose CT were displayed in a multiformat display (3×3) on three high-spatial-resolution (1600×1200 pixels) liquid crystal display color monitors (CCL202; Totoku Electric, Tokyo, Japan). For cine-mode display, the same 27 CT sections for each study were magnified and stacked on one monitor. The speed or sequence of the image display for cine-mode display was controlled manually by the observer. The windowing in the two image display formats was initially set at lung settings but could be adjusted by the observer to bronchial or mediastinal settings. Two clinical parameters (age and sex) were provided to the observer on the monitor.

The 14 radiologists who participated in this observer study were classified into two groups according to type of display. Observers who used multiformat display (group 1) consisted of five general radiol-

ogists with 7–18 years of experience (mean, 12 years) and one 3rd-year radiology resident. Observers who used cine-mode display (group 2) consisted of three chest radiologists with 16, 17, and 45 years of experience (mean, 26 years), four general radiologists with 5–16 years of experience (mean, 13 years), and one 4th-year radiology resident. The observers in group 2 had more experience than did the observers in group 1.

Radiologists were given the following instructions: (a) We wish to evaluate radiologists' performance in detecting lung cancer without and with a CAD scheme on low-dose CT scans obtained from a screening program. (b) The role of the CAD output is that of a ‘second opinion.’ (c) Twenty-seven low-dose CT studies (with 10-mm-thick sections) that did not or did contain lung cancer and/or non-cancerous abnormalities such as benign nodules and scars are included in this observer study. (d) The observer in this study will be blinded to the number of patients with lung cancer and the performance level of the CAD scheme. (e) Click on the screen by using a mouse (i) to indicate on a bar your confidence level regarding the presence (or absence) of a lung cancer and (ii) to locate the most likely position on each CT scan. You may indicate the cancer location first and (f) click on one of the following four clinical actions: (i) Return to annual screening, (ii) diagnostic thin-section CT in 6 months, (iii) diagnostic thin-section CT in 3 months, or (iv) diagnostic thin-section CT immediately.” The radiologists made their judgments first without and then with the CAD scheme.

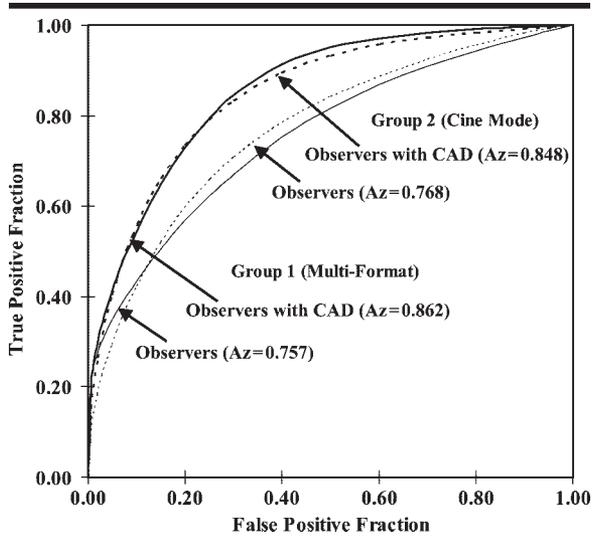


Figure 2. Graph shows ROC curves for detecting cancers missed at CT without and with use of the CAD scheme and for the two display modes. With the CAD scheme, the average A_z value improved significantly from 0.757 to 0.862 for group 1 ($P = .04$) and from 0.768 to 0.848 for group 2 ($P = .01$).

For a training session before the test, we provided five different cases (that were not part of the study set of 27) so that radiologists could learn how to operate the cine-mode interface and how to take into account the computer output in their decision. The reading time was not limited in this study. The average reading time was 48 minutes (range, 27–61 minutes; 1.8 minutes per case).

Statistical Analysis

The confidence level ratings from each observer were analyzed with use of the receiver operating characteristic (ROC) method, and a quasi-maximum-likelihood estimation of the binormal distribution was fitted to the radiologists' confidence ratings (11). The statistical significance of the difference in the area under the ROC curve (A_z) between observer readings without and with the CAD scheme was tested with use of the Dorfman-Berbaum-Metz method (12), which included both reader variation and case sample variation by means of an analysis of variance approach. Localization ROC (LROC) curves (13) for observers without and with the CAD scheme were also determined for each reading condition.

The "proper" binormal model (14) was used to fit the ROC and LROC curves (Metz CE, written communication, 2004). In this study, localization was considered correct if the center of the cancer lesion was located within 15 mm from the point marked by the observer. The distance cri-

terion of 15 mm was based on the fact that our database contained lesions with diameters as large as 26 mm. The distance was computed automatically by the user interface program. The sensitivity in this study was defined on the basis of the number of cancer lesions that were correctly located by an observer regardless of the confidence level ratings. The statistical significance of the difference in sensitivities between the computer outputs and the observer readings without and with the CAD scheme was tested by means of a confidence interval variation alone (15). The statistical significance of the difference in sensitivities between radiologists without and with the CAD scheme and in clinical actions between a beneficial and a detrimental effect of the CAD scheme for each of the studies that did or did not contain a lung cancer was estimated with use of the Student paired t test for the 14 radiologists. In general, $P < .05$ was considered to indicate a statistically significant difference.

RESULTS

Radiologist Performance

With use of the CAD scheme, the average A_z value improved significantly from 0.763 to 0.854 for the 14 radiologists ($P = .002$), from 0.757 to 0.862 for group 1 ($P = .04$), and from 0.768 to 0.848 for group 2 ($P = .01$) (Table 1, Fig 2). No significant difference in the aver-

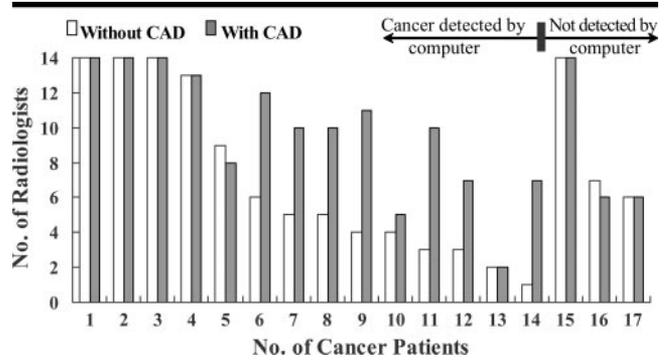


Figure 3. Bar graph shows the number of radiologists who correctly detected cancer in each of 17 patients with lung cancer with and without the use of the CAD scheme. In eight patients, the CAD scheme had a beneficial effect for one to seven radiologists. In two patients, the use of CAD had a detrimental effect for two radiologists. (See details in Discussion.)

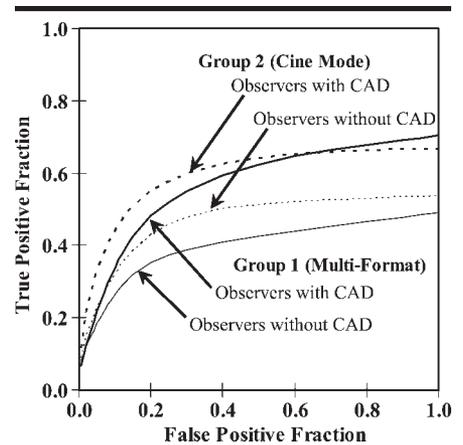


Figure 4. Graph shows LROC curves in the detection of cancers missed at CT for radiologists with and without use of the CAD scheme and with two display modes. The LROC curve was improved for groups 1 and 2 with use of the CAD scheme.

age A_z values between the two groups was found for radiologists without ($P = .82$) and with ($P = .63$) CAD.

In eight of the 17 patients with lung cancer, the CAD scheme helped from one to seven radiologists find the cancers (Fig 3). In two patients, CAD had a detrimental effect for two radiologists. The average LROC curves for the 14 radiologists without and with the CAD scheme in the two groups are shown in Figure 4. Figure 5 shows images from a patient in whom the use of CAD helped seven radiologists detect a cancer lesion.

With use of the CAD scheme (sensitivity, 82% [14 of 17 cancers]), the average sensitivity in the detection of 17 cancers improved significantly—from 52% (124 of 238 observations) to 68% (163 of 238 observations) for the 14 radiologists ($P <$

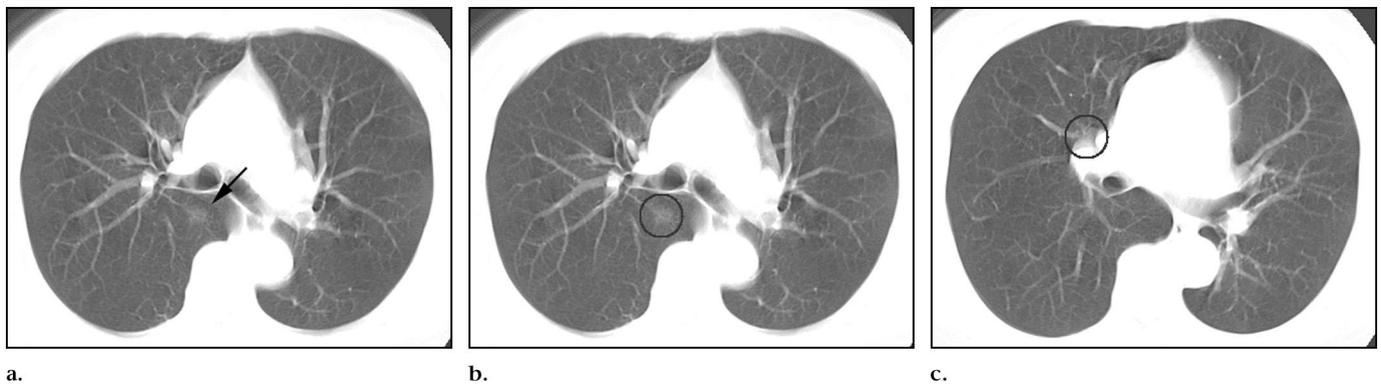


Figure 5. Images obtained in a 69-year-old woman in whom CAD was helpful. Transverse CT scans show (a) a missed lung cancer (arrow) with pure GGO in the right upper lobe, (b) the cancer, and (c) a false-positive finding. Circles in **b** and **c** indicate the computer detections. In this patient, 10 radiologists did not detect the cancer without CAD, whereas CAD helped seven radiologists find the cancer.

TABLE 2
Number of Patients in Whom Important Clinical Action Related to Follow-up Was Changed by 14 Radiologists Owing to CAD

Observer No.	Patients with CAD-revealed Lung Cancer*		Patients without Lung Cancer†	
	Beneficial Effect	Detrimental Effect	Beneficial Effect	Detrimental Effect
1	2	0	0	1
2	2	1	0	0
3	3	1	0	0
4	5	0	1	0
5	0	0	0	0
6	6	0	1	0
7	2	0	0	0
8	2	2	0	0
9	0	0	0	0
10	0	0	0	0
11	5	0	0	0
12	1	0	0	1
13	0	0	0	1
14	2	0	1	0

Note.—“Beneficial effect” indicates a change to follow-up for a patient with lung cancer and a change to screening for a patient without lung cancer; “detrimental effect” indicates a change to screening for a patient with lung cancer or a change to follow-up for a patient without lung cancer.

* The mean number of such patients in whom the use of CAD had a beneficial effect was 2.1 ± 2.0 (standard deviation); the mean number in whom it had a detrimental effect was 0.3 ± 0.6 . The difference between a beneficial effect and a detrimental effect among these patients was statistically significant ($P = .005$).

† Patients in whom CAD revealed a condition other than lung cancer. The mean number of such patients in whom the use of CAD had a beneficial effect was 0.2 ± 0.4 ; the mean number in whom it had a detrimental effect was also 0.2 ± 0.4 .

.001), from 49% (50 of 102 observations) to 71% (72 of 102 observations) for group 1 ($P = .02$), and from 54% (74 of 136 observations) to 67% (91 of 136 observations) for group 2 ($P = .006$). The sensitivity of the CAD scheme alone was greater than that of the radiologists alone ($P < .001$) and that of the radiologists with CAD ($P < .001$), although the specificity was lower. No significant difference in sensitivities was found between the two viewing modes for radiologists without ($P = .44$) and with ($P = .71$) the use of CAD.

Clinical Actions

For the four clinical actions described earlier (ie, return to annual screening or perform diagnostic thin-section CT in 6 months, 3 months, or immediately), we attempted to quantify the changes in clinical action attributable to use of the CAD scheme. For patients with a lung cancer, the average number for whom clinical actions were changed for a beneficial effect (ie, a “step up”) (3.3) was greater than the number for whom clinical actions were changed for a detrimen-

tal effect (ie, a “step down”) (0.4) ($P < .001$). For patients without a lung cancer, the average numbers affected by the CAD scheme for a beneficial effect (step down) and a detrimental effect (step up) were 0.5 and 0.3, respectively ($P = .27$).

Table 2 shows the number of patients for whom the important clinical action related to follow-up was influenced positively or negatively by the 14 radiologists. For these patients, the difference between the mean number of patients in whom the action was changed from screening to follow-up (2.1 patients) and the mean number of patients in whom the action was changed from follow-up to screening (0.3 patients) was significant ($P = .005$). For patients without a lung cancer, no statistically significant difference between a beneficial effect (a change from follow-up to screening [in 0.2 patients]) and a detrimental effect (a change from screening to follow-up [in 0.2 patients]) owing to use of the CAD scheme was found for the radiologists ($P > .99$).

DISCUSSION

It has been reported (16–21) that the use of CAD has the potential to improve diagnostic accuracy in the detection of lung nodules on chest radiographs and CT scans. In previous studies with chest radiography, however, some abnormal cases—each with one lung nodule—and some normal cases were used for the observer test, and the radiologists’ performance in terms of their confidence level regarding the presence or absence of a nodule was evaluated by means of ROC analysis without localization (16,17). For developing CAD schemes for use with relatively thick (18,19) or thin CT sections (20,21), the number of lung nod-

ules was generally not limited to a single nodule in each examination, and the sensitivity with which radiologists correctly detected the nodule, regardless of the confidence level, was commonly used as a measure of the radiologists' performance. In previous CT-based studies (18–21), the truth for the nodules was established by radiologist consensus—not according to pathologic results—because most small nodules are benign and do not undergo biopsy or resection.

There were some differences between the present study and the previous studies, as follows: In the present study, (a) the CAD scheme was developed by using missed lung cancers, which were confirmed at surgery; (b) the mean diameter of the cancers was 12 mm (all were at least 6 mm), and the CT findings for the cancers included lesions with pure GGO, mixed GGO, and solid opacity; and (c) ROC, LROC, and sensitivity analyses were used to evaluate radiologists' performance in the detection of subtle cancers without and with CAD. The importance of these differences is discussed in the next paragraphs.

Missed lung cancers include the most difficult cases for detection in clinical work and mass screening programs, and several investigators have reported the possible reasons for missing lung cancers on CT scans (4,5,22,23). In our series (5), lung cancers were missed mainly because they had low attenuation (eg, they were of small size and/or were faint lesions with GGO) or because of the presence of large structured noise elements (normal structures and/or complex backgrounds caused by other disease) or both. In addition, the cancers had poor conspicuity as defined by Kundel and Revesz (24). In general, the missed cancers corresponded to earlier visible findings in the same locations at previous examinations—findings that had been identified as abnormal according to radiologists' consensus. However, in a previous study by Austin et al (25) of radiologists' performance alone, each of six radiologists, who were biased by knowledge that the patients had lung cancers that were missed on chest radiographs, missed cancer in a mean of 26% of 22 patients. The main purpose of our study was to identify whether unassisted radiologists could identify these previously missed cancers in the context of an observer study and to evaluate whether a CAD scheme could help them detect the cancers missed on CT scans.

Diederich et al (26) reported that more than 70% of noncalcified nodules are 5 mm or smaller, and no lung cancers were

found among those small lesions in CT screening programs for lung cancer at baseline. Similar findings have also been reported by Swensen et al (27). Henschke et al (28) reported that the frequency with which malignancy was or could have been diagnosed when the largest noncalcified nodule was smaller than 5 mm in diameter was very low (0 of 378). The nodules with pure or mixed GGO on CT scans in lung cancer screening programs were more likely to be malignant than were solid nodules (29,30). Although there was a limitation in the low-dose CT protocol with 10-mm-thick sections used in our study, the cancers were at least 6 mm in diameter, and the CT findings for the cancers included lesions with pure GGO, lesions with mixed GGO, and lesions with solid opacity. We believe, therefore, that it may be more important for a CAD scheme used as a "second opinion" to detect relatively large nodules with or without GGO; such nodules include primary lung cancers more frequently than small nodules, most of which are benign lesions, do.

Basically, ROC analysis without localization (11,12) can help correctly evaluate observer performance in the detection of the presence (or absence) of a lesion on medical images when each image does not include obvious false-positive findings, provided that the number of patients is sufficiently large. However, because chest CT scans may contain pulmonary vessels or focal lung diseases that have an appearance that is similar to that of nodular lesions, high positive confidence level ratings by radiologists for a given CT study do not always correspond to true-positive findings (lung cancers) but instead sometimes correspond to false-positive findings. With use of LROC analysis (13), only the responses with correct localization are evaluated for each reading condition, although a proper statistical test for practical use in evaluating the difference between the curves is still unavailable. The shortcoming of LROC analysis for estimating sensitivity is that the radiologist's performance is evaluated only for patients with true-positive findings and not for patients with true-negative findings. Therefore, in this study, we decided to evaluate the performance with three methods—that is, ROC, LROC, and sensitivity analysis—and the results obtained with all three methods showed that the diagnostic accuracy of the radiologists improved with use of the CAD scheme.

Although the radiologists in our study were able to recognize the presence of some subtle lung cancers, they could not be sure whether the CT features of the lesion were indicative of malignancy even when the computer marked the lesion. The possible reasons why the sensitivity for radiologists who used CAD did not reach at least 82% include the fact that the radiologists were not familiar with the appearance of early lung cancers at CT, especially at thick-section CT. In addition, the sensitivity of the radiologists for detecting cancer lesions was affected by some findings such as scars and vertically oriented pulmonary vessels, which had an appearance similar to that of nodular lesions on CT scans in this observer study. In addition, false-positive computer findings would have an effect on radiologists' performance in the detection of lung cancer. We noted that radiologists tended to ignore the CAD output more frequently for studies with a large number of false-positive findings (eight per study, the largest in our scheme) than for those with a small number of false-positive findings. In a previous observer study of the use of LROC analysis in the detection of clustered microcalcifications on mammograms, Chan et al (31) reported that radiologists' diagnostic accuracy with CAD was further improved by reducing the computer's false-positive rate (from four to one false-positive finding per image).

In this observer study, the use of CAD had a detrimental effect in two patients for two radiologists. In one patient, a radiologist detected a cancer lesion without CAD with a confidence level of 0.46 and made a recommendation to follow up the cancer with diagnostic thin-section CT in 3 months. The computer indicated the cancer lesion and eight false-positive findings. With use of CAD in the same patient, a different radiologist changed the location from cancer to a false-positive finding (vertical pulmonary vessel) with a confidence level of 0.59 and did not change the clinical action. In another patient, another radiologist detected a cancer lesion without CAD with a confidence level of 0.31 and recommended follow-up with CT in 6 months. The computer did not mark the cancer lesion but indicated three false-positive findings. The radiologist who used CAD also changed the location from cancer to a false-positive finding (vertical pulmonary vessel) with a confidence level of 0.46 and did not change the clinical action. Therefore, when the CAD scheme yields false-positive findings that are very

similar to true-positive findings, it may have a detrimental effect on the observers' performance when the task involves the detection of only one lesion at CT in an observer study. If radiologists were allowed to identify more than one lesion in an observer study, however, it is possible that they might elect to keep the cancer as detected initially and add the false-positive finding as a further suspicious area.

In recent CT screening programs, most images were reviewed in a multiformat display (film- or monitor-based viewing) and/or a cine-mode display (1–3,26,27). The cancers in this observer study were missed in the Nagano lung cancer screening project, in which a multiformat display (3×4 or 4×4) on two high-spatial-resolution (1728×2304 matrix) monitors was used (5). A similar multiformat viewing mode was used in our study by the radiologists in group 1. In general, cine viewing of CT scans of the chest is believed to improve radiologists' ability to detect lung nodules compared with film-based viewing (32,33). Tillich et al (33), however, found no significant difference between cine and film-based viewing in the detection rate of pulmonary nodules (metastases) larger than 5 mm in diameter. We also did not find a significant difference between the two viewing modes in the detection of primary lung cancers (≥ 6 mm) missed in a CT screening program. The limitations of this study include the facts that the low-dose CT sections were thick (10 mm), rather than thin, and the radiologists differed in the two groups. It was not the purpose of our study to compare diagnostic accuracy with the cine or multiformat mode but rather to determine that the benefits of CAD were substantial, independent of the display mode used.

In summary, lung cancers missed at low-dose CT screening were very difficult to detect, even in an observer study; the use of CAD, however, improved the radiologists' performance in the detection of these subtle cancers. In addition, CAD can help radiologists make recommendations for follow-up.

Acknowledgments: The authors are grateful to Ulrich Bick, MD, Alexandra Funaki, DO, John Fennessy, MD, Gen Iinuma, MD, Edward Michals, MD, Masaki Matsusako, MD, Peter MacEneaney, MD, Sidney Regalado, MD, Christopher Straus, MD, Gregory Scott Stacy, MD, Shuji Sakai, MD, Taylor Stroud, MD, Ira Wolke, MD, and Chaotong Zhang, MD, for participating as observers; to Charles E. Metz, PhD, and Lorenzo Pesce, PhD, for the use of proper binormal model in ROC and LROC

curves; and to Elisabeth Lanzl, AM, for improving the manuscript.

References

- Kaneko M, Eguchi K, Ohmatsu H, et al. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. *Radiology* 1996;201:798–802.
- Sone S, Takashima S, Li F, et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. *Lancet* 1998;351:1242–1245.
- Henschke CI, MacCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999;354:99–105.
- Kakinuma R, Ohmatsu H, Kaneko M, et al. Detection failures in spiral CT screening for lung cancer: analysis of CT findings. *Radiology* 1999;212:61–66.
- Li F, Sone S, Abe H, MacMahon H, Armato SG III, Doi K. Missed lung cancers in low-dose helical CT screening obtained from a general population. *Radiology* 2002;225:673–683.
- Armato SG III, Li F, Giger ML, MacMahon H, Sone S, Doi K. Performance of automated CT nodule detection on missed cancers from a lung cancer screening program. *Radiology* 2002;225:685–692.
- Arimura H, Katsuragawa S, Suzuki K, et al. Computerized scheme for automated detection of lung nodules in low-dose CT images for lung cancer screening. *Acad Radiol* 2004;11:617–629.
- Suzuki K, Armato SG III, Li F, Sone S, Doi K. Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose CT. *Med Phys* 2003;30:1602–1616.
- Doi K, MacMahon H, Katsuragawa S, Nishikawa RM, Jiang Y. Computer-aided diagnosis in radiology: potential and pitfalls. *Eur J Radiol* 1999;31:97–109.
- Doi K. Computer-aided diagnosis in digital chest radiography. In: Samei E, Flynn M, eds. 2003 Syllabus: categorical course in diagnostic radiology—physics. Oak Brook, Ill: Radiological Society of North America, 2003; 227–236.
- Metz CE, Herman BA, Shen JH. Maximum-likelihood estimation of receiver operating (ROC) curves from continuously distributed data. *Stat Med* 1998;17:1033–1053.
- Dorfman DD, Berbaum KS, Metz CE. ROC rating analysis: generalization to the population of readers and cases with the jackknife method. *Invest Radiol* 1992;27:723–731.
- Starr SJ, Metz CE, Lusted LB, Goodenough DJ. Visual detection and localization of radiographic images. *Radiology* 1975;116:533–538.
- Metz CE, Pan X. "Proper" binormal ROC curves: theory and maximum-likelihood estimation. *J Math Psychol* 1999;43:1–33.
- Metz CE. Quantification of failure to demonstrate statistical significance: the usefulness of confidence intervals. *Invest Radiol* 1993;28:59–63.
- MacMahon H, Engelmann R, Behlen FM, et al. Computer-aided diagnosis of pulmonary nodules: results of a large-scale observer test. *Radiology* 1999;213:723–726.
- Kakeda S, Moriya J, Sato H, et al. Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system. *AJR Am J Roentgenol* 2004;182:505–510.
- Awai K, Murao K, Ozawa A, et al. Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. *Radiology* 2004;230:347–352.
- Wormanns D, Fiebich M, Saidi M, Diederich S, Heindel W. Automatic detection of pulmonary nodules at spiral CT: clinical application of a computer-aided diagnosis system. *Eur Radiol* 2002;12:1052–1057.
- McCulloch CC, Kaucic RA, Mendonca PR, Walter DJ, Avila RS. Model-based detection of lung nodules in computer tomography exams. *Acad Radiol* 2004;11:258–266.
- Brown MS, Goldin JG, Suh RD, McNitt-Gray MF, Sayre JW, Aberle DR. Lung micronodules: automated method for detection at thin-section CT—initial experience. *Radiology* 2003;226:256–262.
- Gurney JW. Missed lung cancer at CT: imaging findings in nine patients. *Radiology* 1996;199:117–122.
- White CS, Romney BM, Mason AC, Austin JH, Miller BH, Protopoulos Z. Primary carcinoma of the lung overlooked at CT: analysis of findings in 14 patients. *Radiology* 1996;199:109–115.
- Kundel HL, Revesz G. Lesion conspicuity, structured noise, and film reader error. *AJR Am J Roentgenol* 1976;126:1233–1238.
- Austin JH, Romney BM, Goldsmith LS. Missed bronchogenic carcinoma: radiographic findings in 27 patients with potentially resectable lesion evident in retrospect. *Radiology* 1992;182:115–122.
- Diederich S, Wormanns D, Semik M, et al. Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers. *Radiology* 2002;222:773–781.
- Swensen SJ, Jett JR, Sloan JA, et al. Screening for lung cancer with low-dose spiral computed tomography. *Am J Respir Crit Care Med* 2002;165:508–513.
- Henschke CI, Yankelevitz D, Naidich DP, et al. CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans. *Radiology* 2004;231:164–168.
- Henschke CI, Yankelevitz DF, Mirtcheva R, McGuinness G, McCauley D, Miettinen OS. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. *AJR Am J Roentgenol* 2002;178:1053–1057.
- Li F, Sone S, Abe H, MacMahon H, Doi K. Malignant versus benign nodules at CT screening for lung cancer: comparison of thin-section CT findings. *Radiology* 2004;233:793–798.
- Chan HP, Doi K, Vyborny CJ, et al. Improvement in radiologists' detection of clustered microcalcifications on mammography: the potential of computer-aided diagnosis. *Invest Radiol* 1990;25:1102–1110.
- Seltzer SE, Judy PF, Adams DF, et al. Spiral CT of the chest: comparison of cine and film-based viewing. *Radiology* 1995;197:73–78.
- Tillich M, Kammerhuber F, Reittner P, Riepl T, Stoeffler G, Szolar DH. Detection of pulmonary nodules with helical CT: comparison of cine and film-based viewing. *AJR Am J Roentgenol* 1997;169:1611–1614.