Screening for Lung Cancer with Digital Chest Radiography: Sensitivity and Number of Secondary Work-up CT Examinations[1]

Bartjan de Hoop, MD
Cornelia Schaefer-Prokop, MD, PhD
Hester A. Gietema, MD, PhD
Pim A. de Jong, MD, PhD
Bram van Ginneken, PhD
Rob J. van Klaveren, MD, PhD
Mathias Prokop, MD, PhD

Purpose: To estimate the performance of digital chest radiography for detection of lung cancer.

Materials and Methods: The study had ethics committee approval, and a nested case-control design was used and included 55 patients with lung cancer detected at computed tomography (CT) and confirmed with histologic examination and a sample of 72 of 4873 control subjects without nodules at CT. All patients underwent direct-detector digital chest radiography in two projections within 2 months of the screening CT. Four radiologists with varying experience identified and localized potential cancers on chest radiographs by using a confidence scale of level 1 (no lesion) to 5 (definite lesion). Localization receiver operating characteristic (ROC) analysis was performed. On the basis of the assumption that suspicious lesions seen at chest radiography would lead to further work-up with CT, the number of work-up CT examinations per detected cancer (CT examinations per cancer) was calculated at various confidence levels for the screening population (cancer rate in study population, 1.3%).

Results: Tumor size ranged from 6.8 to 50.7 mm (median, 11.8 mm). Areas under the localization ROC curve ranged from 0.52 to 0.69. Detection rates substantially varied with the observers' experience and confidence level: At a confidence level of 5, detection rates ranged from 18% at one CT examination per cancer to 53% at 13 CT examinations per cancer. At a confidence level of 2 or higher, detection rates ranged from 94% at 62 CT examinations per cancer to 78% at 44 CT examinations per cancer.

Conclusion: A detection rate of 94% for lung tumors with a diameter of 6.8–50.7 mm found at CT screening was achievable with chest radiography only at the expense of a high false-positive rate and an excessive number of work-up CT examinations. Detection performance is strongly observer dependent.

[1] From the Department of Radiology (B.d.H., H.A.G., P.A.d.J., M.P.) and Image Sciences Institute (B.v.G.), University Medical Center, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands; Department of Radiology, Meander Medical Center, Amersfoort, the Netherlands (C.S., H.A.G.); Department of Radiology, Academic Medical Center, Amsterdam, the Netherlands (C.S.); and Department of Pulmonology, Erasmus Medical Center, Rotterdam, the Netherlands (R.J.v.K.). Received July 20, 2009; revision requested August 28; final revision received October 5; accepted October 27; final version accepted November 16. **Address correspondence to** B.d.H. (e-mail: *B.J.dehoop@umcutrecht.nl*).

Chest radiography is still the most commonly used technique in clinical practice to rule out chest disease, to study the effects of treatment, and to monitor patients with chest abnormalities. Computed tomography (CT) has a much higher sensitivity for the detection of small intrapulmonary lesions than does chest radiography, but chest radiography has the advantage of low cost, low radiation dose, and easy accessibility.

Historically, lung cancer screening studies by using cytologic evaluation and/or conventional screen-film chest radiography have yielded disappointing results (1). Screening with conventional chest radiography was therefore considered inappropriate. However, in these studies, analog screen-film techniques were used for chest radiography. The use of modern digital equipment with highly quantum-efficient detectors and elaborate processing tools improves visualization of pulmonary structures with chest radiography (2–5) and may, therefore, be a more suitable screening tool than is conventional chest radiography.

Up to now, little has been known about the performance of modern digital radiography for lung cancer screening. We used a nested case-control setup that was based on data from the Dutch-Belgian Randomized Lung Cancer Screening (NELSON) trial (6) to study how the confidence level for the presence of a lesion affected observer performance. Assuming that a positive chest radiographic screening result would lead to initiation of a work-up CT examination, we also estimated the number of work-up CT examinations needed to detect one lung cancer in the NELSON study cohort. The aim was to estimate the performance of digital chest radiography (hereafter referred to as chest radiography) for detection of lung cancer.

Advances in Knowledge

- Direct-detector digital chest radiography can have a detection rate of up to 94% at lung cancer screening but only at the cost of an excessive rate of false-positive results and subsequent additional work-up CT examinations.
- Confidence levels and observer experience have a substantial effect on the diagnostic efficacy of digital chest radiography for the detection of early lung cancers.

Materials and Methods

Study Population

We recruited our patients from two screening sites (Utrecht and Groningen, the Netherlands) and included all 4938 patients who underwent baseline and 1-year follow-up screening until July 2007. The NELSON trial was approved by the ethics committes of both participating institutions, and a waiver was received for our part of the study. All participants in the NELSON trial were former or current heavy smokers (7). A total of 65 lung cancers were detected at baseline screening and at 1-year follow-up in this group, for a cancer rate of 1.3%.

Case Cohort

We recruited our case cohort from all 65 patients in whom a pulmonary malignancy was detected with low-dose CT at one of the two screening sites. All malignancies were histologically proved. The NELSON study setup does not call for obligatory acquisition of a chest radiograph at inclusion time. However, chest radiography is still part of clinical practice during the diagnostic work-up of lesions suspicious for malignancy. Therefore, the cases with positive findings in our study group comprise a selection of patients who underwent chest radiography in two projections as part of the clinical practice during the diagnostic work-up or for preoperative screening. Consequently, all chest radiography was performed after detection of the malignancy by using CT. Ten patients who did not undergo chest radiography within 6 weeks after detection of a suspicious lung nodule were excluded. Thus, the study group consisted of 55 patients with at least one malignant nodule.

The study group included 12 patients whose cancer was detected at 1-year follow-up but was retrospectively visible at baseline CT screening. All 12 nodules were also reported at baseline but did not meet the criteria for referral at that time (6).

Control Subjects

From all participants who were not patients and who were screened at both study sites (n = 4873), we included all participants in whom CT did not show nodules larger than 5 mm in diameter and in whom chest radiography had been performed within 2 months from CT screening for reasons other than a suspicious lung nodule. Chest radiographs

Implication for Patient Care

- Digital chest radiography is substantially less efficient than low-dose CT for detection of tumors at the same stage of disease in a setting of lung cancer screening. With digital chest radiography, detection rates for lung cancer as high as 94% are achievable only at the expense of an excessive number of work-up CT examinations.

Published online
10.1148/radiol.09091308

Radiology 2010; 255:629–637

Abbreviations:
CI = confidence interval
COPD = chronic obstructive pulmonary disease
NELSON = Dutch-Belgian Randomized Lung Cancer Screening trial
PPV = positive predictive value
ROC = receiver operating characteristic

Author contributions:
Guarantors of integrity of entire study, B.d.H., C.S., R.J.v.K., M.P.; 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, B.d.H., H.A.G., B.v.G.; clinical studies, B.d.H., H.A.G., B.v.G, R.J.v.K., M.P.; experimental studies,; statistical analysis, B.d.H., C.S.; and manuscript editing, B.d.H., C.S., H.A.G., B.v.G., R.J.v.K., M.P.

Authors stated no financial relationship to disclose.

for which the radiology report mentioned pulmonary abnormalities other than those related to chronic obstructive pulmonary disease (COPD) were excluded. Seventy-two participants met these criteria. Indications for performance of chest radiography were exclusion of acute cardiovascular disease (n = 18), follow-up for COPD (n = 17), screening for lung abnormalities because of rheumatoid arthritis (n = 13), preoperative screening for cardiovascular surgery (n = 11), unexplained fever (n = 11), trauma (n = 1), and malaise (n = 1).

We tested for differences in the prevalence of COPD in the case and control groups because this disease may affect visibility of nodules. We were able to compare prevalence of COPD because a large subsample of subjects in the whole screening population underwent pulmonary function testing as part of a substudy of the lung cancer screening trial. Pulmonary function testing results were available in 43 (78%) of our patients with lung cancer, 46 (64%) of our control subjects, and 2547 (52%) of participants without cancer in the lung cancer screening trial. The remaining participants did not undergo pulmonary function testing. This testing included forced expiratory volume in 1 second and forced vital capacity. A participant whose ratio of forced expiratory volume in 1 second to forced vital capacity was less than 0.7 was considered to have COPD (8).

Acquisition and Evaluation of Screening CT Scans

All CT scans were acquired and evaluated for nodules according to the NELSON protocol (6). Volume and mean diameter of detected nodules were assessed by using volumetric software (Lung CARE; Siemens, Erlangen, Germany) (6).

Acquisition and Evaluation of Chest Radiographs

Acquisition technique was identical to that for conventional chest radiography performed in our hospital (University Medical Center Utrecht, Utrecht, the Netherlands). All chest radiographs were obtained by using a cesium iodide amorphous silicon flat-panel-detector (ie, direct-detector) unit (DigitalDiagnost; Philips, Best, the Netherlands). Images were processed by using nonlinear multifrequency-band processing (9); parameters recommended by the manufacturer were used. For all patients, posteroanterior and lateral projections were available. Images were evaluated on liquid crystal display monitors (MFGD 3220D; Barco, Kortrijk, Belgium), with a resolution 2048 × 1536 pixels, without and with gray-scale reversal. Options for magnification and adaptation of window settings were used only when observers were unsure about a region.

Cases and control subjects were presented in alphabetical order on the basis of patient name to four independent observers with varying levels of experience: two chest radiologists (C.S. and M.P.) with more than 20 years of experience (observers A and B), a general radiologist with more than 20 years of experience (observer C), and a 3rd-year resident with special experience and interest in chest radiology (observer D). Observers were aware of the study population but did not know the number of malignancies in the group. Nodules smaller than 5 mm in diameter and calcified granulomas were ignored. Posteroanterior and lateral radiographs were evaluated. The observers scored the presence of focal opacities that were suspicious for malignancy by using a five-point confidence scale: level 1, no lesion; level 2, irregularity, probably no lesion; level 3, indeterminate for the presence of a lesion; level 4, lesion probably present; and level 5, lesion definitely present. Readers had to manually localize the lesion on the radiograph. If more than one suspicious area was detected, the observer had to mark the most suspicious area. The chest radiographic reading was considered true-positive only when the localization of the lesion was correct. Observers were not forced to place a marking; they could also rate a radiograph as normal (ie, no nodule present, confidence level 1). Reading time was unlimited and ranged from 140 to 175 minutes for the different observers for the total study; the mean time per patient examination ranged from 70 to 97 seconds.

After acquisition of all reading data, and with knowledge of the CT findings, observer A retrospectively determined whether the lesions that had not been seen by any of the observers were visible on the chest radiograph. In addition, the same observer determined whether lesions were obscured by anatomic structures on the posteroanterior radiograph.

Statistical Analysis

A nested case-control setup in the NELSON trial was used. This design uses a selection of the control subjects to represent all control subjects in the full cohort, enabling reconstruction of the results for the full cohort (10). For that purpose, the results for the control group are multiplied by the quotient of one divided by the sample fraction. The cases of cancer in which no chest radiograph was available (10 [15.4%] of 65) were excluded. To determine the sample fraction, the number of noncases, therefore, had to be adjusted accordingly (4873 − 15.4% = 4123) to match the 1.3% cancer rate in the NELSON cohort at the time of this study. The sample fraction in our study, therefore, was 72/4123.

We tested whether the control subjects in the observer study were representative of all noncases in the full cohort (a requirement of a nested case-control study). Categorical variables were evaluated by using a χ^2 test, and continuous variables were evaluated with a Student t test.

In all calculations, we assumed 100% sensitivity for CT. Confidence intervals (CIs) were calculated by using the Wilson score. The following four parameters were used to assess the performance of chest radiography as a screening tool for lung cancer:

1. Localization receiver operating characteristic (ROC) curve analysis. This analysis summarizes the reader's sensitivity and specificity in a single value. Localization ROC analysis differs from normal ROC analysis in that it takes into account the correct localization of a reader's marking. Localization ROC analysis was performed as described by Swensson (11). Jackknife free-response

ROC software (Chakraborty D, University of Pittsburgh, Pittsburgh, Pa) (12,13) was used to test for significant differences between the localization ROC areas.

2. Sensitivity and specificity. These statistics were calculated individually for each reader and each confidence level. Only correctly localized tumors were considered true-positive findings.

3. Number of work-up CT examinations per chest radiography–detected cancer. This number describes how many CT scans need to be obtained in the whole screening cohort to find one case of a positive finding that was correctly suspected from a finding at chest radiography. It is based on the assumption that a suspicious chest radiographic finding leads to initiation of CT for further diagnostic work-up. The nested case-control setup allows us to estimate the positive predictive value (PPV) of chest radiography for the total screening population. Because PPV describes the proportion of true-positive chest radiographs among all positive radiographs, PPV is equal to the proportion of work-up CT examinations with positive results (cancer at the suspected location) among all work-up CT examinations performed for a positive chest radiograph. Therefore, the number of work-up CT examinations per chest radiography–detected cancer can be calculated as 1/ PPV.

4. Total percentage of malignancies detected during CT work-up. There is a small chance that work-up CT will reveal cancer in a different location than is suspected at chest radiography. This chance increases with the total number of work-up CT examinations performed. The total percentage of detected malignancies, therefore, includes the reader's true-positive findings at chest radiography plus an estimated number of malignancies incidentally found at the work-up CT examinations performed for a false-positive density on a chest radiograph. To calculate the total percentage of malignancies detected at CT, we also determined the chance that a CT examination, on the basis of a false-positive chest radiography report, would incidentally reveal a malignancy.

Table 1

Demographic Characteristics of Study Participants

Characteristic	Cases ($n = 55$)	Control Subjects ($n = 72$)	All Noncases Minus Control Subjects ($n = 4801$)	*P* Value, Cases versus Control Subjects	*P* Value, Control Subjects versus Noncases
Mean age (y)	64.5 ± 6.4*	62.4 ± 5.3*	61.4 ± 5.5*	.52	.46
No. of men/women	49/6	59/13	4291/510	.32	.02
No. of patients with COPD (%)	63	38	38	.14	.88

*Data are the mean ± standard deviation.

Table 2

Diameters of Detected and Undetected Nodules

Observer	Median Diameter of Detected Nodules (mm)	Median Diameter of Undetected Nodules (mm)	*P* Value
A	12.2 (7.6–50.7)	10.6 (4.2–18.5)	.1
B	11.8 (7.6–50.7)	11.4 (4.2–21.0)	.09
C	11.8 (8.1–50.7)	10.7 (6.8–35.4)	.36
D	17.1 (8.1–50.7)	11.4 (3.5–21.0)	.001*

Note.—Only probably present and definite nodules (confidence levels 4 and 5) were counted. Numbers in parentheses are ranges.

*For observer D, the detected nodules were significantly larger than the nodules that were not detected by this observer. Observer D detected a significant difference between the nodule groups.

This chance is equal to the number of nondetected malignancies divided by the number of participants for whom chest radiography had negative or false-positive results. This chance was multiplied by the number of CT examinations performed for false-positive findings in radiography reports to calculate the number of malignancies incidentally detected at CT. The number of true-positive chest radiographs plus malignancies incidentally detected at CT formed the total number of detected malignancies. These incidentally detected lesions will also result in a slightly different number of CT examinations per detected cancer, which hereafter we call CT examinations per cancer. *P* values less than .05 were considered to indicate a significant difference.

Results

Study Population

Our case cohort did not significantly differ from our control subject cohort with respect to age, sex, and prevalence of COPD. Compared with the noncases in the full cohort, age and prevalence of COPD were not different in our control subject cohort; however, the control subject cohort contained relatively more women ($P = .02$) than the noncase group in the full cohort (Table 1).

Malignancies

The diameter of the malignancies ranged from 6.8 to 50.7 mm (median, 11.8 mm). Four malignancies manifested as ground-glass opacity on a CT scan: One was nonsolid and three were part-solid lesions. Two of the part-solid lesions with ground-glass opacity were detected by three observers; the other two were not detected by any observer. Most lesions were located in the right upper lobe ($n = 25$). The remaining lesions were located in the right middle lobe ($n = 3$), right lower lobe ($n = 9$), left upper lobe ($n = 8$), and left lower lobe ($n = 10$). On the posteroanterior radiograph, 26 lesions were at least partially obscured by overlying anatomic structures, such as hilar vascular structures ($n = 7$), the clavicle ($n = 9$), the heart ($n = 3$), a rib ($n = 6$), or the diaphragmatic recess

(n = 1). Obscured lesions were responsible for a mean of 43% of all undetected lesions; individual rates were 36% for observer A, 41% for observer B, 39% for observer C, and 55% for observer D.

Twenty-four (44%) malignancies were correctly localized by all observers. Seven (13%) malignancies were localized by none of the observers. Three of these seven were not visible on chest radiographs even with knowledge of the CT findings (Fig 1).

The median diameter of correctly localized lesions ranged from 11.8 to 17.1 mm, depending on the reader. The median diameter of undetected lesions ranged from 10.6 to 11.1 mm. The difference in size was significant (P = .001) for only one reader (Table 2).

Localization ROC Analysis

The area under the localization ROC curve ranged from 0.52 for observer C to 0.69 for observer A. Localization ROC analysis results indicated better performance for the two chest radiologists (observers A and B) compared with the performance of observers C and D, but differences were significant ($P < .05$) only for observers A and C (Fig 2).

Sensitivity

At the highest level of confidence (level 5, lesion definitely present), sensitivity for correctly localized malignant lesions on digital chest radiography varied from 18% (95% CI: 10%, 32%) to 49% (95% CI: 36%, 63%) at a specificity of 100% (95% CI: 94%, 100%) to 92% (95% CI: 82%, 97%). The false-positive rates in the control group at this level of confidence ranged from 0% (0 of 72) to 13% (9 of 72) for observers D and C, respectively.

When lesions rated as probably present (level 4) were taken into account, sensitivity increased to 36% (95% CI: 24%, 50%) for observer D and 73% (59%, 84%) for observer A, with specificities of 99% (95% CI: 50%, 76%) and 82% (95% CI: 71%, 90%), respectively (Table 3).

Most lesions that had been rated as definitely present at chest radiography turned out to be correctly localized malignancies: The PPVs of nodules rated as definitely present were 82% (95% CI: 64%, 92%) for observer A, 89% (95% CI: 71%, 97%) for observer B, 71% (95% CI: 52%, 85%) for observer C, and 100% (95% CI: 66%, 100%) for observer D.

Number of Work-up CT Examinations per Cancer Detected at Chest Radiography

When only definite lesions (confidence level 5) were taken into account, we calculated that positive calls would have led to initiation of one (95% CI: 1, 2) to 24 (95% CI: 16, 38) work-up CT examinations per lung cancer detected at chest radiography in the total screening cohort. When lesions rated as probably present (level 4) were taken into account, the number of CT examinations per cancer detected at chest radiography ranged from four (95% CI: 3, 6) to 46 (95% CI: 32, 68). Table 3 summarizes the relationship between sensitivity and number of work-up CT examinations.

Total Percentage of Malignancies Detected at CT Work-up

The percentage of additional, incidentally detected malignancies ranged from 0% to 8% of the total number of malignancies for the various observers at

Figure 1

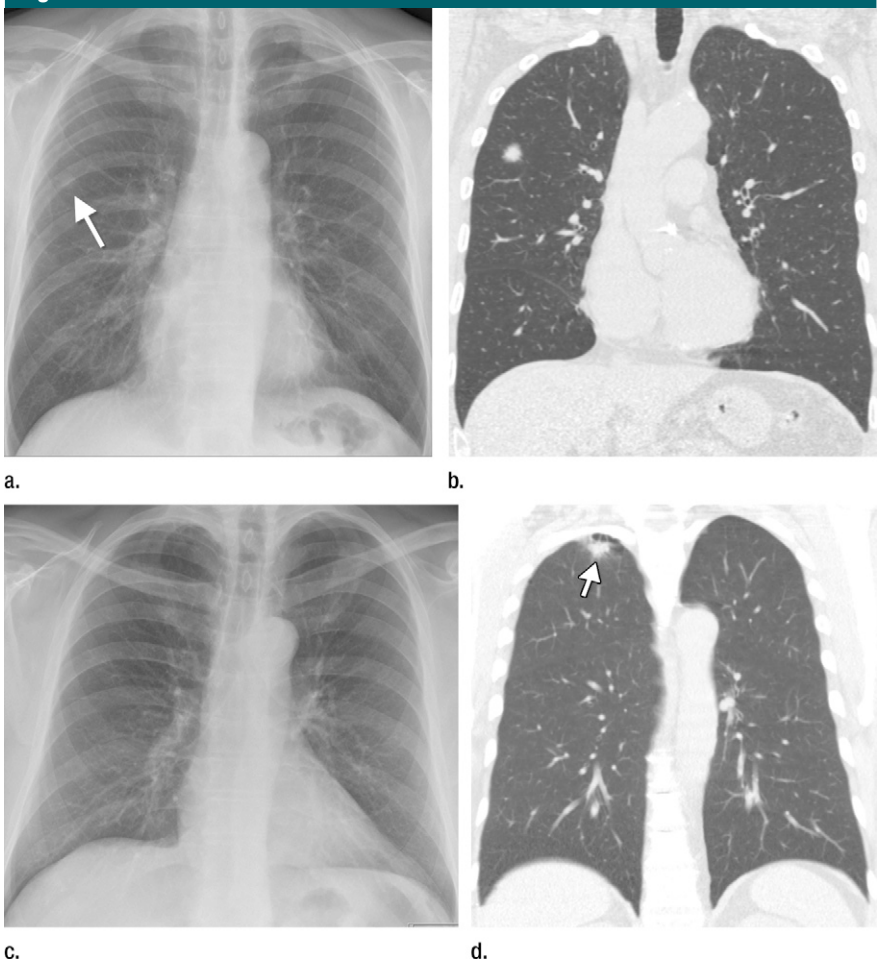

Figure 1: **(a, b)** Small 11.4-mm- diameter malignancy that was correctly localized and scored as definitely present by all observers (arrow) at **(a)** digital chest radiography and **(b)** CT. **(c)** Digital chest radiograph and **(d)** CT scan of malignancy (arrow) that was not detected by any of the observers during the study or retrospectively with knowledge of CT findings.

Figure 2

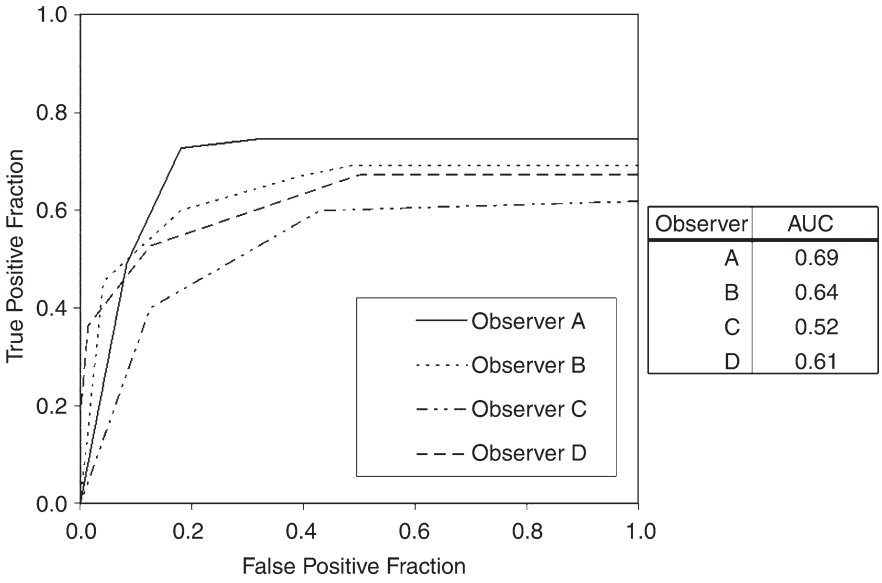

Figure 2: Localization ROC curves and areas under the curve *(AUC)* for the different observers. Note the better performance of the chest radiologists (observers A and B). The difference between observers A and C was significant ($P < .05$).

the highest level of confidence. At this confidence level, the total percentage of malignancies detected during work-up varied from 53% with 13 work-up CT examinations per cancer to 18% with one work-up CT examination per cancer. For all observers, the percentage of additional incidentally detected malignancies increased as the confidence level decreased (Table 4).

Discussion

Even in the era of digital imaging, detection of lung cancer with chest radiography is a challenging task that shows high interreader variability. When only lesions rated as probably present and definitely present were considered to require further diagnostic work-up, the number of detected malignant lesions ranged from 37% to 78% (Table 4), depending on the reader. Thus, in instances in which we would rely on chest radiography alone, 22%–63% of the lung cancers would be missed at a stage of disease at which they could be detected with CT. Even for the most experienced chest radiologist, a detection rate beyond 90% (eg, 94%) could be achieved only when the threshold level for performing CT was reduced to the lowest confidence level (irregularity, probably no lesion) and malignancies present at a work-up CT examination for false-positive chest radiographs were counted as well. This, however, would have resulted in 62 work-up CT examinations per cancer, adding up to 3191 CT examinations in the whole population and still leaving three cancers undetected. The PPV of CT during the first round of the NELSON trial was 35.7% (14), resulting in approximately three referrals to a pulmonologist to diagnose one cancer. This finding shows that even modern digital chest radiography has a much worse performance than does CT for lung cancer screening and is hampered by large numbers of false-positive readings or low detection rates.

On the other hand, despite the lower detection rate of chest radiography, about half of the CT-proved malignancies were detected with chest radiography by all readers. This may substantially affect the power of randomized lung cancer screening trials if chest radiography is used in the control arm (15). However, screening with chest radiography commonly involves only frontal radiographs; we used posteroanterior *and* lateral images in our study. Whether using chest radiography as a primary screening tool would also influence the outcome of patients in terms of mortality and survival cannot be determined on the basis of these data alone because the biological behavior of primary bronchogenic carcinoma greatly varies, and the prognosis of lesions differs according to their time of diagnosis. On average, lesions detected at chest radiography were larger than undetected lesions, although the differences in size were small.

We showed that the area under the curve for the localization ROC curve increased substantially with the observer's years of experience and subspecialization. However, the difference in performance was significant only for observers A and C. The two chest radiologists (observers A and B) had the greatest area under the curve, but the resident with special experience in chest radiology (observer D) performed better than did the general radiologist with more than 20 years of experience (observer C) . These results suggest that special training is advantageous for reading chest radiographs in a lung cancer screening setting.

Observer D showed a generally lower level of confidence and, therefore, a stronger increase in sensitivity than did the other three observers when lesions detected with lower levels of confidence were also taken into account. Observer behavior, therefore, has a stronger effect on performance with chest radiography than it does with CT (16).

Sensitivity of chest radiography for malignancies detected in a lung cancer screening trial has previously been studied by using conventional radiography (17). The diameter of the malignancies was similar to those in our study. Similar percentages of nodules were obscured by anatomic structures. The main difference refers to the imaging technique (digital versus conventional) and the inclusion of lateral radiographs in our study. Conventional chest radiography had a sensitivity of 23% and a specificity of 96%. While maintaining the same level of specificity,

Table 3

Relationship between Sensitivity, Specificity, and Number of Work-up CT Examinations per Cancer Detected

Observer and Minimum Level of Confidence Included*	Sensitivity (%)	Specificity (%)	No. of Work-up CT Examinations per Cancer Detected†
Observer A			
Irregularity, probably no lesion; levels 2 + 3 + 4 + 5	75 (61, 85)	24 (15, 35)	78 (57, 107)
Indeterminate; levels 3 + 4 + 5	75 (61, 85)	68 (56, 78)	33 (24, 45)
Lesion probably present; levels 4 + 5	73 (59, 84)	82 (71, 90)	20 (14, 27)
Lesion definitely present; level 5	49 (36, 63)	92 (82, 97)	14 (9, 20)
Observer B			
Irregularity, probably no lesion; levels 2 + 3 + 4 + 5	69 (55, 80)	51 (39, 63)	54 (39, 75)
Indeterminate; levels 3 + 4 + 5	67 (53, 79)	60 (47, 71)	46 (33, 64)
Lesion probably present; levels 4 + 5	60 (46, 73)	82 (71, 90)	24 (17, 33)
Lesion definitely present; level 5	45 (32, 59)	96 (87, 99)	8 (5, 12)
Observer C			
Irregularity, probably no lesion; levels 2 + 3 + 4 + 5	60 (46, 73)	56 (43, 67)	57 (40, 81)
Indeterminate; levels 3 + 4 + 5	60 (46, 73)	57 (45, 68)	55 (39, 78)
Lesion probably present; levels 4 + 5	53 (39, 66)	68 (56, 78)	46 (32, 68)
Lesion definitely present; level 5	40 (27, 54)	88 (77, 94)	24 (16, 38)
Observer D			
Irregularity, probably no lesion; levels 2 + 3 + 4 + 5	67 (53, 79)	50 (38, 62)	57 (40, 79)
Indeterminate; levels 3 + 4 + 5	53 (39, 66)	88 (77, 94)	19 (13, 27)
Lesion probably present; levels 4 + 5	36 (24, 50)	99 (50, 76)	4 (3, 6)
Lesion definitely present; level 5	18 (10, 32)	100 (94, 100)	1 (1, 2)

Note.—The cancer rate in the lung cancer screening cohort was 1.3%. While for readers A, B, and C, inclusion of the lower levels of confidence (levels 2 and 3) did not substantially increase sensitivity, observer D had a generally low level of confidence and would gain in sensitivity if low levels of confidence were included. Numbers in parentheses are 95% CIs.

* Five-point ROC scale: level 1, no lesion; level 2, irregularity, probably no lesion; level 3, indeterminate for the presence of a lesion; level 4, lesion probably present; and level 5, lesion definitely present.

† Numbers of work-up CT examinations per cancer detected at chest radiography.

digital radiography and inclusion of lateral images showed a sensitivity approximately twice as high for each observer. Researchers in several other studies (18–22) have assessed the sensitivity of conventional and digital chest radiography; sensitivities have ranged from 36% to 84% depending on the study population. These studies retrospectively assessed the performance of chest radiography for CT-proved lesions but did not separately quantify the number of lesions that were detected at CT but were impossible to visualize on a projection radiograph (18,21,23).

Because of the procedure we used to select participants, the control patients may have been skewed toward COPD. It is known that the increased nodular and reticulonodular markings on a chest radiograph, often seen in smokers with COPD, affect the ability of an observer to spot focal opacities (24). The results of pulmonary function testing, however, showed that the prevalence of COPD did not differ between patients in our control cohort and the noncase group in the full cohort.

The main limitation of our study was the absence of an independent reference standard. Only malignancies detected at CT were included. Sensitivity of CT in the NELSON trial, defined as the ratio between cancers detected at CT and all pulmonary cancers diagnosed during the 1st year after CT screening, was greater than 94% (14). Within the screening trial, only one interval cancer was misinterpreted at CT; this cancer turned out to be retrospectively visible at screening CT. Thus, our cohort of radiologically detectable cases would have been practically the same had we used CT or an independent reference standard.

The cancer rate was higher in our study than it was in the NELSON cohort. Although the observers were not aware of the exact disease frequency, they were aware of the higher prevalence, which could have led to overdiagnosis (25). Conversely, most lesions rated as definitely present were indeed malignant. Furthermore, the number of CT examinations per detected cancer may have been underestimated because control subjects in our study did not have any nodules larger than 5 mm, as proved by using CT. In a usual screening situation, however, a certain percentage of patients without malignancies would have had benign nodules at presentation, which again could have led to false-positive readings. Finally, four observers performed the readings in this study. Although they had varying experience and rating behavior, they still represent a selected group. More observers are needed to quantify the effect of reader behavior if chest radiography is to be used as a screening tool on a large scale.

In conclusion, high rates of lung cancer detection can be achieved with digital chest radiography at a stage when lesions are seen at CT screening, but only at the expense of a low specificity that results in an excessive number of work-up CT examinations. The detection performance with chest radiography strongly depends on the

Table 4

Estimated Total Percentages of Malignancies Detected during Screening Work-up in Screening in Which Positive Chest Radiograph Would Lead to Initiation of Work-up CT

Observer and Minimum Level of Confidence Included *	Sensitivity of Chest Radiography (%)	Malignancies Incidentally Detected during Work-up CT Examination (%)	Total Malignancies Detected during Work-up Examination (%)†	No. of Work-up CT Examinations per Cancer
Observer A				
Irregularity, probably no lesion; levels 2 + 3 + 4 + 5	75	19	94	62
Indeterminate; levels 3 + 4 + 5	75	8	83	30
Lesion probably present; levels 4 + 5	73	5	78	18
Lesion definitely present; level 5	49	4	53	13
Observer B				
Irregularity, probably no lesion; levels 2 + 3 + 4 + 5	69	15	84	44
Indeterminate; levels 3 + 4 + 5	67	14	80	38
Lesion probably present; levels 4 + 5	60	7	67	21
Lesion definitely present; level 5	45	3	48	7
Observer C				
Irregularity, probably no lesion; levels 2 + 3 + 4 + 5	60	18	78	44
Indeterminate; levels 3 + 4 + 5	60	17	77	43
Lesion probably present; levels 4 + 5	53	15	68	36
Lesion definitely present; level 5	40	8	47	21
Observer D				
Irregularity, probably no lesion; levels 2 + 3 + 4 + 5	67	17	84	46
Indeterminate; levels 3 + 4 + 5	53	6	59	17
Lesion probably present; levels 4 + 5	36	1	37	4
Lesion definitely present; level 5	18	0	18	1

* Five-point ROC scale: level 1, no lesion; level 2, irregularity, probably no lesion; level 3, indeterminate for the presence of a lesion; level 4, lesion probably present; level 5, lesion definitely present.

† This number includes the truly localized malignancies plus the malignancies detected incidentally at CT performed for false-positive findings at chest radiography.

observer's confidence level and experience. Therefore, even the use of modern digital technology instead of analog screen-film technique does not make chest radiography as efficient as low-dose CT for lung cancer screening.

References

1. Flehinger BJ, Melamed MR. Current status of screening for lung cancer. Chest Surg Clin N Am 1994;4(1):1–15.
2. Hennigs SP, Garmer M, Jaeger HJ, et al. Digital chest radiography with a large-area flat-panel silicon X-ray detector: clinical comparison with conventional radiography. Eur Radiol 2001;11(9):1688–1696.
3. Fink C, Hallscheidt PJ, Noeldge G, et al. Clinical comparative study with a large-area amorphous silicon flat-panel detector: image quality and visibility of anatomic structures on chest radiography. AJR Am J Roentgenol 2002;178(2):481–486.
4. Garmer M, Hennigs SP, Jäger HJ, et al. Digital radiography versus conventional radiography in chest imaging: diagnostic performance of a large-area silicon flat-panel detector in a clinical CT-controlled study. AJR Am J Roentgenol 2000;174(1):75–80.
5. Redlich U, Hoeschen C, Effenberger O, et al. Comparison of four digital and one conventional radiographic image systems for the chest in a patient study with subsequent system optimization [in German]. Rofo 2005;177(2):272–278.
6. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. Lung Cancer 2006;54(2):177–184.
7. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). Int J Cancer 2007;120(4):868–874.
8. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2007;176(6):532–555.

9. Stahl M, Aach T, Dippel S. Digital radiography enhancement by nonlinear multiscale processing. Med Phys 2000;27(1):56–65.

10. Biesheuvel CJ, Vergouwe Y, Oudega R, Hoes AW, Grobbee DE, Moons KG. Advantages of the nested case-control design in diagnostic research. BMC Med Res Methodol 2008;8:48.

11. Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. Med Phys 1996;23(10):1709–1725.

12. Zheng B, Chakraborty DP, Rockette HE, Maitz GS, Gur D. A comparison of two data analyses from two observer performance studies using Jackknife ROC and JAFROC. Med Phys 2005;32(4):1031–1034.

13. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. Med Phys 2004;31(8):2313–2330.

14. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. N Engl J Med 2009;361(23):2221–2229.

15. National Lung Cancer Screening Trial. http://www.nci.nih.gov/nlst. 2008. Accessed July 2, 2009.

16. 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(2):347–352.

17. Sone S, Li F, Yang ZG, et al. Characteristics of small lung cancers invisible on conventional chest radiography and detected by population based screening using spiral CT. Br J Radiol 2000;73(866):137–145.

18. Quekel LG, Kessels AG, Goei R, van Engelshoven JM. Detection of lung cancer on the chest radiograph: a study on observer performance. Eur J Radiol 2001; 39(2):111–116.

19. Potchen EJ, Cooper TG, Sierra AE, et al. Measuring performance in chest radiography. Radiology 2000;217(2):456–459.

20. Gavelli G, Giampalma E. Sensitivity and specificity of chest X-ray screening for lung cancer: review article. Cancer 2000;89(11 suppl):2453–2456.

21. Li F, Arimura H, Suzuki K, et al. Computer-aided detection of peripheral lung cancers missed at CT: ROC analyses without and with localization. Radiology 2005;237(2): 684–690.

22. Toyoda Y, Nakayama T, Kusunoki Y, Iso H, Suzuki T. Sensitivity and specificity of lung cancer screening using chest low-dose computed tomography. Br J Cancer 2008;98(10):1602–1607.

23. 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.

24. Samei E, Flynn MJ, Eyler WR. Detection of subtle lung nodules: relative influence of quantum and anatomic noise on chest radiographs. Radiology 1999;213(3): 727–734.

25. Egglin TK, Feinstein AR. Context bias. A problem in diagnostic radiology. JAMA 1996;276(21):1752–1755.