



Published in final edited form as:

J Thorac Imaging. 2012 July ; 27(4): 249–254. doi:10.1097/RTI.0b013e318256951e.

READER VARIABILITY IN IDENTIFYING PULMONARY NODULES ON CHEST RADIOGRAPHS FROM THE NATIONAL LUNG SCREENING TRIAL

Satinder Singh, M.D.¹, David S Gierada, M.D.⁴, Paul Pinsky, M.D.³, Colleen Sanders, M.D.¹, Naomi Fineberg, Ph.D.², Yanhui Sun, Ph.D.², David Lynch, M.D.⁵, and Hrudaya Nath, M.D.¹

¹University of Alabama Hospitals in Birmingham, Department of Radiology-JT N370, 619 South 19th Street, Birmingham, AL 35233, Phone - (205) 934-5345, Fax - (205) 934-5688

²University of Alabama School of Public Health, Department of Biostatistics - 343B Ryals, 1665 University Boulevard, Birmingham, AL 35294, Phone – (205) 934-5971, Fax – (205) 975-2540

³EPN – Executive Plaza North, 305F, 6130 Executive Boulevard, Rockville, MD 20850 Phone – (301) 402-6480

⁴Mallinckrodt Institute of Radiology, Department of Radiology, 510 South Kingshighway Boulevard, St. Louis, MO 63110, Phone – (314) 362-2927, Fax – (314) 362-6218

⁵National Jewish Health Hospital, Division of Radiology, 1400 Jackson Street, Denver, CO 80206 Phone – (303) 270-2219

Abstract

Purpose—To measure reader variability related to the evaluation of screening chest radiographs (CXR) for findings of primary lung cancer.

Materials and Methods—From the National Lung Screening Trial (NLST), 100 cases were randomly selected from baseline CXR examinations for retrospective interpretation by nine NLST radiologists; images with non calcified lung nodules (NCN) or other abnormalities suspicious for lung cancer as determined by the original NLST reader were over-sampled. Agreement on the presence of pulmonary nodules and abnormalities suspicious for cancer and recommendations for follow-up were assessed by the multirater Kappa statistic.

Results—The multi-rater Kappa statistic for inter-reader agreement on the presence of at least one NCN was 0.38. Rates at which readers reported the presence of at least one NCN ranged from 32 to 63% (mean 41%); among 16 subjects with NCN and cancer diagnosis within one year of the CXR examination, a mean of 87% (range 81–94%) of cases were classified as suspicious for cancer across all readers. The multi-rater Kappa for agreement on follow-up recommendations was 0.34; pair wise Kappa values ranged from 0.15 to 0.64 (mean 0.36). For all subjects, readers recommended a follow-up procedure classified as high level (CT, FDG-PET, or biopsy) 42% of the time on average (range 30–67%); this increased to 84% (range 52–100%) when readers reported a NCN and 88% (range 82–94%) for subjects with cancer.

Conclusion—Reader agreement for screening CXR interpretation and follow-up recommendations is fair overall, but high for malignant lesions.

Keywords

Lung cancer; chest radiograph; reader variability; screening

Introduction

Many pulmonary nodules or findings suspicious for malignancy in asymptomatic individuals are identified incidentally through chest radiography. This may occur in various common clinical settings, including periodic health surveillance, preoperative assessment, and screening for metastatic disease. The reporting of findings suspicious for malignancy has several potential consequences. In the true positive setting, it may lead to diagnosis and treatment of an unsuspected malignancy, which may be beneficial if it prolongs the life of the patient. However, if the examination is falsely positive, potential detrimental consequences include the generation of unnecessary additional health care expenses and patient anxiety. With falsely negative examinations, the diagnosis of lung cancer may be delayed. Regardless of the eventual outcome, reporting findings suspicious for malignancy necessitates further diagnostic evaluation.

The detection of radiographic abnormalities, their interpretation as being suspicious for malignancy, and the radiologist's recommendations for follow-up are in part subjective, and thus are in part reader-dependent. Such reader variability will influence the rates of true positive detection of malignancy, the rates of unnecessary evaluations due to false positive interpretations, and false negative rates. Reader variability in the evaluation for malignancy also may influence the results of radiographic lung cancer screening studies, which have not been shown to reduce lung cancer mortality.^{1,2} To our knowledge, this variability among a large number of readers has not been systematically evaluated, particularly in the lung cancer screening setting. Accordingly, the purpose of this study was to measure variability among readers in the assessment of chest radiographs for lung nodules and other findings of malignancy.

Materials and Methods

Subjects and Case Selection

The subjects of this study were participants in the National Lung Screening Trial (NLST), which compared lung cancer mortality between more than 53,000 individuals randomized to undergo 3 screening tests at annual intervals by either postero-anterior chest radiography (CXR) or low radiation dose CT (LDCT).⁴⁻⁶ The NLST participants were between the ages of 55-74 years with a smoking history of 30 or more "pack-years". NLST enrollment occurred between September 2002 and April 2004. Informed consent was obtained, and HIPAA guidelines were followed.

The chest radiographs used in this reader study were retrospectively selected from the set of about 17,000 Year 0 (baseline) CXR screening examinations performed at the 10 centers in the Lung Screening Study (LSS) trial network of the NLST (see Appendix). A subject's baseline exam was eligible for possible inclusion if the subject either completed the Year 1 examination or was diagnosed with lung cancer before the Year 1 examination. The study design called for a stratified random sample composed of 100 distinct baseline screening radiographs. The strata were set up to achieve the same 37:63 ratio of soft copy to hard copy format of the chest radiographs actually reviewed in the entire LSS trial network of the NLST during the baseline screening year. For the selection of the test cases, all eligible year 0 images were divided into the 6 strata shown in Table 1, categorized by non-calcified nodule (NCN) status (according to the reading of the original radiologist who interpreted the study for the NLST), cancer status, and hard/soft copy format. Then the stated number of images for each stratum were randomly selected (using a random number generator) from all eligible images in each stratum.

The number of 100 total exams was chosen to represent a large enough sample to give reasonably precise estimates for the quantities of interest and while still allowing readers to complete the readings in a single reading day. The strata were designed to enrich the sample with 33 exams for which the original NLST radiologist had noted a NCN, to allow for comparisons between images with and without NCN, while maintaining NCN images as a minority of the total number as in the actual screening setting. Among these 33 exams, the sample was further enriched to obtain 16 subjects for whom a lung cancer diagnosis was obtained within a year of the exam, in order to have adequate numbers of NCN cases with and without cancer to statistically compare the two groups. Note the NCNs in the sampling design were based on the original NLST radiologist interpretation. Only one case with an original report of findings other than a NCN that led to a diagnosis of lung cancer (hilar and mediastinal lymphadenopathy) was obtained through the selection process.

CXR Technique

All radiographic equipment met ACR guidelines for chest radiography.³ Computed, digital, and film-screen radiography were permitted.⁶ Twenty-two of the hard copy images had been obtained as digital radiographs and originally interpreted on hard copy. Technical requirements included high kVp equipment (100–150 kVp) with a grid ratio of 10:1 or higher and up to 20 mAs; a value higher than 20 was permitted for very large patients.

Readers

Nine ABR-certified radiologists, each from a different LSS-NLST screening center and involved in NLST CXR interpretation, participated in the study (Appendix). The mean (SD) number of years of experience interpreting CXRs of these readers was 22.4 (6.7) with a range of 12–40. The mean (SD) proportion of the past calendar year's Relative Value Units (RVU) that had been generated by these readers from chest imaging (CXR, CT, and MR), used to describe the level of reader experience, was 0.47 (0.29), with a range of 0.25 to 0.98. Six of the nine radiologists were thoracic radiology sub-specialists in academic practice and the other three were members of large group outpatient and hospital based private practices. The readers knew that the images came from the NLST, but were not aware of the case selection criteria or the test set composition.

Image display and interpretation

All protected health information, including participant name, birth date, date of examination, identification numbers, and screening center identifiers contained in the images were either removed electronically (soft copy) or masked (hard copy). All readers interpreted the same 100 images in the same order during a single visit to a central location. Soft copy images were viewed on a clinical PACS work station (Centricity, GE Medical systems) using a clinical 2K × 2K gray scale monitor. The hard copy images were interpreted on a standard view box.

Readers were instructed to record all NCN regardless of size. Other abnormalities indirectly suggestive of the presence of lung cancer, such as hilar and/or mediastinal lymphadenopathy, and lobar atelectasis, were also recorded. The protocol also required a reader recommendation for diagnostic follow up of abnormal findings. Reader follow-up recommendation options included no diagnostic intervention; additional chest radiographs; LDCT or diagnostic chest CT; FDG-PET; biopsy; and/or other investigation (specified by the reader). In addition, each reader was asked to mark the location and approximate size of any radiographic abnormality on an 8" × 11" paper copy of a normal chest radiograph, and to label any calcified nodules (CN). Finally, the readers were asked to indicate their level of confidence that an abnormality was present that should be investigated further for lung

cancer. A suspicion score ranging from 1 (total confidence no abnormality present) to 10 (total confidence abnormality was present) was utilized for this purpose.

One author (HN- who was not a reader), an academic thoracic radiology sub-specialist with more than 30 years' experience, reviewed all the radiographs in which the readers identified one or more NCN, the paper copies of the mock CXR marking the abnormalities, and their interpretations; he recorded whether the outlined NCN was ill-defined or obscured by superimposed anatomic structures in order to gain an understanding of the reasons for any disagreement in interpretation and follow-up recommendations.

Data analysis

Multi-rater Kappa statistics were utilized to measure agreement among readers for reporting at least one NCN and for recommending different categories of diagnostic follow-up; pair-wise Kappa values also were analyzed. Kappa values for the level of agreement were interpreted in the following manner: 0–0.20 (slight agreement); 0.21–0.40 (fair agreement); 0.41–0.60 (moderate agreement); 0.61–0.80 (good agreement); 0.81–1.00 (almost perfect agreement).⁷ Recommendations for diagnostic follow-up were divided into three categories for the purposes of assessing reader agreement – no diagnostic intervention necessary, low level follow-up (additional radiographic views, chest fluoroscopy or chest radiography in 3 months) or high level follow-up (CT, LDCT or contrast enhanced diagnostic CT, FDG-PET, biopsy). CT and PET imaging were included in the high level follow-up category because these recommendations would be expected to signify a higher level of concern over an abnormal finding than a recommendation for additional radiography, and because in practice a CT scan would be obtained before proceeding to a biopsy in virtually all instances. Pairwise Spearman rank correlations of readers' suspicion scores were computed to assess the concordance among readers in these scores. Permutation tests were used to compute the statistical significance of differences in Kappa values, mean reader pair-wise correlations, mean suspicion scores and mean reader NCN reporting rates between various subsets of images.⁸

Results

Of the 100 subjects, 65% were between age 55 and 64 years; mean (SD) age was 62.3 (4.9) years. The mean (standard deviation) number of pack years smoked was 59 (23). From the original NLST radiologists' reports, the median (inter-quartile range) size of the 33 NCN, in terms of maximum dimension, was 15 mm (7–25mm); for one NCN, no dimensions had been recorded by the original NLST radiologist. Median (25th–75th) nodule size was 22.5 mm (15–39mm) for the 16 subjects with an NCN and cancer as compared to 7.0 mm (5.5–16mm) for the subjects with an NCN and no cancer ($p=0.004$; Wilcoxon rank-sum test).

Agreement among readers on screening results

With respect to inter-reader agreement on the presence of at least one NCN, the multi-rater Kappa statistic was 0.38 (95% CI 0.29–0.47). Pairwise Kappa statistics ranged from 0.13 to 0.60 (mean=0.38). The multi-rater Kappa was 0.43 for the 37 images read as soft copy and 0.34 for the 63 images read as hard copy ($p=0.31$). Across readers, the percentage of images reported as containing an NCN ranged from 32% to 63%, with a mean of 44% (Table 2); the number of reported NCNs per image ranged from 0.33 to 0.83 across readers, with a mean of 0.53. For the 16 images with an NCN linked to a subsequent diagnosis of cancer, this range was 81 to 94% (mean 87%). Additionally, for the 16 images with an NCN as reported by the original radiologist to be greater than or equal to the median size (15mm), the mean rate of reporting an NCN for the readers was 87.5%; this contrasts to a mean of 54.2% for the 16 images with an NCN less than the median size ($p=0.002$).

The pair-wise Spearman correlation coefficients for the readers' suspicion scores ranged from 0.39 to 0.81 (mean 0.59). The mean pair-wise correlation was higher for the 33 images in which a nodule had been reported by the original reader (mean 0.64, range 0.44–0.86), than for the other 67 images (mean 0.45, range 0.24–0.67) ($p = 0.033$). Readers' suspicion scores were associated with the number of readers reporting an NCN. The average suspicion score, among readers reporting an NCN, was 6.1 for those images in which six or fewer readers reported an NCN as compared to 8.5 for those images where 7–9 readers reported an NCN ($p < 0.0001$). Readers' average suspicion scores for the 17 images (16 with NCN, and 1 with hilar/mediastinal lymphadenopathy) with cancer ranged from 7.6 to 8.9 (mean 8.4); in contrast, readers' average scores for the 83 non-cancer images ranged from 1.9 to 5.5 (mean 3.4) ($p < 0.0001$).

Agreement among readers on follow-up recommendations

Readers often made multiple recommendations for follow-up; the highest level recommended was used for the analysis. The rate of recommending high level follow-up ranged across readers from 30 to 67% of cases, with a mean of 42%, and was higher when a NCN was reported (Table 3). For those images in which readers judged a NCN to be present, the rate of recommending high-level follow-up ranged from 52% to 100% (mean 84%), while among the 17 cancer images, the rate of recommending high-level follow-up ranged from 82 to 94% (mean 88%). The multi-rater Kappa for agreement on level of follow-up (none/low/high) was 0.34 (95% confidence interval 0.26, 0.42); pair wise Kappa values ranged from 0.15 to 0.64 (mean 0.36). For none/low versus high level of follow-up, agreement by the multi-rater Kappa was 0.41 (95% CI: 0.32–0.50).

Review of the images containing NCN

Based on the readers' interpretation and notation of NCN and CN on the paper copy, the readers identified 120 distinct NCN in 88 images. Among these 120 NCN, 43 (35.8%) were identified by only 1 out of 9 readers and had been measured by the test reader as less than 10 mm in size. Thirty-two of the 120 NCN (26.7%) were ill-defined and/or obscured by a rib, hilum, or mediastinal structures. Four nodules (3.3%) were variably interpreted as CN and NCN by different readers. In 3 instances (2.5%), a rib – either a callus or costal cartilage – was interpreted as an NCN by some readers. Nineteen of the 120 NCN (15.8%) were identified by all nine readers, and readers agreed about the number and location of all NCN in 12 cases (10%).

The review demonstrated that agreement improved with increasing size and conspicuity of the abnormality. It also became apparent that in some cases readers interpreted the same abnormality differently, particularly larger, poorly defined abnormalities, some interpreting them as NCN and others as consolidation, or as lymphadenopathy or pleural thickening when the opacity was adjacent to the hilum or chest wall, respectively (Figs 1,2). Such differences in interpretation of an abnormality (as opposed to its detection) resulted in differences in the level of suspicion for the presence of cancer, as well as the level of the follow-up recommendations.

Discussion

It is known that some degree of observer variability exists in the interpretation of chest radiographs.⁹ For example, differences in the performance of radiologists and other physicians in the ability to distinguish normal from abnormal chest radiographs have been quantified previously by ROC analysis.⁹ In addition, the issues of radiographically missed lung cancer and the possible causes have been well described.^{10–12} However, to our knowledge, there has been limited specific assessment of interobserver agreement by Kappa

analysis for potentially malignant findings on chest radiographs. In this study, which compared the performance of a large number of radiologists in the interpretation of screening chest radiographs obtained among participants randomized to the CXR arm of the NLST, mean pairwise Kappa values for pulmonary nodule identification corresponded to fair interobserver agreement.

A previous study¹³ found a higher Kappa value of 0.49 for nodule detection on posteroanterior radiographs, which is considered moderate agreement. However, agreement in that comparison study was assessed between only two readers, both from the same institution. In contrast, our result is based on a wider cross-section of radiologists, each from a different institution. This revealed a fairly wide range of pairwise Kappa values, likely reflecting inherent differences and similarities in reader sensitivity and specificity. We are unaware of any other comparison studies on reader variability in the radiographic assessment for pulmonary nodules.

There was moderate correlation between the suspicion scores of the various reader pairs. As would be expected, the average suspicion scores were higher when more readers judged a nodule to be present. Suspicion was about the same for the subset of cancer cases as for all cases in which most readers reported a nodule as present. This may be related to the limited ability to characterize nodules by CXR, so that all findings suggestive of a nodule must be considered suspicious. Since the test group consisted of NLST participants at increased risk of lung cancer, and the readers were aware of this characteristic of the study population, the suspicion score in this study may be higher than it would have been for patients with lower risk profiles.

While overall agreement was only fair, the rates of reported nodules for the cancer cases were high for all readers, with a narrow range. This suggests that reader variability for pulmonary nodule identification is low when it is most important for it to be low, i.e. when malignancy is truly present. For the non-cancer cases and cases in which fewer readers reported a nodule, observer variability was higher. However, the clinical consequences of reader variability are expected to be less in these situations, the likely result being greater variability in the rates of further assessment of findings that ultimately have no clinical importance. We also acknowledge that some cancers may not be detectable on CXR. Therefore, a nodule reporting rate of 100% and perfect inter observer agreement would not be expected, even for cases with proven cancer.

Follow-up recommendations may potentially be as important as detecting an abnormality, as they convey a measure of the reader's suspicion regarding the malignant potential for a nodule. Agreement for nodule detection follow-up recommendations as measured by Kappa was only fair, probably because agreement for nodule detection was also only fair. However, the rates for recommending high-level follow-up corresponded appropriately to the number of nodules reported, and to the level of suspicion for all nodule cases and for the subset of cancer cases.

Subsequent review of the chest radiographs for which a NCN was identified sheds some light on the relationship between the readers' interpretation of a radiographic abnormality, their follow-up recommendation, and interobserver agreement. Agreement that an image contained a NCN tended to be low when the abnormality was ill-defined, or obscured by overlying anatomic structures. Agreement increased for larger NCNs, and was higher for those larger than 15 mm compared to those smaller than 15 mm. In addition, findings reported as NCNs by some readers were considered benign CNs or costal calcification by others, which in turn influenced follow-up recommendations and agreement. The difficulties in distinguishing CN and NCN on CXR have been previously described.¹⁴ These types of

variability in reader behavior have also been found for other screening methods such as mammography,¹⁵ and for lung cancer screening with low dose CT.¹⁶ Our observations suggest the need for certain ancillary radiographic technologies, such as dual-energy subtraction imaging¹⁷ or post-processing subtraction of osseous structures¹⁸ to reduce observer variability by making it easier to discriminate true nodules from summation of normal structures, or CNs from NCNs. Nodule detection software also has the potential to help improve interobserver agreement,¹⁹ although the propensity for false-positive nodule identification could be an additional source of disagreement.

This study has several limitations. Any reader variability study is an artificial exercise, and radiologists' behavior may be different in an actual clinical reading environment. This may have been a factor in our study, since the overall rate of reported nodules was a little higher for the study readings than for the original radiologist readings used to select the test set. Similarly, the original radiologists' readings on which the case selection was based were made within the context of a lung cancer screening trial, and may have been different in a routine clinical setting. In addition, different readers with varying degrees of experience interpreting chest imaging studies, or different image sets may produce different results. Due to the source of the radiographs, there were no corresponding CT scans or other reference standard to establish the true presence or absence of lung nodules in the test set. This precluded any assessment of reader accuracy, but does not affect assessment of reader variability.

In summary, interpretations of chest radiographs from a large number of radiologists in a lung cancer screening study demonstrated fair agreement for detecting NCNs and for recommending high-level follow up. However, agreement was high for detection and suspicion level for nodules proven to be malignant.

Acknowledgments

Grant Support: National Cancer Institute; National Lung Screening Trail (NLST) (<http://www.cancer.gov/nlst>, clinicaltrials.gov identifier NCT00047385)

This research was supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. The authors thank Drs. Christine Berg, Richard Fagerstrom, and Pamela Marcus, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff of the National Lung Screening Trial (NLST), Mr. Tom Riley and staff, Information Management Services, Inc., Ms. Brenda Brewer and staff, Westat. Online staff listing at: http://www.nejm.org/doi/suppl/10.1056/NEJMoa1102873/suppl_file/nejmoa1102873_appendix.pdf Most importantly, we acknowledge the study participants, whose contributions made this study possible

References

1. Oken MM, Hocking WG, Kvale PA, et al. Screening by chest radiograph and lung cancer mortality: the Prostate, Lung, Colorectal, and Ovarian (PLCO) randomized trial. *Jama*. 2011; 306:1865–73. [PubMed: 22031728]
2. Humphrey LL, Teutsch S, Johnson M. Lung cancer screening with sputum cytologic examination, chest radiography, and computed tomography: an update for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2004; 140:740–53. [PubMed: 15126259]
3. ACR-SPR. Practice Guideline for the Performance of Pediatric and Adult Chest Radiography. 2011.
4. Aberle DR, Berg CD, Black WC, et al. The National Lung Screening Trial: overview and study design. *Radiology*. 2011; 258:243–53. [PubMed: 21045183]
5. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365:395–409. [PubMed: 21714641]
6. Aberle DR, Adams AM, Berg CD, et al. Baseline characteristics of participants in the randomized national lung screening trial. *J Natl Cancer Inst*. 2010; 102:1771–9. [PubMed: 21119104]

7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–74. [PubMed: 843571]
8. Edgington, ES. *Randomization Tests*. 3. New York: Marcel-Dekker; 1995.
9. Potchen EJ, Cooper TG, Sierra AE, et al. Measuring performance in chest radiography. *Radiology*. 2000; 217:456–9. [PubMed: 11058645]
10. Shah PK, Austin JH, White CS, et al. Missed non-small cell lung cancer: radiographic findings of potentially resectable lesions evident only in retrospect. *Radiology*. 2003; 226:235–41. [PubMed: 12511696]
11. Quekel LG, Kessels AG, Goei R, van Engelshoven JM. Miss rate of lung cancer on the chest radiograph in clinical practice. *Chest*. 1999; 115:720–4. [PubMed: 10084482]
12. Muhm JR, Miller WE, Fontana RS, Sanderson DR, Uhlenhopp MA. Lung cancer detected during a screening program using four-month chest radiographs. *Radiology*. 1983; 148:609–15. [PubMed: 6308709]
13. Karacin, O.; Ibis, A.; Akcay, S.; Akkoca, O.; Eyuboglu, FO.; Coskun, M. Chest radiography and the solitary pulmonary nodule: interobserver variability, and reliability for detecting nodules and calcification. *Journal of Radiology*. 2002 Mar. www.jradiology.org
14. Berger P, Perot V, Desbarats P, Tunon-de-Lara JM, Marthan R, Laurent F. Airway wall thickness in cigarette smokers: quantitative thin-section CT assessment. *Radiology*. 2005; 235:1055–64. [PubMed: 15833982]
15. Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. *J Natl Cancer Inst*. 1998; 90:1801–9. [PubMed: 9839520]
16. Gierada DS, Pilgram TK, Ford M, et al. Lung cancer: interobserver agreement on interpretation of pulmonary findings at low-dose CT screening. *Radiology*. 2008; 246:265–72. [PubMed: 18024436]
17. Kuhlman JE, Collins J, Brooks GN, Yandow DR, Broderick LS. Dual-energy subtraction chest radiography: what to look for beyond calcified nodules. *Radiographics*. 2006; 26:79–92. [PubMed: 16418245]
18. Freedman MT, Lo SC, Seibel JC, Bromley CM. Lung nodules: improved detection with software that suppresses the rib and clavicle on chest radiographs. *Radiology*. 2011; 260:265–73. [PubMed: 21493789]
19. White CS, Flukinger T, Jeudy J, Chen JJ. Use of a computer-aided detection system to detect missed lung cancer at chest radiography. *Radiology*. 2009; 252:273–81. [PubMed: 19561261]

APPENDIX

The 10 screening centers of the Lung Screening Study Component of the National Lung Screening Trial were: University of Alabama at Birmingham, University of Colorado, Georgetown University, Henry Ford Hospital, Marshfield Clinic, University of Minnesota, Pacific Health Research Institute, University of Pittsburgh, University of Utah, and Washington University.

Readers	
Peter Balkin, M.D.	Pacific Health Research Institute, Honolulu, HI
Matthew Freedman, M.D.	Georgetown University Medical Center, Washington, DC
David Gierada, MD	Mallinckrodt Institute of Radiology, Washington University, St. Louis, MO
David Lynch, MD	National Jewish Hospital, University of Colorado, Denver, CO
Ian Malcolm, M.D.	University of Alabama at Birmingham, Birmingham, AL
Howard Mann, M.D.	University of Utah, Salt Lake City, UT
William Manor, M.D.	Marshfield Medical Research & Education Foundation, Marshfield, WI

Readers

Karla Myhra-Bloom, M.D.	University of Minnesota, Minneapolis, MN
Carl Zylak, M.D.	Henry Ford Health System, Detroit, MI

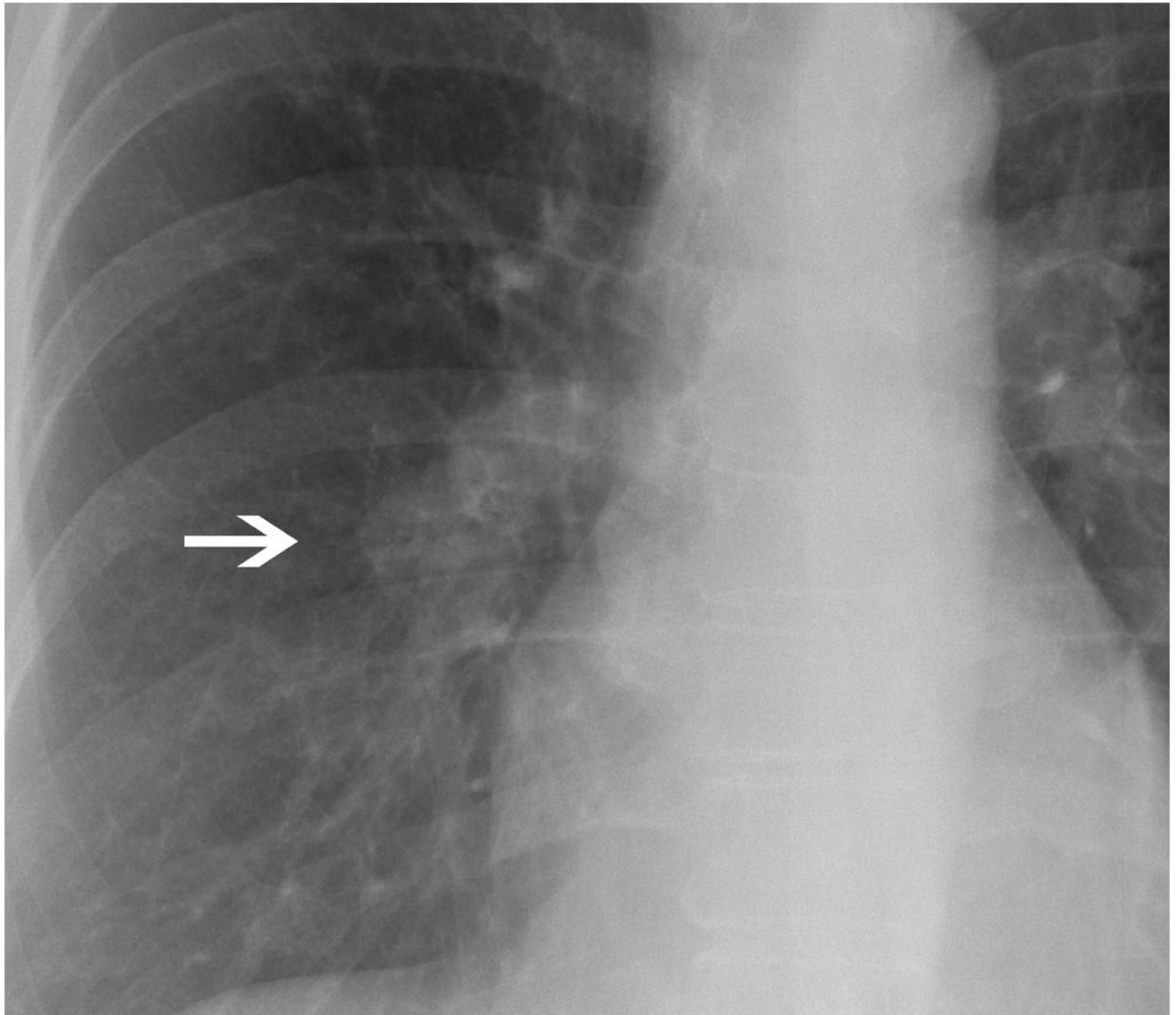


Fig. 1. There is a poorly defined opacity overlying the inferior aspect of the right hilum (arrow). Some readers considered this abnormality a non calcified nodule while others hilar adenopathy. However, all the readers considered this radiograph highly suspicious for cancer.

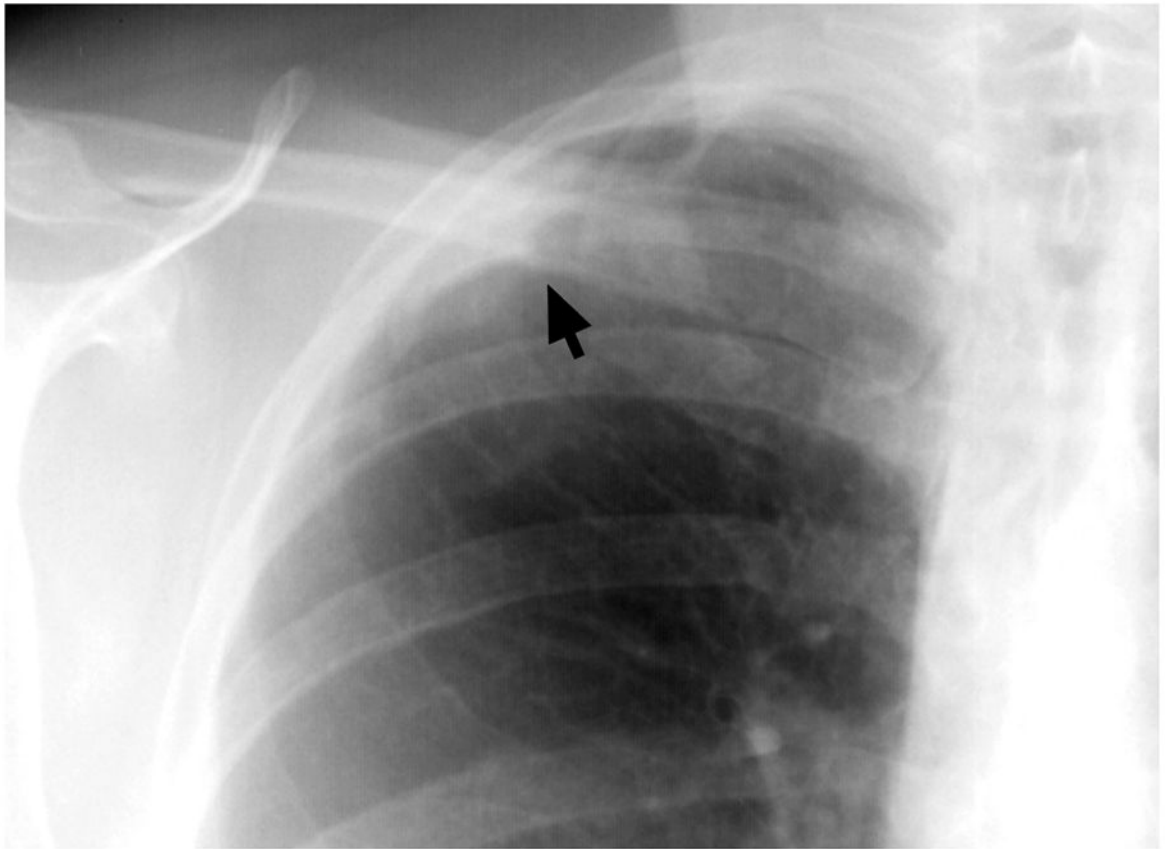


Fig 2. There is an ill-defined opacity in the right apex adjacent to the lateral pleural surface (arrow). All the readers recognized the abnormality, but some readers considered this as a focal opacity (NCN), others as apical pleural thickening. Their level of suspicion and follow up recommendation correlated with their interpretation of the abnormality.

Table 1

Sampling design for exams selected for reader study

Findings on Image ¹	Image Format		Total
	Read as Soft Copy	Read as Hard Copy	
NCN without lung cancer ²	7	10	17
NCN w/lung cancer ²	6	10	16
No NCN	24	43	67
Total	37	63	100

¹ Findings (NCN or not) based on the original NLST radiologist's read

² Lung cancer within year of exam

Table 2

Reported non-calcified nodules (NCNs) by readers

	Number (% of images for which reader reported at least one NCN)	Mean number of NCN reported per image	Number (% of images with cancer present for which reader reported an NCN)
Reader #	N=100	N=100	N=16 ^I
1	46 (46)	0.54	13 (81)
2	63 (63)	0.83	14 (88)
3	43 (43)	0.59	13 (81)
4	33 (33)	0.46	14 (88)
5	44 (44)	0.59	15 (94)
6	32 (32)	0.33	13 (81)
7	34 (34)	0.40	14 (88)
8	44 (44)	0.61	15 (94)
9	34 (34)	0.38	14 (88)
Mean (SD)	41 (9.8)	0.53 (0.15)	13.9 (87)

^IWith cancer present and a NCN according to original reader Note: SD = Standard Deviation; NCN = non-calcified nodule

Table 3

Follow-up recommendations by readers

Reader #	All Exams (N=100)		Exams where NCN judged by reader to be present		Exams where cancer was present (N=17)	
	No diagnostic intervention	Low Level F/U	High Level F/U	High Level F/U	High Level F/U	High Level F/U
1	29 (29)	40 (40)	31 (31)	24 (52)	14 (82)	
2	26 (26)	7 (7)	67 (67)	58 (92)	15 (88)	
3	50 (50)	0 (0)	50 (50)	43 (100)	15 (88)	
4	64 (64)	6 (6)	30 (30)	28 (85)	15 (88)	
5	57 (57)	10 (10)	33 (33)	31 (70)	14 (82)	
6	53 (53)	9 (9)	38 (38)	27 (84)	15 (88)	
7	56 (56)	1 (1)	43 (43)	33 (97)	15 (88)	
8	47 (47)	9 (9)	44 (44)	37 (84)	16 (94)	
9	56 (56)	1 (1)	43 (43)	31 (91)	15 (88)	
Mean [SD]	(49) [12.9]	(9) [12.1]	(42) [11.5]	(84) [14.8]	(88) [3.5]	

Note: Data are number (%) of readings. NCN = non calcified nodule, SD = Standard Deviation