

Published in final edited form as:

Invest Radiol. 2012 August ; 47(8): 457–461. doi:10.1097/RLI.0b013e318250a5aa.

Computer-aided nodule detection and volumetry to reduce variability between radiologists in the interpretation of lung nodules at low-dose screening CT

Kyung Nyeo Jeon^{1,6}, Jin Mo Goo^{2,3}, Chang Hyun Lee², Youkyung Lee², Ji Yung Choo², Nyoung Keun Lee², Mi-Suk Shim², In Sun Lee², Kwang Gi Kim⁴, David S. Gierada⁵, and Kyongtae T. Bae⁶

¹Department of Diagnostic Radiology, College of Medicine, Gyeongsang National University, Jinju, Korea

²Department of Radiology, Seoul National University College of Medicine, and Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Korea

³Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea

⁴Biomedical Engineering Branch, Division of Convergence Technology, National Cancer Center, Gyeonggi-Do, Korea

⁵Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, MO

⁶Department of Radiology, University of Pittsburgh School of Medicine, Pittsburgh, PA

Abstract

Objective—To evaluate whether a computer-aided diagnosis (CAD) system improves interobserver agreement in the interpretation of lung nodules at low-dose CT screening for lung cancer.

Materials and Methods—Baseline low-dose screening CT examinations from 134 participants enrolled in the National Lung Screening Trial were reviewed by seven chest radiologists. All participants consented to the use of their de-identified images for research purposes. Screening results were classified as positive when noncalcified nodules larger than 4 mm in diameter were present. Follow-up evaluation was recommended according to the nodule diameter: 4 mm; >4–8 mm; > 8 mm. When multiple nodules were present, recommendations were based on the largest nodule. Readers initially assessed the nodule presence visually and measured the average nodule diameter manually. Revision of their decisions after reviewing the CAD marks and size measurement was allowed. Interobserver agreement evaluated using multirater κ statistics was compared between initial assessment and that with CAD.

Results—Multirater κ values for the positivity of the screening results and follow-up recommendations were improved from moderate ($\kappa=0.53$; 0.54) at initial assessment to good ($\kappa=0.66$; 0.67) after reviewing CAD results. The average percentage of agreement between reader pairs on the positivity of screening results and follow-up recommendations per case was also increased from 77% and 72% at initial assessment to 84% and 80% with CAD.

Correspondence to: Jin Mo Goo, M.D., Department of Radiology, Seoul National University Hospital, 101 Daehangno, Jongno-gu, Seoul 110-744, Korea, Tel: 82-2-2072-2624, Fax: 82-2-743-7418, jmgoo@plaza.snu.ac.kr.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusion—CAD may improve the reader agreement on the positivity of screening results and follow-up recommendations in the assessment of low-dose screening CT.

Keywords

computer-aided diagnosis (CAD); lung nodule; computed tomography; screening

Introduction

The benefits of early detection of lung cancer to improve clinical outcomes have been suggested by previous studies.^{1, 2} Recently, the results of the National Lung Screening Trial (NLST) showed a 20% reduction in mortality due to lung cancer among those who underwent low-dose CT screening as compared to the control group screened by chest radiography.³ Although this is exciting and highly promising news for improving outcomes of lung cancer diagnosis and treatment, there are a number of practical issues to be addressed in the implementation of low-dose CT screening for lung cancer.

Although a high false positive screening rate of 96.4% with low-dose CT was reported for the NLST,³ small nodules should not be missed to achieve a high cure rate. Several studies, however, showed that newly detected nodules on repeat CT could be demonstrated in retrospect on the prior CT,^{4, 5} and 39% of these nodules were 4 mm or larger.⁵

Another issue is that in the NLST no consensus or standardized guideline was established on how to clinically manage the detected nodules. There is considerable variability among radiologists in the interpretation of screening CT, for both baseline and follow-up scans.^{6, 7} In order to appropriately follow and manage CT findings and analyze outcomes, it is crucial to standardize radiologists' interpretations to reduce variability and errors.

Multiple studies have shown that computer-aided diagnosis (CAD: both detection and volumetry) systems are beneficial as a second reader to improve the detection of small pulmonary nodules.^{8–10} CAD systems can also provide more precise volumetric measurement of nodules, which improves the characterization of detected lung nodules and detection of interval changes in follow-up scans.¹¹ Although intermeasurement agreements have been reported in several studies,^{12, 13} the effect of CAD on the positivity or follow-up recommendations on screening CT has never been evaluated. We postulated that the use of CAD may help reduce variability and improve agreement among radiologists in the interpretation of low-dose screening CT. Thus, the purpose of this study was to evaluate whether a CAD system improves interobserver agreement in the classification and recommended follow-up of low-dose CT screening examinations for lung cancer.

Materials and Methods

Subjects and Case Selection

All subjects were enrolled in the NLST (<http://www.cancer.gov/nlst>, clinicaltrials.gov identifier NCT00047385), a multicenter research protocol approved by the institutional review boards of all participating centers, and each participant consented to the use of their de-identified images for research purposes at the time of enrollment. The overview and study design of the NLST has been previously described.¹⁴

A total of 134 screening CT examinations in 134 NLST participants who each underwent one baseline screening examination were used in this study. This study population is identical to that of a study by Gierada et al;⁶ there is one fewer case in the current study because one of the CT examinations provided two different lesions for analysis in the

previous study. Note that all CT images from each CT examination were evaluated in our study, while subsets of only 12 to 40 CT images were presented to readers in the previous study by Gierada et al.⁶ Among 134 cases, 74 cases were classified as having at least one noncalcified nodule 4 mm or larger in greatest transverse diameter (52 cases with single nodule, 12 cases with two nodules, 7 cases with three nodules, 2 cases with four nodules, 1 case with five nodules); 60 cases were classified as having a negative result (20 cases with less than 4 mm nodules, 20 cases with calcified nodules, and 20 cases with no abnormality or other abnormalities like emphysema). All protected health information and screening site identifiers were removed from the images.

All low dose CT scans were acquired with the use of multidetector CT scanners with a minimum of four channels. No intravenous contrast material was administered. Acquisition parameters included 120–140 kVp and 20–60 mAs (effective). Reconstruction section thickness was 2 mm in 51 examinations, 2.5 mm in 82, and 1.25 mm in 1. Reconstructed sections were contiguous or overlapping in the transverse plane.

Readers and Image Viewing

Seven radiologists (J.M.G., C.H.L., Y.L., J.Y.C., N.K.L., M.S., I.S.L.) with average experience in interpreting CT scans of 10 years \pm 6 (range, 5–20 years) participated as readers. None of the readers had participated in the study by Gierada et al.⁶

The image sets in Digital Imaging and Communication in Medicine format were loaded into the workstation (Z600; HP, Palo Alto, CA) with two flat-panel liquid crystal display color monitors (UltraSharp 2007FP; Dell, Round Rock, TX). Images were displayed using a viewer system (Xelis; Infinitt Healthcare Co., Ltd., Seoul, Korea) which provides tools for CAD for lung nodules and nodule volumetry in addition to conventional tools for image magnification, electronic caliper measurement, and image scrolling. The images were examined with lung (window width, 1500H; window level, –700H) and soft tissue windows (window width 400H; window level, 20H), but readers were then free to adjust settings to their preferences.

Cases were presented in a predetermined randomized order, which was the same for each reader. Each reader was provided with instructions for the use of the workstation, and a demonstration of the system and tools was performed before the reading session. No limit was imposed on the interpretation time. Clinical, demographic, or database information was not provided to readers.

Case Interpretation and Recommendations for Follow-up

In the NLST protocol, a positive screening result was defined as the presence of at least one noncalcified nodule (NCN) \geq 4 mm in greatest transverse diameter. For the follow-up evaluation of patients with a positive screening examination, no specific algorithms were prescribed and the readers' recommendations for follow-up were based on current standards of clinical practice.

In our study, however, the definition of a positive screening examination and the follow-up recommendations were modified by adopting the guidelines of the Fleischner Society,¹⁵ because of the following considerations. First, all of the radiologists participating in the study were familiar with the Fleischner Society's guidelines and used them in their daily clinical practice. Therefore, no new study guidelines or additional training was required to instruct the radiologists for the study. Second, the diameter of a nodule calculated by the CAD system represents the effective diameter of a nodule (i.e., the diameter computed from a sphere whose volume is the same as the nodule volume measured by the CAD). This

diameter more closely approximates the average of length and width used by the Fleischner Society's guidelines than the greatest transverse diameter used in the NLST protocol.

When NCNs greater than 4 mm in effective diameter were detected in any CT images, a screening study was classified as a positive. Conversely, when a NCN was smaller than or equal to 4 mm in the effective diameter, or when nodules were calcified or absent, it was classified as a negative screening CT. Readers were asked to provide recommendations for follow-up mainly based on the effective nodule diameter: (1) negative screening CT, continuation or routine screening without any intervention (no follow-up); (2) $4\text{mm} < \text{NCN} < 8\text{mm}$, repeat low-dose CT at a specified interval before the next annual screening examination (low-level follow-up); (3) $\text{NCN} > 8\text{mm}$, a diagnostic imaging study and/or lung biopsy (high-level follow-up). However, the recommendations could be modified according to the reader's discretion when a nodule was characterized as subsolid. Readers were also allowed to modify their decisions according to nodule characteristics such as nodule margins. When multiple nodules were present, screening positivity and recommendations were based on the largest nodule.

Each reader independently reviewed and interpreted the CT images twice. Initially, readers assessed the presence of nodules visually, and recorded bi-dimensional manual measurements of the nodule to the nearest 0.1 millimeter using an electronic caliper. After the initial assessment, readers clicked a button on the viewer program to see the results of CAD marks and nodule volumetry with a color overlay over the segmented nodule. Revision of their decisions after reviewing the CAD marks and measurements was allowed. Readers were instructed to reject false positive marks and unsatisfactory segmentation results of the nodule at their discretion. When a decision was altered after reviewing CAD results regarding the screening positivity or recommendations for follow-up, readers were asked to record the reason.

Data Analysis

The average percentage of agreement between reader pairs on the positivity of screening results and follow-up recommendations per case was calculated. The number of cases with complete agreement across all readers was determined. Interobserver agreement for the positivity of the screening test and recommendations for follow-up was assessed by using multirater κ , which measures the level of agreement after taking chance agreement into account.¹⁶ A κ value of less than 0.20 indicated poor agreement; a κ value of 0.21–0.40, fair agreement; a κ value of 0.41–0.60, moderate agreement; a κ value of 0.61–0.80, good agreement; and a κ value of more than 0.81, excellent agreement. Interobserver agreement between reader pairs (for the positivity of the screening test and recommendations for follow-up) was also calculated. All analyses were compared between the initial assessment and the assessment augmented by CAD. Statistical analysis was performed using SAS software (version 9.2, SAS institute, Cary, NC).

Results

Positive versus Negative Screening Result

The number of cases per reader classified as positive screening examinations ranged from 73 to 98 cases (mean, 81) at the initial assessment and from 65 to 91 cases (mean, 80) after the review of CAD results. The number of cases that showed a difference regarding screen positivity between reader pairs ranged from 24 to 39 cases (mean, 30) at the initial assessment and from 13 to 30 cases (mean, 22) with CAD. The mean percentage of agreement between reader pairs on the positivity of screening results increased from 77% at initial assessment to 84% with CAD. The number of cases for which there was complete

agreement over all readers also increased from 61 cases (46%) to 80 cases (60%) (Fig. 1). The multirater κ value increased from 0.53 (95% confidence interval: 0.49, 0.56) to 0.66 (95% confidence interval: 0.62, 0.70) (Table 1). The κ values among all reader pairs ranged from 0.37 to 0.61 at the initial assessment and ranged from 0.56 to 0.79 with CAD (Table 2).

Follow-up Recommendations

The number of cases for which there was a difference regarding the follow-up recommendations between reader pairs ranged from 31 to 45 cases (mean, 37) at the initial assessment and from 15 to 36 cases (mean, 20) with CAD. The mean percentage of agreement between reader pairs on follow-up recommendations per case was increased from 72% at initial assessment to 80% with CAD.

The number of cases for which there was complete agreement over all readers regarding the follow-up recommendations also increased from 50 cases (37%) to 72 cases (54%) (Fig. 1). Multirater κ value increased from 0.54 (95% confidence interval: 0.51, 0.57) to 0.67 (95% confidence interval: 0.64, 0.70) (Table 1). The κ values among all reader pairs ranged from 0.49 to 0.59 at the initial assessment and ranged from 0.60 to 0.87 with CAD (Table 3).

Causes for the Decision Change

The readers responded that they changed their decisions after review of CAD results when new noncalcified nodules larger than 4 mm were detected by CAD in 0 to 13 cases (mean, 5) and when the CAD measurement of nodule size was reviewed in 5 to 28 cases (mean, 14) (Table 4).

Discussion

The role of CT screening to detect early lung cancer is one of the most important yet contentious issues in dealing with lung cancer. Recent results of NLST break new ground in this area, and efforts should be made towards developing better protocols for screening programs.

Small pulmonary nodules are a routine finding on low-dose screening CT scans, and present a management dilemma. Because the NLST provided no mandated diagnostic evaluation algorithm, the decision as to follow-up was left to the individual radiologists. The Fleischner guidelines were designed to standardize the management of small nodules incidentally detected at CT and to reduce the overuse of short intervals.¹⁵ In a survey of 834 radiologists, 77.8% of responders were aware of the Fleischner Society guidelines and 58.8% worked in practices that employed them or similar guidelines.¹⁷ Therefore, we employed the modified Fleischner guidelines in the follow-up recommendations in this study, even though this was proposed for the management of small nodules at nonscreening CT.

With the expansion of low-dose CT screening, radiologists may have to analyze a large number of CT images in a limited time. Radiologists' interpretation efficiency and the consistency of results would be crucial to maximize the clinical benefits of low-dose CT screening for lung cancer. Double reading (independent readings by two radiologists) would improve the detection rate of pulmonary lesions. However, interobserver variability in the interpretation (lesion detection, characterization, and measurement) of screening CT scans remains a serious concern.^{6, 7, 18} Previous studies demonstrated the difficulty of detecting small pulmonary nodules and a large interobserver variability in the detection of pulmonary nodules among readers.¹⁸⁻²⁰ Leader et al¹⁸ analyzed the reader variability per nodule among three reviewers and reported that interobserver agreement for the detection of individual nodules improved with increases in nodule size and that the fraction of missed

nodules was higher in smaller nodules. A large interobserver variability has been also reported on lesion characterization including the determination of changes in nodule attenuation and margins as well as the estimation of likelihood of malignancy.^{7, 21}

Studies have reported large interobserver variation in the manual measurement of nodule size in two dimensions with electronic calipers.^{21, 22} The small size, indistinct margins, or irregular shape of certain nodules may affect placement of the electronic cursors. Since nodule size is used as a criteria both for the determination of screen positivity and for follow-up guidelines in lung cancer screening CT, relatively small difference in size measurement may critically affect clinical management decisions on nodules. A study by Gierada et al,⁶ using case-based evaluation of interobserver variability in lung cancer screening CT, revealed that much of the disagreement occurred because some readers classified cases with nodules near the threshold (4-mm) as positive, and others classified them as negative.

CAD is likely to play a promising role in the improvement of nodule detection and reduction in interobserver variability in the interpretation of low dose CT for lung cancer screening. The NELSON trial (Dutch-Belgian lung cancer screening trial) is the first large lung cancer screening trial in which automated, volumetric nodule assessment is routinely applied and forms part of the nodule management protocol.²³ In this trial, to limit the number of follow-up CT examinations and the overuse of invasive procedures, recommendations for follow-up or management of the nodules were provided based on volumetric nodule assessment or the percent volume change and estimated volume doubling time in consecutive scans. However, the effect of CAD on interobserver variability in dealing with CT screen positivity and follow-up recommendations has not been studied.

Multiple studies have shown that CAD is useful in detecting lung cancers that had been missed at CT.^{24, 25} It has been shown that CAD offers a useful second opinion in detection of small lung nodules.^{8, 26-28} Lung nodule volumetric measurement with CAD facilitates a reduction in interobserver variability in the evaluation of indeterminate nodules in low-dose CT.^{12, 13} In the present study, readers changed their decisions on average 19 times in assessing 134 cases (14%) after reviewing CAD results. Decisions were altered more frequently because of CAD nodule size measurement than the detection of new nodules. Interobserver agreement on the positivity of screening results and follow-up recommendations improved in all reader pairs with CAD.

The study by Gierada et al,⁶ which used CT examinations of the same NLST participants as in our study, investigated interobserver agreement among 16 radiologists who regularly interpreted NLST screening CT images. In their study, the level of agreement for the positivity of screening results, expressed as multirater κ value, was higher than that in our initial assessment without CAD; 0.64 vs. 0.53. We postulate that this difference could be explained by the fact that our study radiologists reviewed and assessed the entire CT examination in each case, whereas only a selected subset of CT images were presented to the readers in the study by Gierada et al. In reviewing the entire set of CT images, each reader had to deal with many more nodule candidates, consequently increasing the chance of disagreement. We believe our study set-up more closely simulates the clinical setting in interpreting low dose CT for lung cancer screening than review of selected subsets of CT images. Gierada et al⁶ showed a lower level of agreement for follow-up recommendations (multirater κ of 0.35) than that in our study either before (0.54) or after CAD (0.67). This difference may reflect the fact that the follow-up recommendations in our study were based on the guidelines of the Fleischner Society, whereas in their study, the follow-up recommendation was left to the reader's discretion and clinical judgment.

Of the 134 CT examinations used in our study, 74 had been classified as having positive screening results by the radiologists who originally interpreted the CT examinations in the NLST. In contrast, an average of 81 cases (range, 73–98) at the initial assessment and 80 cases (range, 65–91) with CAD were interpreted as positive examinations in our study. However, the positivity rates of our study cannot be directly compared to the NLST original reader rates, due to differences in methodology. Rather than unidimensional longest diameter, we used the average diameter of length and width of a nodule because that is the basis for the Fleischner Society follow-up guidelines, and it is more comparable to the equivalent diameter of a sphere with the same volume. In addition, a study by Gierada et al which used the same dataset also showed that the number of noncalcified nodules ≥ 4 mm identified by readers was $93 \text{ nodules} \pm 22$.⁶ This indicates 19 more cases were categorized as positive on average compared to the original reading.

There are several limitations in our study. First, because the data analysis was based on the interpretation of each case, it is plausible that two readers may have identified different nodules but conveyed the same opinions, thereby resulting in the same outcome. However, this also may occur in actual clinical practice. Second, section thickness is an important parameter in nodule volume measurement, and smaller nodules tend to have higher measurement errors related to section thickness and location than do larger nodules.^{29, 30} Because a substantial number of the nodules in this study were small, segmentation in CAD was likely affected. Third, though CT reconstruction kernel is reported to possibly influence the performance of a CAD system,³¹ the effect of the image reconstruction kernel on the performance of CAD system did not addressed in this study because CT examinations used in our study were performed in 10 different LSS (Lung Screening Study)-NLST centers using various MDCT scanners. Forth, no specific guidelines were provided for the follow-up recommendations of subsolid nodules because of lack of such recommendations in the current Fleischner guidelines. By employing recently proposed guidelines,^{32, 33} there may be room to reduce interobserver variability further.

In summary, we demonstrated the potential usefulness of CAD for reducing interobserver variability in detecting nodules and making follow-up recommendations during low-dose CT for lung cancer screening. Further work is necessary to determine the practical value of CAD in the clinical setting, and successful integration of lung nodule CAD into the clinical workflow is required.

Acknowledgments

This work was supported by Industrial Strategic technology development program, 10038419, Intelligent image diagnosis and therapy-support system funded by the Ministry of Knowledge Economy (MKE, Korea).

REFERENCES

1. Flehinger BJ, Kimmel M, Melamed MR. The effect of surgical treatment on survival from early lung cancer. Implications for screening. *Chest*. 1992; 101:1013–1018. [PubMed: 1313349]
2. Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med*. 2006; 355:1763–1771. [PubMed: 17065637]
3. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N Engl J Med*. 2011
4. Henschke CI, Naidich DP, Yankelevitz DF, et al. Early lung cancer action project: initial findings on repeat screenings. *Cancer*. 2001; 92:153–159. [PubMed: 11443621]
5. 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. [PubMed: 11850344]

6. Gierada DS, Pilgram TK, Ford M, et al. Lung cancer: interobserver agreement on interpretation of pulmonary findings at low-dose CT screening. *Radiology*. 2008; 246:265–272. [PubMed: 18024436]
7. Singh S, Pinsky P, Fineberg NS, et al. Evaluation of reader variability in the interpretation of follow-up CT scans at lung cancer screening. *Radiology*. 2011; 259:263–270. [PubMed: 21248232]
8. Brown MS, Goldin JG, Rogers S, et al. Computer-aided lung nodule detection in CT: results of large-scale observer test. *Acad Radiol*. 2005; 12:681–686. [PubMed: 15935966]
9. Goo JM. A computer-aided diagnosis for evaluating lung nodules on chest CT: the current status and perspective. *Korean J Radiol*. 2011; 12:145–155. [PubMed: 21430930]
10. Goo JM, Lee JW, Lee HJ, Kim S, Kim JH, Im JG. Automated lung nodule detection at low-dose CT: preliminary experience. *Korean J Radiol*. 2003; 4:211–216. [PubMed: 14726637]
11. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med*. 2009; 361:2221–2229. [PubMed: 19955524]
12. Gietema HA, Wang Y, Xu D, et al. Pulmonary nodules detected at lung cancer screening: interobserver variability of semiautomated volume measurements. *Radiology*. 2006; 241:251–257. [PubMed: 16908677]
13. Revel MP, Lefort C, Bissery A, et al. Pulmonary nodules: preliminary experience with three-dimensional evaluation. *Radiology*. 2004; 231:459–466. [PubMed: 15128991]
14. Aberle DR, Berg CD, Black WC, et al. The National Lung Screening Trial: overview and study design. *Radiology*. 2011; 258:243–253. [PubMed: 21045183]
15. MacMahon H, Austin JH, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology*. 2005; 237:395–400. [PubMed: 16244247]
16. Crewson PE. Reader agreement studies. *Am J Roentgenol*. 2005; 184:1391–1397. [PubMed: 15855085]
17. Eisenberg RL, Bankier AA, Boiselle PM. Compliance with Fleischner Society guidelines for management of small lung nodules: a survey of 834 radiologists. *Radiology*. 2010; 255:218–224. [PubMed: 20308458]
18. Leader JK, Warfel TE, Fuhrman CR, et al. Pulmonary nodule detection with low-dose CT of the lung: agreement among radiologists. *Am J Roentgenol*. 2005; 185:973–978. [PubMed: 16177418]
19. Gruden JF, Ouanounou S, Tigges S, Norris SD, Klausner TS. Incremental benefit of maximum-intensity-projection images on observer detection of small pulmonary nodules revealed by multidetector CT. *Am J Roentgenol*. 2002; 179:149–157. [PubMed: 12076925]
20. 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. [PubMed: 10405721]
21. Bogot NR, Kazerooni EA, Kelly AM, Quint LE, Desjardins B, Nan B. Interobserver and intraobserver variability in the assessment of pulmonary nodule size on CT using film and computer display methods. *Acad Radiol*. 2005; 12:948–956. [PubMed: 16087090]
22. Revel MP, Bissery A, Bienvenu M, Aycard L, Lefort C, Fria G. Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? *Radiology*. 2004; 231:453–458. [PubMed: 15128990]
23. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung cancer*. 2006; 54:177–184. [PubMed: 16989922]
24. Armato SG 3rd, Sensakovic WF. Automated lung segmentation for thoracic CT impact on computer-aided diagnosis. *Acad Radiol*. 2004; 11:1011–1021. [PubMed: 15350582]
25. 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; 237:684–690. [PubMed: 16244277]
26. Sahiner B, Chan HP, Hadjiiski LM, et al. Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: analysis of an observer performance study by nodule size. *Acad Radiol*. 2009; 16:1518–1530. [PubMed: 19896069]
27. Beigelman-Aubry C, Raffy P, Yang W, Castellino RA, Grenier PA. Computer-aided detection of solid lung nodules on follow-up MDCT screening: evaluation of detection, tracking, and reading time. *Am J Roentgenol*. 2007; 189:948–955. [PubMed: 17885070]

28. Park EA, Goo JM, Lee JW, et al. Efficacy of computer-aided detection system and thin-slab maximum intensity projection technique in the detection of pulmonary nodules in patients with resected metastases. *Invest Radiol.* 2009; 44:105–113. [PubMed: 19034026]
29. Goo JM, Tongdee T, Tongdee R, Yeo K, Hildebolt CF, Bae KT. Volumetric measurement of synthetic lung nodules with multi-detector row CT: effect of various image reconstruction parameters and segmentation thresholds on measurement accuracy. *Radiology.* 2005; 235:850–856. [PubMed: 15914478]
30. Winer-Muram HT, Jennings SG, Meyer CA, et al. Effect of varying CT section width on volumetric measurement of lung tumors and application of compensatory equations. *Radiology.* 2003; 229:184–194. [PubMed: 14519875]
31. Hwang J, Chung MJ, Bae Y, Shin KM, Jeong SY, Lee KS. Computer-aided detection of lung nodules: influence of the image reconstruction kernel for computer-aided detection performance. *J Comput Assist Tomogr.* 2010; 34:31–34. [PubMed: 20118719]
32. Godoy MC, Naidich DP. Subsolid pulmonary nodules and the spectrum of peripheral adenocarcinomas of the lung: recommended interim guidelines for assessment and management. *Radiology.* 2009; 253:606–622. [PubMed: 19952025]
33. Goo JM, Park CM, Lee HJ. Ground-glass nodules on chest CT as imaging biomarkers in the management of lung adenocarcinoma. *Am J Roentgenol.* 2011; 196:533–543. [PubMed: 21343494]

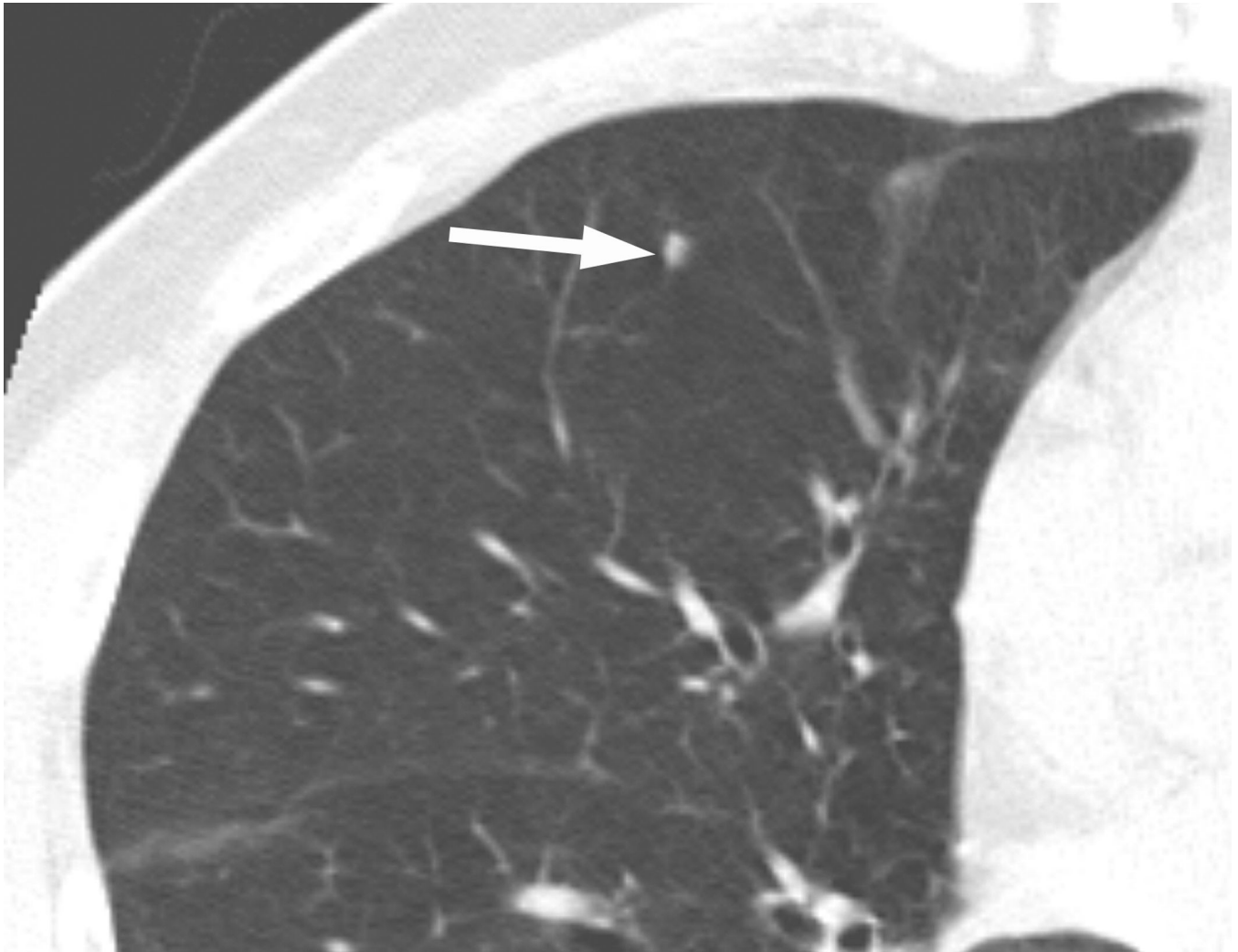


Figure 1.

A nodule in the right middle lobe (arrow) was classified as positive by three readers and negative by four readers at the initial assessment. Three readers recommended a low-level follow-up, while four readers recommended no follow-up. After the review of CAD results, all readers classified this nodule as negative and recommended no follow-up.

TABLE 1

Comparison of multirater generalized kappa on the positivity of screening and follow-up recommendations before and after reviewing CAD results

	Kappa without CAD	Kappa with CAD
Positivity of screening	0.53	0.66
Follow-up recommendations	0.54	0.67

TABLE 2

PreCAD & PostCAD Positivity Kappa (k value among all reader pairs on the positivity of screening results before and after reviewing CAD results)

preCAD postCAD	Readers						
	1	2	3	4	5	6	7
1		0.59	0.61	0.53	0.49	0.58	0.56
2	0.76		0.64	0.46	0.52	0.55	0.43
3	0.72	0.67		0.52	0.53	0.59	0.57
Readers 4	0.71	0.56	0.63		0.56	0.44	0.52
5	0.70	0.56	0.56	0.76		0.54	0.54
6	0.79	0.63	0.65	0.69	0.72		0.37
7	0.66	0.64	0.57	0.59	0.68	0.61	

Note: The κ values for all reader pairs ranged from 0.37 to 0.61 at initial assessment, and from 0.56 to 0.79 after reviewing CAD results.

TABLE 3

PreCAD and postCAD Follow-up Recommendations (κ value among all reader pairs on the follow-up recommendations before and after reviewing CAD results)

preCAD postCAD	Readers							
	1	2	3	4	5	6	7	
1	0.58	0.56	0.54	0.53	0.56	0.56	0.53	
2	0.77		0.52	0.56	0.52	0.57	0.52	
3	0.79	0.72		0.56	0.53	0.54	0.58	
Readers	4	0.69	0.60	0.61		0.53	0.56	0.53
	5	0.79	0.66	0.69	0.73		0.59	0.57
	6	0.87	0.76	0.74	0.69	0.73		0.49
	7	0.77	0.70	0.64	0.61	0.75	0.68	

Note: The κ values among all reader pairs ranged from 0.49 to 0.59 at the initial assessment and ranged from 0.60 to 0.87 after reviewing CAD results.

TABLE 4

Reasons for decision change by readers after reviewing CAD results

Pre-CAD	Post-CAD	Reasons of change	R1	R2	R3	R4	R5	R6	R7	mean
NoFU	LLFU	size	5	2	1	7	1	2	7	4
		detection	5	1	7	13	4	0	3	5
	HLFU	size	0	0	0	0	0	0	0	0
		detection	0	0	0	0	0	0	0	0
LLFU	NoFU	size	9	11	1	0	3	20	3	7
		detection	0	0	0	0	0	0	0	0
	HLFU	size	0	0	0	3	1	0	0	1
		detection	0	0	0	0	1	0	0	0
HLFU	NoFU	size	0	0	0	0	0	0	1	0
		detection	0	0	0	0	0	0	0	0
	LLFU	size	3	3	3	0	5	6	4	3
		detection	0	0	0	0	0	0	0	0
sum		size	17	16	5	10	10	28	15	14
		detection	5	1	7	13	5	0	3	5
		total	22	17	12	23	15	28	18	19

Note: The readers changed their decisions after reviewing CAD results in 12 to 28 cases (mean, 19), due to newly detected nodules by CAD in 0 to 13 cases (mean, 5) and due to change of the nodule size by CAD in 5 to 28 cases (mean, 14).

R1, Reader 1; R2, Reader 2; R3, Reader 3; R4, Reader 4; R5, Reader 5; R6, Reader 6; R7, Reader 7; NoFU, no follow-up; LLFU, low-level follow-up; HLFU, high-level follow-up; size, readers' decision changes due to change of the nodule size by CAD; detection, readers' decision changes due to newly detected nodules by CAD