10.1148/radiol.2016150497

Appendix E1

Subjects enrolled in this clinical trial were referred for, or had recently undergone, CT of the chest for suspicion of a pulmonary nodule or other indications unrelated to pulmonary nodules. The inclusion criteria for the study were broad, so as to represent a range of patients undergoing chest radiographic examinations. The following were the inclusion criteria to qualify a subject for participation in this clinical study: *(a)* subjects scheduled to undergo chest CT as part of their needed medical care, *(b)* subjects 18 years of age and older, and *(c)* subjects in good enough physical condition to stand motionless and hold their breath during the image acquisition procedure.

The following were the exclusion criteria to disqualify a subject from this clinical study: *(a)* subjects who were under 18 years of age, *(b)* female subjects who were pregnant or suspected they may be pregnant, *(c)* subjects with at least one nodule larger than 20 mm (to minimize the possibility of a large nodule obscuring other nodules in the projection view), *(d)* subjects whose previous images showed objects in or around the lungs that might produce substantial artifacts that could obscure pulmonary nodules, and *(e)* subjects whose recent images showed active lung or pleural disease that could obscure pulmonary nodules.

In addition, there were two exclusion criteria to prevent the recruitment of subjects with an excessive number of nodules in one lung or in both lungs. These criteria were designed to address the potential bias against conventional projection radiographic imaging by minimizing the probability of one nodule obscuring another by means of superposition and to minimize the possibility of satisfaction of search (1) errors arising from scoring multiple nodules in a single case: *(a)* subjects with more than five pulmonary nodules between 5 mm and 20 mm in diameter in either the right or left lung and *(b)* subjects suspected to have more than 15 total nodules between 3 mm and 20 mm in the initial interpretation of prior images for inclusion determination. Note that up to 20 nodules between 3 mm and 20 mm would be allowed in the final study sample after the thorough truth panel assessment procedure.

There were 187 subjects who provided informed consent and were enrolled in the study. All but one of these subjects underwent all of the imaging procedures (case no. 163 withdrew before completion). Prior to the truth panel process, 27 unique cases were excluded from the study for reasons that included minor protocol deviations and enforcement of the inclusion and exclusion criteria as summarized below (note that some cases were excluded for multiple reasons):

- More than 15 nodules (case no. 4)
- More than five nodules of 5–20 mm in each lung (case no. 81)
- Artifacts: patient motion during tomosynthesis image acquisition (case nos. 36, 43, 89, 90, and 100)
- Diseases that could obscure the solitary pulmonary nodule, such as fibrosis, emphysema, fluid-obscuring bases, compressed lung, mass, scarring, severe lung disease,

and patchy airspace disease (case nos. 12, 15, 16, 42, 45, 89, 90, 92, 98, 100, 106, 159, and 169)

• Missing CT sections or incomplete scan coverage (case nos. 61, 75, 76, 135, 161, and 182)

• Incorrect CT reconstruction: only bone window or 5-mm sections available (case nos. 12, 34, 36, 74, and 84)

• Missing tomosynthesis data (case #34)

• Subject dropped out of study (case no. 163)

The demographic characteristics of the 160 subjects that underwent the truth panel process are listed in Tables E1 and E2.

Of the 160 cases that underwent the truth panel process, two additional cases (nos. 26 and 35) were then excluded from the reader study because there were more than 20 nodules present; this resulted in 158 cases included in the reader study. The final ratio of positive to negative cases was 2.67:1.

There were also two additional minor deviations from the protocol. One case (case no. 5) was included in the study even though it contained a single nodule larger than 20 mm. There were two cases in which some nodule information was not recorded by the truth panel. In case no. 150, one calcified nodule was inadvertently not reconciled by the truth panel into coordinates on the PA views for conventional radiography or tomosynthesis, but no readers recorded any marks near this nodule location and it therefore did not affect the results. In case no. 180, one truth panelist noted a nodule with coordinates, but no characteristics were recorded; in later review, this nodule could not be found by another truth panelist. No readers marked this nodule, so it also did not affect the results.

Table E1: Age and Sex of Subjects That Fulfilled the Inclusion and Exclusion Criteria

Age	No. of Men	No. of Women	Total
20–29 y	4	3	7
30–39 y	9	3	12
40–49 y	10	9	19
50–59 y	24	21	45
60–69 y	25	22	47
70–79 y	16	12	28
80–89 y	1	1	2
Total	89	71	160

Table E2: Racial Distribution of Subjects Included in the Clinical Trial

Race	No. of Subjects
White	147
Black or African American	7
Asian	3
Native Hawaiian or other Pacific Islander	1
Other	2

Total	160

Appendix E2

Method to Establish Ground Truth

Three thoracic radiologists (H.P.M., E.A.K., and G.P.R., all American Board of Radiology certified, with 19, 18, and 14 years of experience, respectively) served as the truth panel and reviewed the CT images to identify the correct locations of all visible pulmonary nodules. Freely available image display software (OsiriX; Pixmeo, Geneva, Switzerland) was used to display and measure nodule diameters. The truth panel viewed CT images reconstructed with filtered back-projection methods at 1.25-mm or thinner section thickness by using the standard of care at each institution, with a standard axial reconstruction kernel. Locations of all nodules larger than 3.0 mm were marked, and diameters were measured. All truth panel assessment was performed at the contract research organization (ACR Image Metrix, Philadelphia, Pa) hired to oversee the collection, interpretation, and analysis of observer data.

Each CT case was reviewed by two truth panelists. If the first two truth panelists did not completely agree on the presence, location (greater than 1-cm discrepancy), size (greater than 20% disparity in diameter), or characterization of a nodule or the recommended management for the case, a third truth panelist arbitrated what would be considered ground truth for the case. The arbitrator decided which of the two prior truth panelists was correct and could not change any findings or insert new findings. For size disparities less than 20%, the true diameter was deemed to be the mean diameter measured by the two truth panelists. During the truth panel process, a rotation was implemented such that each of the three truth panelists reviewed approximately two-thirds of the cases and arbitrated the remaining third of the cases so that no truth panelist arbitrated a case for which that they had already established ground truth findings.

After the determination of ground truth on the CT images, the arbitrating truth panelist then localized each confirmed nodule on the PA and lateral radiographs and the tomosynthesis images (locations on the DE images were the same as on the conventional PA radiographs) to reconcile locations with the subsequent reader markings. For nodules identified on the CT images that could not be visualized on the conventional radiographs or tomosynthesis images, the reconciling truth panelist marked the approximate location of the nodule in question and recorded that the location was an estimate. The truth panel also recorded an interpretation of nodule characteristics and recommended case management on the basis of whether additional imaging studies should be ordered according to Fleischner Society guidelines for management of pulmonary nodules (2).

Training Images for Observers

Prior to participating in the observer study, a set of training images (prepared by the study sponsor and approved by the chief medical officer of the contract research organization) was shown to each reader to illustrate the corresponding appearance of nodules with each of the three test modalities and CT. None of these cases was included in the analysis. Observers were provided with four image sets of an anthropomorphic chest phantom (Model N1 “Lungman”; Kyoto Kagaku, Kyoto, Japan) with superimposed low- and high-density nodules 3–15 mm in diameter. The four phantom cases had one, three, seven, and eight simulated nodules randomly distributed throughout the lung parenchyma. In addition to the phantom cases, the readers were

presented with nine clinical cases for review, including three normal cases (with no nodules) and cases with one to six nodules. For the phantom and clinical cases, the readers were presented with the true nodule findings and other relevant information about the case, including description of normal and pathologic features that could mimic or obscure nodules.

Reconciliation

Analysis of nodule detection accuracy required reconciling reader markings of suspected nodule locations on the conventional radiographs, DE images, and tomosynthesis images with locations of known nodules on CT images. The three-dimensional distance from each marked lesion to each reconciled true lesion was then computed in the conventional radiography, DE, and tomosynthesis image sets. If a given mark was located within 12 mm of a true nodule, independent of nodule size, then that mark was counted as a correct lesion localization, with the 12-mm criterion determined as optimal after reviewing the asymptotic accuracy by using multiple distance thresholds over the full set of lesion marks.

Adjudication

The study sponsor and contract research organization conducted a data image-quality control process to review all markings to ensure accuracy and consistency. Adjudication of less than 10% (63 of 790 reader-case combinations) of markings was required, owing to factors such as observer error in recording (eg, markings placed on both the lateral and PA radiographs for the same nodule or errors in nodule naming). This adjudication process was performed by the contract research organization and involved careful review of each mark to ensure that the observer's intended target was correctly identified and recorded. An expert radiologist who did not participate as a reader or truth panelist gave the final professional opinion on the adjudicated locations of the few marks in question.

Appendix E3

Questions Asked of Readers

The following data were collected from the readers for each case. This report summarizes findings relative to detection and case management only. Further analysis based on other nodule characteristics will be the subject of future reports.

Nodule Localization

1. Confidence of presence (1: 1–5%; 2: 6–20%; 3: 21–50%; 4: 51–95%; 5: >95%)

2. Best view (PA or lateral for conventional radiography; soft-tissue or bone for DE; slice number for tomosynthesis)

3. Location (right upper lobe, right middle lobe, right lower lobe, left upper lobe, left lower lobe); nodule location markings were also recorded

4. Location within the parenchyma (freestanding, juxtapleural, juxtavascular)

5. Margins (smooth, lobulated, irregular, spiculated, indeterminate)

6. Density (solid, part-solid or semisolid, nonsolid or ground-glass opacity, indeterminate)

7. Calcification (none, central, diffuse solid, laminated, popcorn-like)

8. Distinguishing internal characteristics (none, cavity, fat, other)

Case Management

1. Would you refer this patient for further imaging based on nodule findings only? (yes, no)

2. If yes, in what timeframe? (3 months or less, 6 months or less, 12 months or less)

3. If yes, based on your practice, which modality? (select all that apply: biopsy, CT, DE, PET, plain film, tomosynthesis, other)

4. If yes, further imaging recommended due to? (nodule findings indeterminate, nodule findings suggestive of follow-up)

5. Confidence that case requires follow-up, based on nodule findings only? (1: 0–5%; 2: 6–20%; 3: 21–50%; 4: 51–95%; 5: >95%)

6. Additional specific findings (none, adenopathy, atelectasis, bronchiectasis, cardiomegaly, compression of vertebral bodies, coronary artery calcium, emphysema, enlarged pulmonary arteries, hyperinflation, interstitial lung disease, pleural effusion or thickening, pulmonary fibrosis, pulmonary venous HTN or congestion, other)

HTN = hypertension.

Appendix E4

Details of Primary and Secondary Analyses

Primary Analysis 1: Accuracy of Nodule Detection and Localization on a Nodule-Level Basis (3)

• Maximum LLF: the number of correct lesion localizations, at any FROC rating (ie, 1–5), divided by the total number of true nodules

• Maximum NLF: the number of false marks for normal cases, at any rating, divided by the total number of normal cases

• Weighted JAFROC FOM: the probability that a correct lesion localization rating exceeds that of any mark for a normal case (4–6)

These measures are conceptually similar to sensitivity, false-positive rate, and overall accuracy, respectively, in the FROC context (7). Multiple nodules for the same case were assigned equal weights summing to unity, thus ensuring that each abnormal case received equal importance in the analysis, irrespective of the number of lesions it contained. For the nodule analysis according to size, diameter categories that corresponded to the Fleischner Society guidelines (2) were used (3–4 mm, >4 mm to 6 mm, >6 mm to 8 mm, and >8 mm to 20 mm); only markings indicated by the readers to be within a given size range were included in the analyses.

Primary Analysis 2: Accuracy of Nodule Detection on a Case-Level Basis by Ignoring Localization

Such case-level FOMs were inferred from the FROC data.

- Highest-rating inferred sensitivity: the number of abnormal cases with one or more marks at any rating, divided by the total number of abnormal cases
- Highest-rating inferred specificity: the number of normal cases with no marks, divided by the total number of normal cases
- Highest-rating inferred AUC: the empirical probability that the highest-rated mark for an abnormal case exceeds that for a normal case

Primary Analysis 3: Accuracy of the Case Management Decision for Each Modality by Using as a Reference Standard the Case Management Decisions Based Solely on the Diagnostic-Standard CT Findings

For analysis of the case management question, which yielded ROC ratings of 1–5 with no localization information, the standard ROC method was used. The following FOMs were computed:

- Case management sensitivity (the number of abnormal images rated 2 or more, divided by the total number of abnormal cases)
- Case management specificity (the number of normal images rated 1, divided by the total number of normal cases)
- Case management AUC for each modality combination

Secondary Analyses

Secondary analyses were performed by evaluating nodule localization FOMs on the following subsets of nodules: *(a)* nodules visible to the truth panel on conventional radiographs (ie, nodules with locations that could be identified by the truth panel on conventional radiographs), *(b)* nodules not visible to the truth panel on conventional radiographs (ie, nodules with locations on conventional radiographs that had to be estimated), *(c)* solid nodules, *(d)* calcified nodules, and *(e)* nodules located in the upper lung (both upper lobes). For the secondary analyses *c*–*e*, only markings indicated by the readers to be in the specified category (eg, calcified, solid, or upper lung) were included in the analyses. For the secondary analyses *a* and *b* according to subtlety, true lesion localizations were counted if the correctly marked lesion was in the category of visible or not visible on conventional radiographs, as specified by the truth panel; however, nonlesion localization markings included the same full set of false-positive marks in both the visible and not visible on conventional radiograph categories, due to the impossibility of the readers knowing whether a given mark would have been visible to the truth panel or not on conventional radiographs or only on CT images.

Appendix E5

Previous authors have investigated the use of tomosynthesis for the detection of pulmonary nodules by using a number of different study designs. Dobbins et al (8) found a 3.2-fold improvement in nodule sensitivity with tomosynthesis relative to conventional radiography for expert readers when the locations of the CT-confirmed nodules were known to the readers. Vikgren et al (9) demonstrated a 3.5-fold improvement in LLF for detection of pulmonary

nodules, compared with the 3.6-fold improvement in maximum LLF found in the current study; however, Vikgren et al showed an approximately 50% increase in NLF (similar to false-positive findings), but no significant difference in maximum NLF was found in the current study. Jung et al found a 3.1-fold improvement in detection accuracy with tomosynthesis over conventional radiography for all nodules and a total detection sensitivity of 93% with tomosynthesis for nodules that were proven metastases (10). Yamada et al showed an improvement in true-positive fraction from 0.37 with radiography to 0.80 with tomosynthesis but an increase in false-positive rate from 0.10 with radiography to 0.15 with tomosynthesis (11). The results of this current study demonstrate a similar relative performance increase for nodule detection sensitivity with chest tomosynthesis compared with conventional radiography (roughly threefold) but lower raw performance scores for each modality than those in previous studies. For example, the study by Vikgren et al (9) showed LLF values of 0.16 and 0.56 for conventional radiography and tomosynthesis, respectively, when averaged over observers and nodule sizes, but in our study, we found mean maximum LLF values of 0.038 and 0.135, respectively. Several factors may account for these lower raw performance scores compared with those in previous studies. First, most of the readers in the current study were not experienced thoracic radiologists. The nonlesion localization rate (similar to a false-positive rate) was lower in this study than in previously reported studies for both conventional radiography and tomosynthesis (9), reflecting an apparent conservative bias on the part of the readers that would be consistent with their not being thoracic specialists (although even thoracic radiologists demonstrated a trend toward stricter interpretation thresholds with lower nonlesion localization rate after additional training with tomosynthesis [12]). Second, and probably most important, the distribution of nodule sizes in this study included more small nodules up to and including 6 mm (410 of 516; 79%) than in the Vikgren et al (9) study (84 of 131 nodules; 64%), and the truth panel process involved the use of thinner CT sections in the current study than those in the Vikgren et al study, resulting in a greater preponderance of small nodules being included in the current data set of confirmed nodules.

In this study, we also compared tomosynthesis versus DE imaging, in addition to comparing it versus conventional radiography. In this three-way comparison, conventional radiography plus DE imaging was not found to be superior to conventional radiography alone, unlike most previous studies in which a significant improvement was found (13–18). Our study was powered to demonstrate the larger effect of the primary aim to compare tomosynthesis versus conventional radiography; although an increase in inferred case-level sensitivity of conventional radiography plus DE imaging versus conventional radiography alone was observed (ratio, 1.27), the *P* value ($P = .029$) was not sufficient to demonstrate a significant difference at the level required by the Bonferroni correction for multiple analyses ($P < .017$). In addition, the current study was prospective, unlike many of the previous studies in the literature, and therefore included a range of nodules that were not preselected to be subtle. In the current study, we also found no significant difference in nodule detection when adding DE imaging to tomosynthesis. Such a finding is not surprising because the main advantage of DE imaging for nodule detection is that it diminishes the distracting presence of the ribs and thereby improves conspicuity; with tomosynthesis, the distracting presence of the ribs and that of overlying soft-tissue structures are removed, as well. The only case where DE imaging might be expected to demonstrate superiority to tomosynthesis would be in the characterization of the calcium content of nodules; however, no such improvement was found in this study, most likely because of the small number of calcified nodules in this study cohort (52 of 516 nodules). A study that includes a larger number of

calcified nodules will be required to address specific comparisons of performance to that type of nodule.

References

1. Berbaum KS, Franken EA Jr, Dorfman DD, et al. Satisfaction of search in diagnostic radiology. Invest Radiol 1990;25(2):133–140.

2. 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(2):395–400.

3. Chakraborty DP. JAFROC: software for planning and analyzing FROC studies. http://www.devchakraborty.com. Published 2013. Accessed June 1, 2016.

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

5. Chakraborty DP. A brief history of free-response receiver operating characteristic paradigm data analysis. Acad Radiol 2013;20(7):915–919.

6. Thompson JD, Hogg P, Manning DJ, Szczepura K, Chakraborty DP. A free-response evaluation determining value in the computed tomography attenuation correction image for revealing pulmonary incidental findings: a phantom study. Acad Radiol 2014;21(4):538–545.

7. Popescu LM. Nonparametric signal detectability evaluation using an exponential transformation of the FROC curve. Med Phys 2011;38(10):5690–5702.

8. Dobbins JT 3rd, McAdams HP, Song JW, et al. Digital tomosynthesis of the chest for lung nodule detection: interim sensitivity results from an ongoing NIH-sponsored trial. Med Phys 2008;35(6):2554–2557.

9. Vikgren J, Zachrisson S, Svalkvist A, et al. Comparison of chest tomosynthesis and chest radiography for detection of pulmonary nodules: human observer study of clinical cases. Radiology 2008;249(3):1034–1041.

10. Jung HN, Chung MJ, Koo JH, Kim HC, Lee KS. Digital tomosynthesis of the chest: utility for detection of lung metastasis in patients with colorectal cancer. Clin Radiol 2012;67(3):232–238.

11. Yamada Y, Jinzaki M, Hasegawa I, et al. Fast scanning tomosynthesis for the detection of pulmonary nodules: diagnostic performance compared with chest radiography, using multidetector-row computed tomography as the reference. Invest Radiol 2011;46(8):471–477.

12. Zachrisson S, Vikgren J, Svalkvist A, et al. Effect of clinical experience of chest tomosynthesis on detection of pulmonary nodules. Acta Radiol 2009;50(8):884–891.

13. Ishigaki T, Sakuma S, Ikeda M. One-shot dual-energy subtraction chest imaging with computed radiography: clinical evaluation of film images. Radiology 1988;168(1):67–72.

14. Oestmann JW, Greene R, Rhea JT, et al. “Single-exposure” dual energy digital radiography in the detection of pulmonary nodules and calcifications. Invest Radiol 1989;24(7):517–521.

15. Kelcz F, Zink FE, Peppler WW, Kruger DG, Ergun DL, Mistretta CA. Conventional chest radiography vs dual-energy computed radiography in the detection and characterization of pulmonary nodules. AJR Am J Roentgenol 1994;162(2):271–278.

16. Ricke J, Fischbach F, Freund T, et al. Clinical results of CsI-detector-based dual-exposure dual energy in chest radiography. Eur Radiol 2003;13(12):2577–2582.

17. Fischbach F, Freund T, Röttgen R, Engert U, Felix R, Ricke J. Dual-energy chest radiography with a flat-panel digital detector: revealing calcified chest abnormalities. AJR Am J Roentgenol 2003;181(6):1519–1524.

18. Uemura M, Miyagawa M, Yasuhara Y, et al. Clinical evaluation of pulmonary nodules with dual-exposure dual-energy subtraction chest radiography. Radiat Med 2005;23(6):391–397.