Characterizing Search, Recognition, and Decision in the Detection of Lung Nodules on CT Scans: Elucidation with Eye Tracking[1]

Geoffrey D. Rubin, MD
Justus E. Roos, MD
Martin Tall, MSc
Brian Harrawood, BA
Sukantadev Bag, MSc
Donald L. Ly, MD
Danielle M. Seaman, MD
Lynne M. Hurwitz, MD
Sandy Napel, PhD
Kingshuk Roy Choudhury, PhD

Purpose: To determine the effectiveness of radiologists' search, recognition, and acceptance of lung nodules on computed tomographic (CT) images by using eye tracking.

Materials and Methods: This study was performed with a protocol approved by the institutional review board. All study subjects provided informed consent, and all private health information was protected in accordance with HIPAA. A remote eye tracker was used to record time-varying gaze paths while 13 radiologists interpreted 40 lung CT images with an average of 3.9 synthetic nodules (5-mm diameter) embedded randomly in the lung parenchyma. The radiologists' gaze volumes (GVs) were defined as the portion of the lung parenchyma within 50 pixels (approximately 3 cm) of all gaze points. The fraction of the total lung volume encompassed within the GVs, the fraction of lung nodules encompassed within each GV (search effectiveness), the fraction of lung nodules within the GV detected by the reader (recognition-acceptance effectiveness), and overall sensitivity of lung nodule detection were measured.

Results: Detected nodules were within 50 pixels of the nearest gaze point for 990 of 992 correct detections. On average, radiologists searched 26.7% of the lung parenchyma in 3 minutes and 16 seconds and encompassed between 86 and 143 of 157 nodules within their GVs. Once encompassed within their GV, the average sensitivity of nodule recognition and acceptance ranged from 47 of 100 nodules to 103 of 124 nodules (sensitivity, 0.47–0.82). Overall sensitivity ranged from 47 to 114 of 157 nodules (sensitivity, 0.30–0.73) and showed moderate correlation ($r = 0.62$, $P = .02$) with the fraction of lung volume searched.

Conclusion: Relationships between reader search, recognition and acceptance, and overall lung nodule detection rate can be studied with eye tracking. Radiologists appear to actively search less than half of the lung parenchyma, with substantial interreader variation in volume searched, fraction of nodules included within the search volume, sensitivity for nodules within the search volume, and overall detection rate.

Online supplemental material is available for this article.

[1] From the Duke Clinical Research Institute, Box 17969, 2400 Pratt St, Durham, NC 27715 (G.D.R., K.R.C.); Department of Radiology, Duke University School of Medicine, Durham, NC (G.D.R., J.E.R., M.T., B.H., S.B., D.M.S., L.M.H.); Department of Medical Imaging, University of Toronto, Toronto, ON, Canada (D.L.L.); and Department of Radiology, Stanford University School of Medicine, Stanford, Calif (S.N.). From the 2012 RSNA Annual Meeting. Received January 8, 2014; revision requested February 11; revision received May 1; accepted June 25; final version accepted July 29. **Address correspondence to** G.D.R. (e-mail: *grubin@duke.edu*).

Radiologists rely on their skills as image interpreters to detect and diagnose disease in medical imaging studies. Pulmonary nodule detection is a particularly common and important task facing radiologists who interpret chest computed tomographic (CT) images given the fact that both primary lung cancer and metastases to the lungs typically manifest as pulmonary nodules (1–3). Nevertheless, substantial variability in lung nodule detection has been reported for radiologists who interpret chest CT images (4–7). While the effectiveness of CT scanning in lung cancer detection depends on accurate image interpretation, there is a paucity of investigations focused on assessing and improving radiologists' effectiveness in finding lung nodules on CT images.

Gaze tracking has been used for more than 35 years to gain insights into the detection of lung nodules on chest radiographs (8); however, the study of lung nodule detection on chest CT images presents challenges that are not encountered in the interpretation of radiographs. Specifically, while a chest radiograph involves interpretation of one two-dimensional summation image, CT images presented as contiguous transverse reconstructions present a three-dimensional search space where the search process involves both visual scanning on two-dimensional cross sections and reader-directed navigation through the stack of cross sections.

Investigations of lung nodule detection with gaze tracking during chest radiograph interpretation have led to the hypothesis that there are four components to the detection process: orientation, search, recognition, and decision making (8). As trained observers of chest imaging studies, radiologists should not experience errors of orientation, thereby reducing potential errors to three categories. While this model for lung nodule detection was proposed for radiographic interpretation, we believe that it is equally applicable to the interpretation of chest CT images.

By using a system that maps gaze position onto transverse CT images while the observer navigates the CT volume with cine paging to create a three-dimensional search path, we have studied the process of lung nodule detection on CT images to gain insights into radiologists' search, recognition, and decision (acceptance of detection as a nodule). Our purpose was to determine interreader variation in the search and recognition performance of radiologists with varying levels of experience.

Advances in Knowledge

- Time-varying three-dimensional search paths can be recorded during cine paging CT image interpretation by using eye tracking and by associating gaze position with reader-selected transverse sections.
- Detection of 5-mm lung nodules appears to require that the nodules are within 50 pixels (approximately 3 cm) of the nearest gaze point.
- When gaze volume (GV) is defined as a subvolume of the lung parenchyma within 50 pixels of all gaze points, the GV encompasses an average of 26.7% of the lungs.
- Search effectiveness, nodule recognition, and overall sensitivity of lung nodule detection vary considerably among radiologists.

Materials and Methods

Lung CT Data

This study was performed in accordance with a protocol approved by the Duke University School of Medicine institutional review board. All study subjects provided informed consent, and all private health information was protected in accordance with the Health Insurance Portability and Accountability Act. On the basis of a review by a thoracic radiologist (J.E.R.), three thoracic CT examinations without known lung nodules or other important pulmonary parenchymal abnormalities were selected at random from our clinical population. All patient identifiers were removed from the images. The examinations were performed with a 64 × 0.75 mm CT unit (Sensation 64; Siemens Medical Solutions, Erlangen, Germany), and images were reconstructed contiguously with 1.0-mm thickness and a 30–35-cm field of view. These three examinations are subsequently referred to as substrate CT examinations and comprised 281, 289, and 340 transverse reconstructions, respectively.

A total of 10 lung nodules were synthesized by using a dedicated computer program that allows the user to simulate CT sections of 30-HU nodules that are composed of combinations of primitive three-dimensional shapes (spheres, rods, and cones) of varying sizes and orientations to create spheroid nodules with varying degrees of surface irregularity (Fig 1). All nodules were synthesized with a maximal diameter of 5 mm and a ratio of 1.1 for largest to smallest diameter; thus, the actual nodule diameter varied between 4.8 and 5.0

Implication for Patient Care

- Understanding radiologist performance of lung nodule detection on CT images may contribute to the development of tools to improve consistency and effectiveness in the detection of lung disease.

Published online before print
10.1148/radiol.14132918 **Content code:** CH

Radiology 2015; 274:276–286

Abbreviations:
fT_{GV} = fraction of targets encompassed within the GV
GV = gaze volume
VF = fraction of lung volume encompassed within the GV

Author contributions:
Guarantor of integrity of entire study, G.D.R.; 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, G.D.R., J.E.R., M.T., D.L.L.; clinical studies, J.E.R., D.M.S.; experimental studies, G.D.R., J.E.R., M.T., B.H., S.B., D.L.L., L.M.H., S.N.; statistical analysis, K.R.C.; and manuscript editing, G.D.R., J.E.R., B.H., D.L.L., D.M.S., L.M.H., S.N., K.R.C.

Conflicts of interest are listed at the end of this article.

See also the editorial by Kundel in this issue.

Figure 1

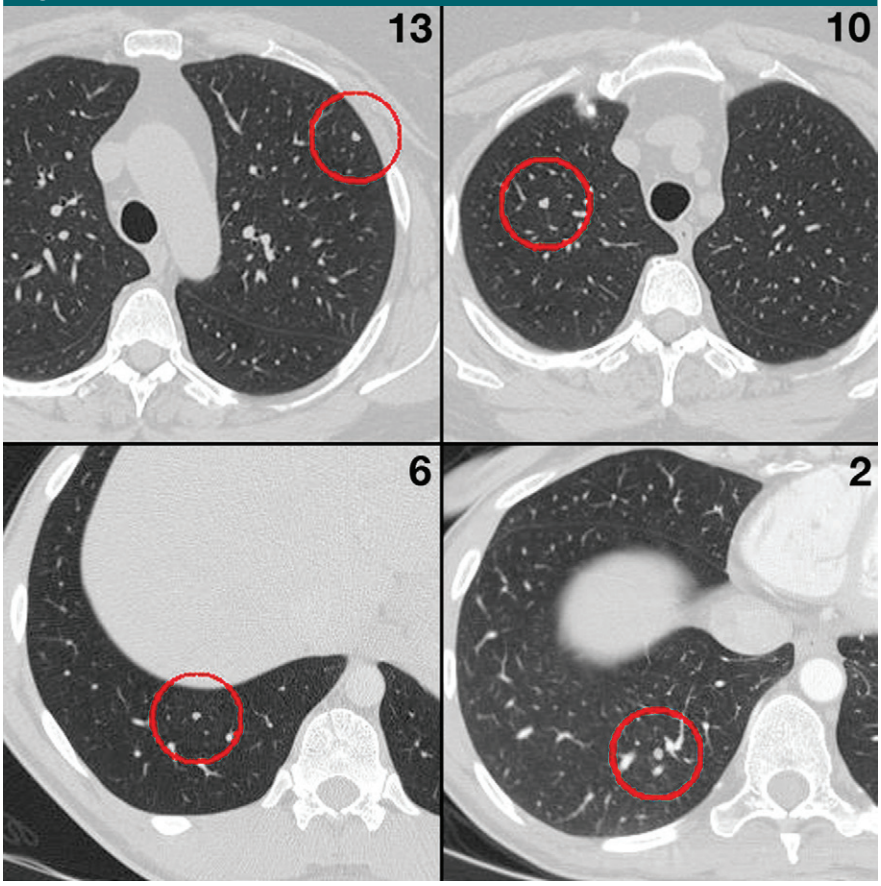


Figure 1: Example CT images show four of the 157 embedded targets. The targets are at the center of the red circles. The number in the upper right corner of each image indicates how many of the 13 readers detected the corresponding nodule.

mm depending on the measurement axis. Volumes consisting of contiguous 1.0-mm-thick CT sections of each of these nodules were then created and merged with the substrate CT sections, as follows. This computer program was developed in our laboratory (9) and is not available commercially.

Forty distinct cases were created by embedding three to six synthetic nodules (mean, 3.9 synthetic nodules) into one of the three substrate patient examinations at different locations within the lungs. Specifically, the five to six synthesized 1-mm-thick CT sections of the nodule models were merged with five to six CT sections centered on one of 157 uniquely selected loci within the pulmonary parenchyma. A thoracic radiologist (G.D.R.) with 22 years of experience interpreting chest CT images positioned the nodules at locations throughout the pulmonary parenchyma that were either free from or contiguous with blood vessels, airways, or other normal structures but did not overlie them (Fig 1). There were 23 nodules in the right upper lobe, 19 in the right middle lobe, 51 in the right lower lobe, 22 in the left upper lobe, and 40 in the left lower lobe. The nature of the 40 cases being composed of three substrate patient CT examinations with 40 unique patterns of nodules embedded was described to each reader prior to the reading sessions. All readers viewed the cases in the same sequence.

Gaze Tracking System

To track and record the reader's gaze, an eye tracking device (SMI IViewX RED; Sensomotoric Instruments, Teltow, Germany) was positioned below the image display. The binocular tracking system relies on dark pupil and corneal reflections created by infrared illumination. Manufacturer specifications state an accuracy of less than 0.5° of visual angle when the participant is positioned within an approximately 40 × 40 × 30 cm (width, height, and depth, respectively) cubic space centered 600 mm from the device. The participant's head is free to move within this cubic region. The device tracks the position of the corneal glint bilaterally with a sampling frequency of 60 Hz and calculates the focal point of the gaze on the surface of the display. The system does not require the participant to wear any equipment and automatically compensates for changes in head position (10).

Participants were seated in front of a 22-inch (55.8-cm) liquid crystal display in a room with stable low ambient light conditions (6 lux with computer display switched off and 13 lux with computer display switched on, as measured at the reader's head position), similar to those encountered in a clinical reading room (Fig 2). At the beginning of each reading session and after every five cases during the reading session, manufacturer-supplied software was used to calibrate reader eye position to an area measuring 278 × 278 mm (1024 × 1024 pixels) centered on the display (10).

Readers sequentially reviewed the 40 simulated CT examinations. Transverse CT sections were displayed with a window level of −500 HU and a window width of 1600 HU. Readers had an option to adjust the window width and level, although none elected to do so. By using the scroll wheel on a standard computer mouse, readers were free to contiguously scroll through the CT sections and were instructed to identify all pulmonary nodules within each case. They were told that the examinations were enriched with lung nodules and to perform a comprehensive search so that they might find them all. They were not told that all examinations had nodules embedded. Also, they were encouraged to take as much time as needed and to

Figure 2

Figure 2: Experimental set-up with the reader positioned in front of the display showing a sample CT section. The eye tracking recorder (arrow) is positioned below the display. Consecutive points along the gaze path are measured 60 times per second. The x and y coordinates are determined by eye position, and the z coordinate is determined by the transverse section selected by using the mouse, which is located under the observer's right hand. Low ambient lighting conditions were used during reading sessions. This photograph was obtained with high ambient lighting for demonstration purposes.

divide the 40 cases into as many reading sessions as felt comfortable, resulting in two to three sessions per reader.

Beginning with the first scrolling action, the x, y, and z (section number) coordinates of the gaze location were recorded at a 60-Hz rate, thereby providing three-dimensional gaze coordinates ($g_x[t]$, $g_y[t]$, $g_z[t]$) at each time point (t) over the duration of the entire search. When readers identified a nodule candidate, they were asked to denote it by clicking the mouse button with the cursor centered on the nodule. They also provided a confidence rating between 1 and 5. The confidence ratings were defined as follows: 5, definitely a nodule; 4, probably a nodule; 3, possibly a nodule; 2, unlikely to be a nodule; and 1, very unlikely to be a nodule. Recording was paused automatically during the reporting process and resumed automatically thereafter.

Readers

Thirteen radiologists with 1–25 years of experience interpreting CT images (mean, 8.6 years ± 8.0 [standard deviation]) were recruited and completed the study. Readers comprised three 1st-year radiology residents, three 4th-year radiology residents, two clinical body imaging fellows, and five attending thoracic radiologists.

Quantification of Gaze Data

A total of 520 complete three-dimensional gaze paths were acquired and analyzed relative to 2041 nodule detection opportunities. For each case and reader, the search path was defined as the union of all gaze points recorded on all CT sections over the entire reading. Two summary measures from the gaze data were computed for each reader and each case pairing: *(a)* the volume fraction (VF) of the lung that was enclosed within the gaze volume (GV) and *(b)* for each embedded nodule or target the nearest distance (in pixels) to the search path. The VF is based on the concept that there is a gaze cone that represents a volumetric region extending from the observer's eyes toward infinity, such that an unobscured object present within the gaze cone is visible to the observer (11). For our analysis, we assume that the gaze cone at any point in time subtends a circular region of interest centered on the gaze point. We define the GV as the volume represented by the union of the intersections of the time-varying gaze cone with the CT sections. For the purpose of calculating VF, the lungs were segmented from the chest wall and mediastinum; the GV was truncated to regions exclusively within the lungs (Fig 3) and normalized by the volume of the segmented lungs. The specific method for calculating these measures is presented in Appendix E1 (online).

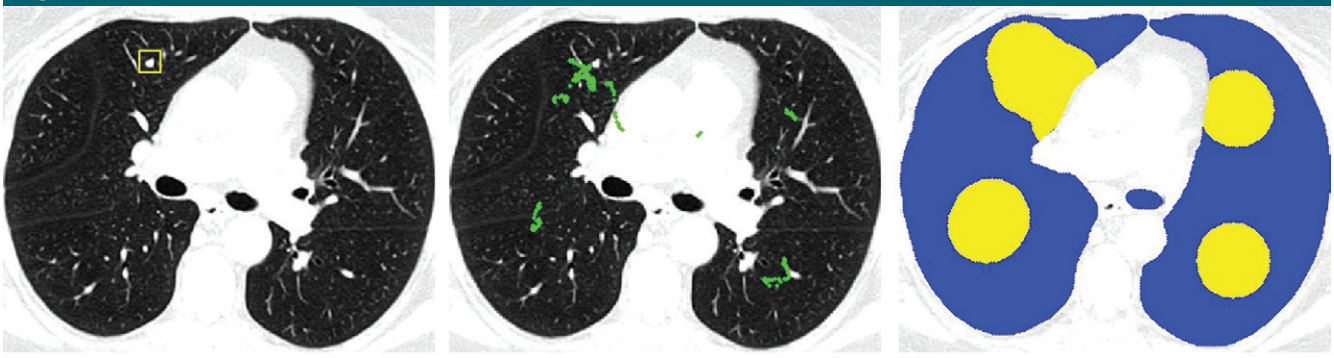
Analysis

Detections of preexisting native nodules within the lung volumes were ignored owing to the absence of a reference standard; thus, specificity was not calculated. The minimum distance between the nearest gaze point and each of the targets, both detected and undetected, was measured and used to determine the GV. The search duration, VF, and trajectories were measured across the 13 readers and 40 cases. Overall observer performance was measured as the fraction of embedded targets detected (overall sensitivity). Performance was further subdivided into *search* and *recognition and acceptance*. The effectiveness and efficiency of search were measured as the fraction of targets encompassed within the GV (fT_{GV}). The efficiency of search was analyzed in two ways: *(a)* as fT_{GV} divided by VF and *(b)* the number of targets enclosed within the GV divided by search time (in minutes). The effectiveness of recognition and acceptance was measured as the fraction of targets present within the GV that were detected.

Statistical Analysis

Correlations between overall reader sensitivity and percentage of lung volume examined and between overall reader sensitivity and search speed were assessed. The dependence of overall reader sensitivity was modeled as a function of VF covered and years

Figure 3



a. b. c.

Figure 3: **(a)** Transverse section of a volumetric CT image shows an example of an embedded nodule (yellow box). The nodule has a maximal diameter of 42 pixels and was detected by 11 of 13 readers. **(b)** Same CT section as in **a**, with all gaze points associated with one observer and corresponding to this CT section shown in green. **(c)** A 50-pixel dilation about each gaze point with truncation at the mediastinal surface results in the yellow regions corresponding to lung regions searched versus the blue regions, which are outside the GV. GV is computed as the sum of the area of the yellow regions across all transverse sections. The proportion of lung volume searched (VF) is computed as the ratio of the GV to the total lung area (yellow and blue areas) after summing these areas across all CT sections.

of experience by using a multivariate linear regression model, which was fit by using the ordinary least squares method and was assessed for covariates with the R software package *(www.r-project.org)*.

Results

Defining GV

For detected targets, the distance between the nearest gaze point and the target was 10.8 pixels ± 10.3 on average and was less than 50 pixels (approximately 3 cm) for 987 (99.8%) of 989 cases (Fig 4). When targets were not detected, the distribution of values for the nearest gaze point to the target distance was much wider, ranging from 0 to more than 100 pixels (mean, 60.4 pixels ± 48.7). On the basis of this result, we calculated the GV by using a circular region with a radius of 50 pixels centered about each gaze point, and VF was derived from a subvolume of the GV where regions outside the lungs were excluded (Fig 3c).

Search Characteristics

Average search duration over the 40 cases for each reader ranged from 1 minute 59 seconds to 5 minutes 55 seconds (mean, 3 minutes 16 seconds [95% confidence interval: 2 minutes 38 seconds, 3 minutes 59 seconds) across the 13 readers, with the shortest and longest individual case searches of 58 seconds and 10 minutes, respectively (Fig 5a). The average percentage of lung volume examined (VF) for the 40 cases ranged from 15% to 43% (mean, 26.7% [95% confidence interval: 22.4%, 30.9%]) across the 13 readers (Fig 5b). Most of the readers had a relatively narrow range of variation across cases (standard deviation <7.5%), with the exception of reader 7 who had a standard deviation of 12%. Reader 7, a fellow, also searched a substantially larger volume of the lungs (43% of the lungs on average) than the other readers. Mean values for VF and search duration by reader were highly correlated (r = 0.89) but ranged from 0.32 to 0.91 within readers.

The temporal trajectories of GV over the course of each reading follow a nearly linear course for all readers and cases, suggesting that over the course of the search period previously unobserved regions of the lung are observed at nearly the same rate throughout the search (Fig 6). Of note, when the search is terminated, there is no indication that the rate of VF growth has begun to plateau, despite an average of 73.7% of the lung remaining unobserved. The average slope of the trajectories shown in Figure 6, representing the average speed with which the search volume is traversed, varied from 6% to 12% of the lung volume per minute of the search. By using linear extrapolation from the observed incremental GV trajectories, the mean time projected to complete the search (coverage of 100% of VF) would be 11 minutes 48 seconds, with a standard deviation of 3 minutes 20 seconds across all readers. For comparison, the actual mean duration of searches was 3 minutes 15 seconds with a standard deviation of 1 minute 39 seconds.

Search Performance

Search performance is characterized in terms of effectiveness and efficiency, as summarized for the 13 observers in Table 1. The observers' GV encompassed 86–143 of the 157 targets, while 14–71 of 157 targets were outside the GV. Thus, search effectiveness as quantified by fT_{GV} ranged from 55% to 91%. Correlation between each reader's average VF and average fT_{GV} was very good (r = 0.79, P = .001), supporting the conclusion that a search that encompasses a greater fraction of the lung results in more nodules within the GV.

Search efficiency is characterized with two metrics, fT_{GV} divided by VF and the fraction of targets encompassed within the GV per minute of search (Table 1). Given the observation

Figure 4

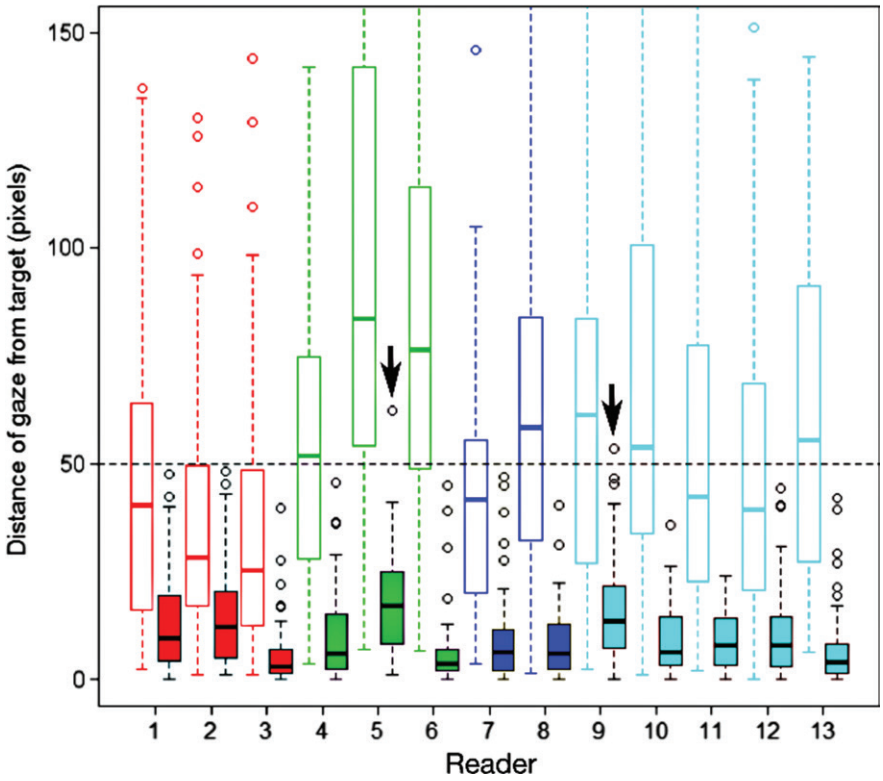


Figure 4: Closest distance from gaze path for detected (solid) and not detected (not solid) targets for each of 13 readers across 158 targets. Boxes span 25th to 75th percentile values, with the internal horizontal bar representing the median. The whiskers extend ± 1.5 × interquartile range from the box. ○ = outlier points. Red, green, blue, and aqua boxes correspond to 1st-year radiology residents, 4th-year radiology residents, fellows, and attending thoracic radiologists, respectively. Across all readers, the 25th, 50th, and 75th percentile distance between a nodule and its nearest gaze point was 3.6, 9.4, and 20 pixels for detected nodules and 23.7, 48.8, and 82 pixels for undetected nodules, respectively. Of the 992 detections made by the 13 readers, only two (arrows) were more than 50 pixels from the gaze path.

that on average GV was less than 50% for all observers, the former measurement characterizes aggregate efficiency by assessing an observer's tendency to capture targets relative to the size of the subvolume searched and on average ranged from 2.27 to 4.54. Readers 4 and 8 encompassed a nearly identical number of targets in their GVs (108 and 107, respectively); however, reader 4's VF was almost twice as large as that of reader 9, corresponding to efficiency metrics (fT_{GV} divided by VF) of 2.55 and 4.54, respectively. The second measure of efficiency reflects the speed with which the reader encompasses targets within their GV and ranged from 0.15 to 0.38 targets per minute. Although a slower search was associated with a greater fT_{GV} across all readers, substantial variations were observed. Reader 1 searched for over twice as long as reader 9; however, reader 1 detected just eight more nodules across the 40 cases.

Diagnostic Performance

Diagnostic performance encompassed the recognition and correct designation of a target as being a lung nodule with a confidence level greater than two of five. While our initial intent was for the confidence ratings to stratify performance at different operating points, there was insufficient variability in the rating selections to do so. Of 1012 true-positive detections, ratings at confidence levels of 5, 4, 3, 2, and 1 were applied across the 13 readers with a frequency of 946 (93.5%), 48 (4.7%), 11 (1.1%), six (0.6%), and one (0.1%), respectively, and seven of 13 readers rated true-positive detections at level 5 exclusively. Overall sensitivity was 989 of 2037 targets (0.49) and ranged from 47 to 114 of 157 targets (0.30–0.73) (Table 1). Sensitivity for targets within the GV ranged from 47 of 100 targets to 103 of 124 targets (range, 0.47–0.82) and totaled 987 of 1518 targets (0.65). These values indicate the frequency with which the readers recognized and accepted nodules within their GV (Table 1). Sensitivity for targets outside of the GV was two of 519 targets (0.4%). As shown in Figure 4, two readers identified one target beyond the designated 50-pixel distance from the nearest gaze point at 54 and 62 pixels from the gaze point, respectively.

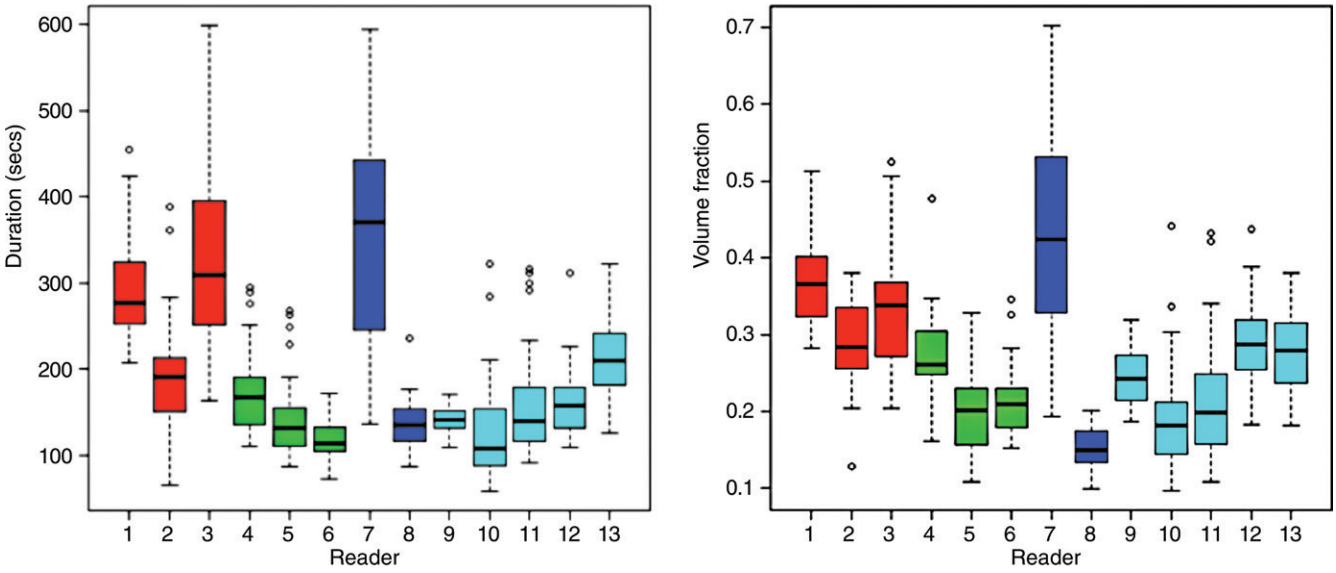
There was moderate correlation (r = 0.62, P = .02) between the average sensitivity for each reader and the average percentage of lung volume examined by the same readers. When we consider that higher search speeds through the GV might be associated with a lower detection rate, we found a weakly positive correlation (r = 0.13, P = .66) between search speed and the likelihood that targets in the search path were detected (sensitivity for targets within the GV). When considering factors contributing to overall sensitivity, we evaluated a multivariate regression model with overall sensitivity as the dependent variable and VF, search duration, and years of experience as independent variables. The model showed a marginally significant increase in overall sensitivity with increasing VF (P = .04), but there was no significant effect due to years of experience (P = .79) (Table 2).

As the 40 cases were presented in the same sequence for all readers, all metrics were assessed to determine if there were order-based trends. None were observed.

Discussion

The detection of 5-mm lung nodules in CT volumes is a difficult task. The

Figure 5



a. b.

Figure 5: **(a)** Duration of search and **(b)** percentage of lung volume encompassed by the gaze path for each reader. Each box comprises 40 values (one per study). The solid black line in each box indicates the median, box limits indicate the interquartile range, and whiskers (dashed lines) indicate the range of data. ○ = outliers.

three substrate CT examinations reconstructed with 1-mm-thick sections had on average 9163089 voxels within the lungs. A 5-mm-diameter lung nodule occupies approximately 130 voxels or only 1.4×10^{-5} of the lung volume. The task of detection is further challenged by the complexity of the underlying lung structure. Only three of 13 radiologists detected more than 50% of the nodules, with a peak performance of 73%, and nine of 13 radiologists detected between 30% and 50% of the nodules. These results are similar to those reported for nodules of similar size, which were detected by using 1-mm-thick multidetector CT images and confirmed with intraoperative palpation and resection (2).

The detection of lung nodules on CT images by experienced interpreters requires a combination of effective search, recognition, and decision making. Our experiment shows that when the center of a radiologist's gaze is never closer than 50 pixels from a lung nodule, there is less than 1% likelihood that the nodule will be detected. The definition of search volume on the basis of this construct results in the observation

Figure 6

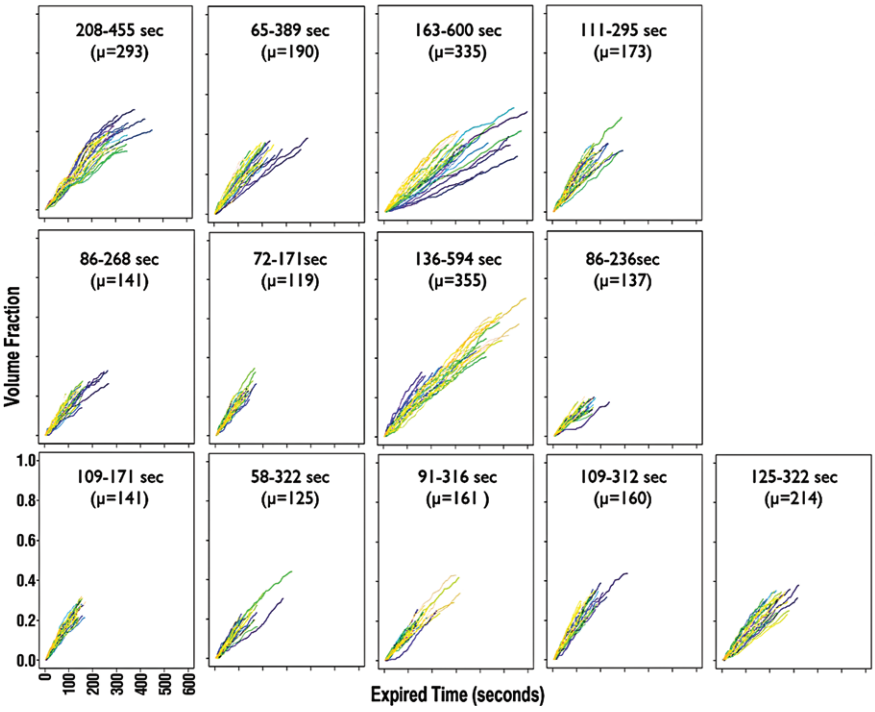


Figure 6: Search trajectories for the 13 readers. Horizontal axes are elapsed time between 0 and 600 seconds, with each increment representing 100 seconds. Vertical axes are VF (lung volume fraction enclosed within the GV). Each line represents one of 40 searches for each observer, with the first through the last searches color coded from dark blue to yellow. Values are the range of search duration, with mean search duration *(μ)* indicated in parentheses.

Table 1

Summary of Performance Metrics for Search, Recognition-Acceptance, and Overall Performance

	Search Performance					Recognition-Acceptance	Overall
Reader No.	VF (%)	Duration (min:sec)	Effectiveness (fT_{GV})	Efficiency (fT_{GV}/VF)	Efficiency (fT_{GV}/min)	Effectiveness (Se_{GV})	Effectiveness (Se_{O})
1	37	4:53	132/157 (84)	2.27	0.17	91/132 (69)	91/157 (58)
2	29	3:10	137/157 (87)	3.01	0.28	74/137 (54)	74/157 (47)
3	33	5:35	132/153 (86)	2.67	0.16	63/132 (48)	63/153 (41)
4	27	2:53	108/157 (69)	2.55	0.24	71/108 (66)	71/157 (45)
5	20	2:21	96/157 (61)	3.06	0.26	78/96 (81)	79/157 (50)
6	21	1:59	86/157 (55)	2.61	0.28	61/86 (71)	61/157 (39)
7	43	5:55	143/157 (91)	2.12	0.15	114/143 (80)	114/157 (73)
8	15	2:17	107/157 (68)	4.54	0.30	72/107 (67)	72/157 (46)
9	25	2:21	124/157 (79)	3.16	0.34	102/124 (82)	103/157 (66)
10	19	2:05	100/157 (64)	3.35	0.31	47/100 (47)	47/157 (30)
11	21	2:41	118/157 (75)	3.58	0.28	67/118 (57)	67/157 (43)
12	29	2:40	122/157 (78)	2.68	0.29	71/122 (58)	71/157 (45)
13	28	3:34	113/157 (72)	2.57	0.20	76/113 (67)	76/147 (48)
All readers*	26.7 [22.4, 30.9]	3:16 [2:38, 3:59]	1518/2037 (74.7) [68.6, 80.7]	2.94 [2.6, 3.3]	0.249 [0.22, 0.28]	987/1518 (65) [58.6, 71.6]	989/2037 (49) [42.3, 54.7]

Note.—Unless otherwise indicated, data in parentheses are percentages. fT_{GV}/min = fraction of all targets encompassed within GV per minute of search, Se_{GV} = sensitivity of target detection for targets within the GV, Se_{O} = overall sensitivity.

* Data are mean values for all readers. Data in brackets are 95% confidence intervals.

Table 2

Multivariate Analysis of Mean Sensitivity per Reader, as a Function of Mean VF and Years of Experience

Term	Estimate	Standard Error	*t* Value	*P* Value
Intercept	0.285	0.109	2.62	.025
VF	0.797	0.345	2.31	.044
Years of experience	−0.001	0.003	−0.26	.797

that, on average, radiologists stopped searching the lungs after only 26.7% of the lung was searched. This is an intriguing observation given that the radiologists were instructed to search the entirety of the lungs to detect all lung nodules. On the basis of the data presented in Figure 6, most searches were accumulating previously unsearched regions of lung at a rate that was consistent throughout the search. The absence of a formal feedback mechanism to indicate if a region of lung has been searched leaves the task of determining when the lungs have been adequately searched to an apparently unreliable internal cue. While there are ample distractions that may interrupt the search process within the context of a clinical reading environment, our experiment was designed to eliminate external distractions.

The lung nodules were distributed throughout the lung parenchyma without regard for location other than to ensure that they were not superimposed on blood vessels or airways. Although the observers searched an average of 26.7% of the lung volume, those search volumes encompassed an average of 75% of the nodules. This apparently paradoxical result might suggest that our definition of the search volume might be overly restrictive and that the gaze cone extends beyond 50 pixels from the gaze point; however the recognition and acceptance of a nodule beyond 50 pixels was extremely rare. One explanation for this apparent paradox might be the existence of a larger gaze cone that influences search and a smaller gaze cone for detailed inspection. Only within the smaller gaze cone is there the possibility that a nodule will be recognized and accepted. Further investigation is required to test this hypothesis.

While the traditional paradigm of lesion detection within images considers lesion recognition and acceptance as two distinct cognitive processes, our current analysis does not allow us to separate these two processes. This is because the act of selecting the target confirmed both recognition and acceptance. Further, we did not evaluate false-positive detections because we did not have a reference standard for the presence or absence of native lesions within the data sets.

While the ability of radiologists to construct a search path that brought their gaze sufficiently close to a nodule for recognition to occur is a critical

Figure 7

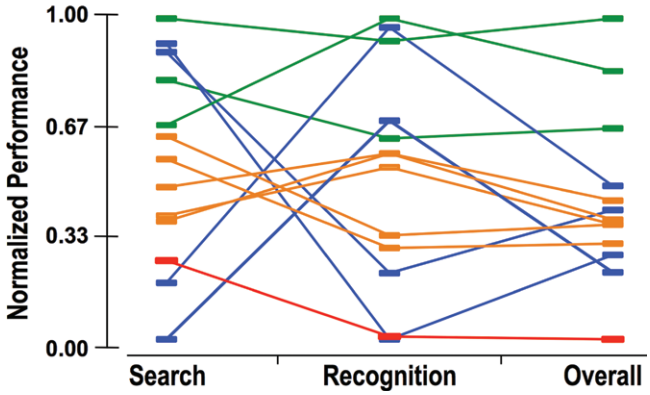


Figure 7: Relative overall performance (overall sensitivity) of 13 observers compared with relative search effectiveness (fT_{GV}) and recognition and acceptance effectiveness (sensitivity for targets within the GV). The values are normalized within each category such that the minimum is set to 0 and the maximum is set to 1. Three observers were in the upper third of performance (green), five observers were in the middle third of performance (orange), one observer was in the lower third of performance (red), and four observers performed inconsistently (blue).

prerequisite for detection, effective search strategies alone (Table 1) did not ensure effective recognition and acceptance of the nodules. While the radiologist with the most effective search (91% fT_{GV}) was also the radiologist with the highest detection rate (73% overall sensitivity), the three next most effective searches (84%–88% fT_{GV}; all 1st-year residents) ranked third, sixth, and 11th in overall sensitivity. In fact, while effectiveness of search (fT_{GV}) and overall effectiveness had a correlation of 0.57 (P = .04), effectiveness of search and effectiveness of recognition and acceptance had a correlation of −0.14 (P = .64), and correlation between sensitivity for targets within the GV and overall sensitivity was 0.68 (P = .01). Figure 7 shows the consistency of performance among readers. Of the four readers with inconsistent performance, two ranked 11th and 12th of 13 readers in search effectiveness and second and fourth of 13 readers in recognition and acceptance effectiveness, respectively, and two ranked second and third of 13 readers in search effectiveness and 11th and 12th of 13 readers in recognition and acceptance effectiveness, respectively.

While initially introduced 36 years ago as a means to study how medical images are searched and interpreted (8), until recently gaze tracking was limited to search of two-dimensional imaging data, most notably chest radiographs and mammograms (12–15) or tiled CT images (16,17). One recent report examines how readers' gaze is influenced when viewing fixed video segments of volume-rendered CT colonograms (18). The association of two-dimensional eye tracking coupled with computer mouse–driven longitudinal translation is a recent innovation used to study three-dimensional search, as is commonly used in the clinical interpretation of CT, magnetic resonance imaging, positron emission tomography, and single photon emission CT data (19). This technique was recently applied to characterize radiologists' search patterns for nodules in lung CT data as that of scanners and drillers. By using a 5° gaze cone angle, Drew and colleagues estimated that drillers and scanners covered an average of 72% and 55% of the lung volume, respectively (20). The fundamental difference between VF calculated by Drew et al and that calculated in the current study is likely related to our current construct for GV, which derives VF based on a 50-pixel radius.

There were several limitations to our study. First, gaze point measurements are subject to error, which introduces variability in the accuracy of gaze measurements and could result in overestimation of the gaze cone as a 50-pixel radius around the gaze point. Second, a 50-pixel (3-cm) radius for defining the GV may be overly restrictive. In our current analysis, we did not attempt to ascertain the initial recognition event, which may have occurred more than 50 pixels from the target center. Moreover, our current analysis of search assesses performance relative to the spatial domain, with limited focus on the temporal domain, such as assessment of dwell times in the proximity of nodules. We anticipate that analyses of gaze behavior over key time segments will bring further insights into the detection process. Third, only 5-mm nodules were assessed; therefore, our results are not generalizable to nodules that are either larger or smaller than 5 mm. Fourth, the effect of proximate lung complexity on nodule detection may be an important influence of nodule detection; however, it was not assessed in this study. Fifth, while performance trends can be gleaned from our analysis, determination of their generalizability and importance are limited by our small sample of 13 radiologists. Sixth, readings were performed in a laboratory setting free of common reading room distractions, such as ringing telephones and unexpected consultations. Readers were given an option to schedule experiments at their convenience, and there were no attempts to measure or control for caffeine intake, timing of prior meal, or fatigue. Seventh, specificity was not measured due to the absence of a reference standard for native nodules. While an experienced thoracic radiologist did not initially identify native nodules in the three substrate CT examinations, after 13 additional radiologists each reviewed the images seven to eight times, some native nodules were identified. This observation is not surprising, given the well-documented

variability with which lung nodules are detected on CT images by observers (4,21) and challenges the establishment of an absolute reference standard for their presence (22). Eighth, all nodules were 5 mm in diameter and solid; thus, it is not possible to extrapolate these findings to subsolid or differently sized nodules. Finally, interpretation of 1-mm transverse sections is just one of several paradigms for lung nodule search in clinical use. Further investigations will be required to assess if alternative means of image display, including 5-mm transverse sections (3), multiplanar reformations (23), or thin-slab maximum intensity projections (24,25) might result in different search patterns or performance improvements. Computer-aided detection, which searches 100% of the lungs, has shown improvements to radiologists' detection rates (6,26). In an era of increasing volumetric spatial resolution and pressures to shorten interpretation times, image interpretation augmented by computer-based search could help overcome the shortcomings in lung nodule detection suggested by this study.

In summary, we have shown that synchronized recording of eye tracking and cine paging of transverse CT sections provides insights into the mechanism of detection of lung nodules, including the relationship between distance from the gaze point and nodule detection, the extent, duration, and effectiveness of search, and the differential effect of search and lesion recognition and acceptance on overall lung nodule detection. The observation that radiologists on average search only 26% of the lung parenchyma yet encompass 75% of nodules in their search volume is a particularly surprising result, and further investigation will be needed to develop a comprehensive model of cognition in lung nodule detection. Moreover, as momentum builds for widespread screening for lung cancer with CT, a greater understanding of the variations and limitations of radiologists' search and detection capabilities will be an important element in assuring consistently effective screening. At this time, we do not know if radiologists can be trained to enhance the effectiveness of their search or if there are specific search strategies or behaviors that enhance detection. Our results represent a new facet in our understanding of the interpretive process of CT images through the use of gaze tracking; however, further study of the relationship between search and CT interpretation is needed prior to deriving recommendations for improving performance.

Disclosures of Conflicts of Interest: G.D.R. Activities related to the present article: none to disclose. Activities not related to the present article: is on the GE Healthcare Comparative Effectiveness Research Advisory Board. Other relationships: none to disclose. **J.E.R.** Disclosed no relevant relationships. **M.T.** Disclosed no relevant relationships. **B.H.** Disclosed no relevant relationships. **S.B.** Disclosed no relevant relationships. **D.L.L.** Disclosed no relevant relationships. **D.M.S.** Disclosed no relevant relationships. **L.M.H.** Disclosed no relevant relationships. **S.N.** Activities related to the present article: none to disclose. Activities not related to the present article: is a consultant for Carestream, is on the scientific advisory board of Fovia and Echopixel. Other relationships: none to disclose. **K.R.C.** Disclosed no relevant relationships.

References

1. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of CT screening for lung cancer: a systematic review. JAMA 2012;307(22):2418–2429.
2. Kang MC, Kang CH, Lee HJ, Goo JM, Kim YT, Kim JH. Accuracy of 16-channel multidetector row chest computed tomography with thin sections in the detection of metastatic pulmonary nodules. Eur J Cardiothorac Surg 2008;33(3):473–479.
3. Ellis MC, Hessman CJ, Weerasinghe R, Schipper PH, Vetto JT. Comparison of pulmonary nodule detection rates between preoperative CT imaging and intraoperative lung palpation. Am J Surg 2011;201(5):619–622.
4. Rubin GD, Lyo JK, Paik DS, et al. Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. Radiology 2005;234(1):274–283.
5. Leader JK, Warfel TE, Fuhrman CR, et al. Pulmonary nodule detection with low-dose CT of the lung: agreement among radiologists. AJR Am J Roentgenol 2005;185(4):973–978.
6. Roos JE, Paik D, Olsen D, et al. Computer-aided detection (CAD) of lung nodules in CT scans: radiologist performance and reading time with incremental CAD assistance. Eur Radiol 2010;20(3):549–557.
7. Pinsky PF, Gierada DS, Nath PH, Kazerooni E, Amorosa J. National lung screening trial: variability in nodule detection rates in chest CT studies. Radiology 2013;268(3):865–873.
8. Kundel HL, Nodine CF, Carmody D. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Invest Radiol 1978;13(3):175–181.
9. Karadi C, Beaulieu CF, Jeffrey RB Jr, Paik DS, Napel S. Display modes for CT colonography. I. Synthesis and insertion of polyps into patient CT data. Radiology 1999;212(1):195–201.
10. Tall M, Choudhury KR, Napel S, Roos JE, Rubin GD. Accuracy of a remote eye tracker for radiologic observer studies: effects of calibration and recording environment. Acad Radiol 2012;19(2):196–202.
11. Gamer M, Hecht H. Are you looking at me? measuring the cone of gaze. J Exp Psychol Hum Percept Perform 2007;33(3):705–715.
12. Tourassi G, Voisin S, Paquit V, Krupinski E. Investigating the link between radiologists' gaze, diagnostic decision, and image content. J Am Med Inform Assoc 2013;20(6):1067–1075.
13. Krupinski EA. Visual scanning patterns of radiologists searching mammograms. Acad Radiol 1996;3(2):137–144.
14. Nodine CF, Liu H, Miller WT Jr, Kundel HL. Observer performance in the localization of tubes and catheters on digital chest images: the role of expertise and image enhancement. Acad Radiol 1996;3(10):834–841.
15. Kundel HL, Nodine CF, Toto L. Searching for lung nodules. the guidance of visual scanning. Invest Radiol 1991;26(9):777–781.
16. Matsumoto H, Terao Y, Yugeta A, et al. Where do neurologists look when viewing brain CT images? an eye-tracking study involving stroke cases. PLoS ONE 2011;6(12):e28928.
17. Ellis SM, Hu X, Dempere-Marco L, Yang GZ, Wells AU, Hansell DM. Thin-section CT of the lungs: eye-tracking analysis of the visual approach to reading tiled and stacked display formats. Eur J Radiol 2006;59(2):257–264.
18. Phillips P, Boone D, Mallett S, et al. Method for tracking eye gaze during interpretation of endoluminal 3D CT colonography: technical description and proposed metrics for analysis. Radiology 2013;267(3):924–931.
19. Tall M, Napel S, Schmitzberger FF, et al. Acquiring and visualizing volumetric eye gaze

paths through chest CT data [abstr]. In: Radiological Society of North America Scientific Assembly and Annual Meeting Program. Oak Brook, Ill: Radiological Society of North America, 2010; 394.

20. Drew T, Vo ML, Olwal A, Jacobson F, Seltzer SE, Wolfe JM. Scanners and drillers: characterizing expert visual search through volumetric images. J Vis 2013;13(10).

21. Armato SG 3rd, McNitt-Gray MF, Reeves AP, et al. The Lung Image Database Consortium (LIDC): an evaluation of radiologist variability in the identification of lung nodules on CT scans. Acad Radiol 2007;14(11): 1409–1421.

22. Bankier AA, Levine D, Halpern EF, Kressel HY. Consensus interpretation in imaging research: is there a better way? Radiology 2010;257(1):14–17.

23. Kwan SW, Partik BL, Zinck SE, et al. Primary interpretation of thoracic MDCT images using coronal reformations. AJR Am J Roentgenol 2005;185(6):1500–1508.

24. Napel S, Rubin GD, Jeffrey RB Jr. STS-MIP: a new reconstruction technique for CT of the chest. J Comput Assist Tomogr 1993;17(5): 832–838.

25. 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. AJR Am J Roentgenol 2002;179(1):149–157.

26. Marten K, Seyfarth T, Auer F, et al. Computer-assisted detection of pulmonary nodules: performance evaluation of an expert knowledge-based detection system in consensus reading with experienced and inexperienced chest radiologists. Eur Radiol 2004;14(10):1930–1938.