

Is It All in the Eye of the Beholder? Comparing Pulmonologists' and Radiologists' Performance

Drayton Hammond, PharmD, MBA, BCPS, BCCCP; and Nikhil Meena, MD

Experienced pulmonologists and radiologists were equally accurate in diagnosing malignant lung nodules and lung masses seen on low-dose computed tomography scans.

Lung cancer remains a leading cause of cancer-related deaths, and screening with low-dose computed tomography (LDCT) has the potential to decrease the mortality rate of patients by 20%.¹ Most major cancer societies have issued lung cancer screening recommendations. For example, the National Comprehensive Cancer Network recommends annual LDCT scans for high-risk patients (those at moderate or low risk need not be screened). High-risk patients are aged between 55 and 74 years (the U.S. Preventive Services Task Force upper age limit is 80 years) and have a smoking history of ≥ 30 pack-years, or if no longer smoking, a quit date within the past 15 years. Although length of screening needed is unclear, it is advised that patients have annual LDCT scans until they have been smoke free for 15 years, develop limited life expectancy, or are no longer eligible for definitive treatment for lung cancer. A

Dr. Hammond is a clinical assistant professor in the department of pharmacology, and **Dr. Meena** is an assistant professor in the department of internal medicine, both at the University of Arkansas for Medical Sciences in Little Rock. Dr. Meena also is a physician at Central Arkansas Veterans Healthcare System in Little Rock.

strong antismoking commitment and a multidisciplinary approach are of paramount importance.^{2,3}

Fleischner Society criteria are the most established guidelines for risk-stratifying pulmonary nodules (Table 1). Nodules are stratified by size and change in size over a 2-year period. There is interest in evaluating change in volume as well, but techniques are still emerging and have not been universally adopted.^{4,5}

Lung nodule screening likely will require significant involvement of radiologists and pulmonologists in the workup of patients with positive screens. Radiologists have demonstrated a fair amount of interobserver agreement with respect to diagnosis, but there are no data comparing pulmonologists with other pulmonologists or with radiologists.⁶⁻⁸ In addition, although health care professionals have access to validated models for predicting risk of malignancy, there is evidence they do not use them.^{9,10}

This study was conducted to determine whether pulmonologists and radiologists experienced in thoracic abnormalities are consistent in accurately diagnosing malignant lung nodules and masses noted on CT scans.

METHODS

After obtaining institutional review board approval for this study, the authors evaluated all the lung nodule or lung mass referrals that had been made to the University of Arkansas for Medical Sciences (UAMS) and Central Arkansas Veterans Healthcare System (CAVHS) interventional pulmonary clinics between March 2009 and March 2013. Of the 1,512 referrals made, 250 were randomly selected for noncontrasted CT image review and data collection. Each CT image was de-identified and then reviewed by a pulmonologist and a radiologist. The study used 4 reviewers—2 board-certified pulmonologists and 2 board-certified radiologists—all with > 3 years of experience. Both radiologists were thoracic specialists, and no residents or fellows participated. For each case, reviewers were given a brief patient history outlining smoking and other malignancies. Data collected included age, sex, race, exposure to cigarette smoking, and the gold standard of final diagnosis (FD).

In each case, a pulmonologist and a radiologist reviewed the patient's CT images from the first visit. Reviewers were asked to

Table 1. Fleischner Society Guidelines for Following Up Newly Detected Pulmonary Nodules in Patients Aged ≥ 35 years^a

Nodule Size, mm ^b	Low-Risk Patient ^c	High-Risk Patient ^d
≤ 4	No follow-up needed ^e	Follow-up CT at 12 mo; if unchanged, no further follow-up ^f
> 4 to ≤ 6	Follow-up CT at 12 mo; if unchanged, no further follow-up ^f	Follow-up CT at 6-12 mo; if unchanged, next follow-up at 18-24 mo ^f
> 6 to ≤ 8	Follow-up CT at 6-12 mo; if unchanged, next follow-up at 18-24 mo	Follow-up CT at 3-6 mo; if unchanged, next follow-ups at 9-12 and 24 mo
> 8	Follow-up CT around 3, 9, and 24 mo; dynamic contrast-enhanced CT, PET, and/or biopsy	Same as for low-risk patient

Abbreviations: CT, computed tomography; PET, positron emission tomography.
^aReprinted with permission from the Radiological Society of North America.¹⁴
^bMean length or width.
^cMinimal or absent history of smoking and other known risk factors.
^dHistory of smoking and presence of other risk factors.
^eRisk of malignancy is very low (< 1%) in this group.
^fNon-solid (ground-glass) opacities may require longer follow-up to rule out indolent malignancies.⁴

determine and document the single most likely diagnosis. Diagnoses were grouped into primary lung cancer, metastatic disease, lymphoma, infectious/inflammatory etiology, benign neoplasm, and other (eg, sarcoma). A lesion with a diagnostic biopsy and stability at 2 years was deemed benign. A lesion that was culture-positive or responded rapidly to antibacterial or antifungal therapy was deemed infectious/inflammatory. Lesions were grouped by size: group 1 (≤ 10 mm), group 2 (11-30 mm), group 3 (31-50 mm), group 4 (≥ 51 mm).

Statistical Analyses

Student *t* tests were used to compare means. Concordance of the pulmonary reviewers and FD was assessed with the κ coefficient. The concor-

dance was also evaluated between the radiology reviewers and FD. These statistical analyses were performed with SAS Version 9.4 (SAS Institute). *P* values were interpreted using the sliding-scale approach of Mendenhall and colleagues: *P* < .01 (highly significant); .01 < *P* < .05 (statistically significant); .05 < *P* < .10 (trending toward significance); *P* > .10 (not significant).¹¹

RESULTS

Of the 250 patients selected for the study, 111 had the pertinent

Table 2. Distribution of Diagnosis by Size and Suspicion: Pulmonology and Radiology Reviewers and Final Diagnosis

Diagnosis by Reviewer	Size-Based Group ^a				Overall (%)
	1	2	3	4	
Pulmonologist					
Primary lung cancer	18	24	2	24	68 (61)
Metastatic disease	7	7	2	3	19 (17)
Lymphoma	0	4	0	4	8 (7)
Infectious/inflammatory	6	4	1	0	11 (9)
Benign neoplasm	1	1	0	1	3 (3)
Other (eg, sarcoma)	0	0	0	2	2 (2)
Radiologist					
Primary lung cancer	18	26	3	25	72 (65)
Metastatic disease	5	6	1	3	15 (14)
Lymphoma	2	4	0	3	9 (8)
Infectious/inflammatory	4	2	1	2	9 (8)
Benign neoplasm	3	1	0	0	4 (4)
Other (eg, sarcoma)	0	1	—	—	—
Final diagnosis					
Primary lung cancer	16	21	2	21	60 (54)
Metastatic disease	3	8	1	3	15 (14)
Lymphoma	0	3	0	4	7 (6)
Infectious/inflammatory	6	5	1	3	15 (13)
Benign neoplasm	7	2	0	2	11 (9)
Other (eg, sarcoma)	0	1	1	1	3 (3)

^aGroup 1 (≤ 10 mm), group 2 (11-30 mm), group 3 (31-50 mm), group 4 (≥ 51 mm).

data available, along with a follow-up appointment > 2 years afterward at the center. The patients included 40 women and 71 men; 79 white patients, 29 black patients, and 3 patients of other races. Mean age was 58 years (range, 21-93 years).

Risk factors for malignancy were older age, larger lesion, and history of smoking. The malignancy rates for women and men were almost identical (53% and 54%, respectively), and the difference was not statistically significant ($P = .40$).

Diagnosis

Table 2 outlines the distribution of the reviewers' diagnoses and the distribution of FD. Primary lung cancer was the dominant suspected diagnosis and accounted for 61%, 65%, and 54% of the cases reviewed by the pulmonologist, the radiologist, and FD, respectively. Metastatic disease was a distant second dominant diagnosis (17%, 15%, and 15%, respectively). There was no statistical difference between the reviews of the pulmonologist and radiologist, and the FD ($P > .05$).

Table 3 lists the κ results for the strength of agreement between pulmonologist and radiologist. Agreement for primary lung cancer was very good: 0.94 (95% confidence interval [CI], 0.89-0.99). With respect to group 1, agreement was perfect: 1.0 (95% CI, 1.000-1.000). Benign neoplasm had the weakest agreement. There was no statistical difference between pulmonologist and radiologist determinations across size-based groups.

Agreement between pulmonologist and FD was almost perfect. The major discrepancy between the sets of reviewers remained benign neoplasm and infectious/inflammatory etiology.

Table 3. κ Scores for Pulmonology and Radiology Reviewers^a

Diagnosis by Size-Based Group ^b	κ	SE for κ	95% CI	Agreement Strength
Group 1				
Primary lung cancer	1.00	0	1.00 to 1.00	Perfect
Metastatic disease	0.71	0.2	0.37 to 1.00	Good
Infectious/inflammatory	0.67	0.2	0.27 to 1.00	Good
Benign neoplasm	0.14	0.14	-0.003 to 0.55	Poor
Group 2				
Primary lung cancer	0.88	0.1	0.74 to 1.00	Very good
Metastatic disease	0.86	0.13	0.59 to 1.00	Very good
Lymphoma	1.00	0	1.00 to 1.00	Perfect
Benign neoplasm	0.5	0.38	-0.23 to 1.00	Moderate
Group 3				
Primary lung	1.00	0	1.00 to 1.00	Perfect
Group 4				
Primary lung cancer	0.88	0.07	0.74 to 1.00	Very good
Metastatic disease	1.00	0	1.00 to 1.00	Perfect
Lymphoma	0.75	0.22	0.31 to 1.00	Good
Overall				
Primary lung cancer	0.94	0.03	0.89 to 0.99	Very good
Metastatic disease	0.79	0.1	0.6 to 0.98	Good
Lymphoma	0.89	0.1	0.67 to 1.00	Very good
Infectious/inflammatory	0.82	0.12	0.58 to 1.00	Very good
Benign neoplasm	0.75	0.22	0.3 to 1.00	Good

Abbreviations: CI, confidence interval; SE, standard error.

^aTests were performed for more than 4 observations.

^bGroup 1 (≤ 10 mm), group 2 (11-30 mm), group 3 (31-50 mm), group 4 (≥ 51 mm).

Of the 111 study patients, 68 (61%) and 72 (65%) were suspected of having primary lung cancer by pulmonologist and radiologist, respectively. However, only 60 (54%) actually had primary lung cancer; the

differences were not statistically significant ($P = .27$ and $.1$, respectively). No cases were reclassified as primary lung cancer on final pathology.

Infectious/inflammatory etiologies did not always have positive

Table 4. Specificity and Sensitivity of Radiologic Examination (CI)

Diagnosis	Specificity		Sensitivity		Positive	Predictive Value
	Pulmonologist	Radiologist	Pulmonologist	Radiologist	Pulmonologist	Radiologist
Malignancy	0.84 (0.77-0.84)	0.77 (0.69-0.77)	1.00 (0.94-1.00)	1.00 (0.94-1.00)	0.89 (0.83-0.88)	0.83 (0.78-0.83)
Metastatic disease	0.96 (0.92-0.96)	1.00 (0.97-1.00)	1.00 (0.79-1.00)	1.00 (0.82-1.00)	0.76 (0.63-0.79)	1.00 (0.82-1.00)
Lymphoma	0.99 (0.96-0.98)	0.98 (0.95-0.98)	1.00 (0.62-1.00)	1.00 (0.69-1.00)	0.89 (0.6-0.89)	0.81 (0.56-0.81)
Infectious/ inflammatory	0.73 (0.52-0.73)	0.6 (0.4-0.6)	1.00 (0.62-1.00)	1.00 (0.69-1.00)	1.00 (0.73-1.00)	1.00 (0.66-1.00)
Benign neoplasm	1.00 (0.98-1.00)	1.00 (0.97-1.00)	0.27 (0.08-0.27)	0.33 (0.13-0.33)	1.00 (0.3-1.00)	1.00 (0.41-1.00)

cultures. Those with positive cultures included *Streptococcus (S) viridans*, *Rhodococcus equi*, *Blastomyces dermatitidis*, *S constellatus*, *S anginosus*, *S intermedius*, and *Histoplasma capsulatum*. Benign neoplasms included radiation injuries, benign fibrous tumor of the pleura, and hamartoma.

Pulmonologists and radiologists had identical high sensitivities for primary lung cancer: 1.0 (95% CI, 0.94-1.00). Specificities were 0.84 (95% CI, 0.77-0.84) for pulmonologists and 0.77 (95% CI, 0.69-0.77) for radiologists, and the difference was not statistically significant ($P = .28$) (Table 4).

DISCUSSION

Computed tomography scans are performed to evaluate a variety of diseases. An estimated 7 million CT scans are performed in the U.S. annually.^{6,12} As the National Lung Screening Trial recommendations are followed more routinely, almost 9 million people could become candidates, adding to the already large number of CT scans to be evaluated.¹³

Radiologists would understandably read most of these patients' scans. However, patients referred to tertiary-care centers usually bring CT images with them; even scans performed at UAMS and CAVHS centers may not be read by a radiologist in time for an appointment. The result is that the clinic pulmonologist often must base decisions on a CT reading, but without the assistance of high-fidelity computer programs or a high-definition scan.⁵ These limitations indicate why it is important to know whether assessment by a pulmonologist compares favorably with assessment by a radiologist and with the eventual diagnosis.

The malignancy rate in the referred population is not insignificant. Halbert and colleagues found a 25% malignancy rate in their study,¹² and the present study had an overall malignancy rate of 54%. The difference may be attributed to the possibility that the patients may have been prescreened prior to referral.

The reviewers overestimated the presence of malignant disease, though not to a level of statistical significance.

About 88% of cases evaluated by a pulmonologist and 83% of cases evaluated by a radiologist were confirmed to be malignant. The reviewers' sensitivity was perfect for all diagnoses except benign neoplasms, likely because these cases were classified malignant, thus increasing sensitivity but decreasing specificity.

This dynamic is important to understand, as it allows for a very high negative predictive value, which has real implications for resource management at VA hospitals, including CAVHS facility, where almost every CT scan with an abnormality is referred for pulmonologist consultation. In these cases, the radiologist not only lists the likely suspicion but includes a recommendation for follow-up or further workup based on Fleischner Society guidelines.^{4,14} The patient should be informed of findings as soon as the radiologist reads the CT scan, and a plan should be made on the basis of the recommendation. The patient should not have to unnecessarily wait—a potential source of anxiety—to see another specialist who would probably make

Negative Predictive Value	
Pulmonologist	Radiologist
1.00 (0.92-1.00)	1.00 (0.9-1.00)
1.00 (0.97-1.00)	1.00 (0.97-1.00)
1.00 (0.97-1.00)	1.00 (0.97-1.00)
0.96 (0.93-0.96)	0.94 (0.91-0.94)
0.92 (0.9-0.92)	0.92 (0.9-0.94)

the same recommendation.

Applying this study's findings could improve workflow and the timing of CT scans. A patient should not be referred to a pulmonologist unless specifically recommended by a radiologist, thus decreasing the scheduling burden on the specialty clinic and allowing for appropriate patients to be scheduled at reasonable intervals. In addition, having only 1 person in charge of ordering CT scans could reduce the chance of duplicating orders and performing CT scans at inappropriate times.

Most important, these results should lead to more detailed physician-patient discussions about radiologic findings, hopefully alleviating any patient anxiety. A patient who still wants to see a specialist may, but with less stress that can accompany being told that there is "something abnormal" on the imaging and that the patient needs to see a lung doctor.

Limitations

This study had a few weaknesses. It was a small trial, and its data were

collected retrospectively. In addition, generalizing its results may be difficult, as its reviewers had less than 5 years of training, and reviewers with more experience likely would be more accurate and have a higher rate of agreement.

Results could have been skewed by the study's unusually large number of patients with malignant disease. Had the study been conducted with a larger population (patients at primary care offices), accuracy and agreement might have been lower.

CONCLUSION

This study answered its 2 questions. Although it is universally accepted that pulmonologists can review patients' scans, to the authors' knowledge this is the first study that asked, "Are pulmonologists as good as radiologists in reading CT scans?" The answer is yes. Also asked was, "Do pulmonologists' and radiologists' diagnoses predict the final path?" The reviewers' were very accurate except in the case of benign neoplasms.

Experienced pulmonologists and radiologists are consistent in accurately diagnosing malignant lung nodules and lung masses noted on CT scans. ●

Author disclosures

The authors report no actual or potential conflicts of interest with regard to this article.

Disclaimer

The opinions expressed herein are those of the authors and do not necessarily reflect those of Federal Practitioner, Frontline Medical Communications Inc., the U.S. Government, or any of its agencies. This article may discuss unlabeled or investigational use of certain drugs.

Please review complete prescribing information for specific drugs or drug combinations—including indications, contraindications, warnings, and adverse effects—before administering pharmacologic therapy to patients.

REFERENCES

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395-409.
2. Wood DE. National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines for Lung Cancer Screening. *Thorac Surg Clin*. 2015;25(2):185-197.
3. Humphrey LL, Deffebach M, Pappas M, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive Services task force recommendation. *Ann Intern Med*. 2013;159(6):411-420.
4. Naidich DP, Bankier AA, MacMahon H, et al. Recommendations for the management of sub-solid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology*. 2013;266(1):304-317.
5. Mehta HJ, Ravenel JG, Shaftman SR, et al. The utility of nodule volume in the context of malignancy prediction for small pulmonary nodules. *Chest*. 2014;145(3):464-472.
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(1):265-272.
7. McCarville MB, Lederman HM, Santana VM, et al. Distinguishing benign from malignant pulmonary nodules with helical chest CT in children with malignant solid tumors. *Radiology*. 2006;239(2):514-520.
8. 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(8):948-956.
9. Schultz EM, Sanders GD, Trotter PR, et al. Validation of two models to estimate the probability of malignancy in patients with solitary pulmonary nodules. *Thorax*. 2008;63(4):335-341.
10. Tanner NT, Aggarwal J, Gould MK, et al. Management of pulmonary nodules by community pulmonologists: a multicenter observational study. *Chest*. 2015;148(6):1405-1414.
11. Mendenhall W, Beaver RJ, Beaver BM. *Introduction to Probability and Statistics*. 13th ed. Belmont, CA: Brooks/Cole, Cengage Learning; 2009.
12. Halbert CL, Madtes DK, Vaughan AE, et al. Expression of human alpha1-antitrypsin in mice and dogs following AAV6 vector-mediated gene transfer to the lungs. *Mol Ther*. 2010;18(6):1165-1172.
13. Ma J, Ward EM, Smith R, Jemal A. Annual number of lung cancer deaths potentially avertable by screening in the United States. *Cancer*. 2013;119(7):1381-1385.
14. MacMahon H, Austin JH, Gamsu G, et al; Fleischner Society. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology*. 2005;237(2):395-400.