Published in final edited form as:

J Thorac Imaging. 2011 February ; 26(1): 48–53. doi:10.1097/RTI.0b013e3181d73a8f.

Influence of Nodule Detection Software on Radiologists' Confidence in Identifying Pulmonary Nodules With Computed Tomography

Paul J. Nietert, PhD^{*}, James G. Ravenel, MD[†], Katherine K. Taylor, MS[‡], and Gerard A. Silvestri, MD[‡]

^{*} Department of Medicine, Division of Biostatistics and Epidemiology, Medical University of South Carolina, Charleston, SC

[†] Department of Radiology and Radiological Science, Medical University of South Carolina, Charleston, SC

[‡] Division of Pulmonary, Critical Care, Allergy and Sleep Medicine, Medical University of South Carolina, Charleston, SC

Abstract

Purpose—With advances in technology, detection of small pulmonary nodules is increasing. Nodule detection software (NDS) has been developed to assist radiologists with pulmonary nodule diagnosis. Although it may increase sensitivity for small nodules, often there is an accompanying increase in false-positive findings. We designed a study to examine the extent to which computed tomography (CT) NDS influences the confidence of radiologists in identifying small pulmonary nodules.

Materials and Methods—Eight radiologists (readers) with different levels of experience examined thoracic CT scans of 131 cases and identified all the clinically relevant pulmonary nodules. The reference standard was established by an expert, dedicated thoracic radiologist. For each nodule, the readers recorded nodule size, density, location, and confidence level. Two weeks (or more) later, the readers reinterpreted the same scans; however, this time they were provided marks, when present, as indicated by NDS and asked to reassess their level of confidence. The effect of NDS on changes in reader confidence was assessed using multivariable generalized linear regression models.

Results—A total of 327 unique nodules were identified. Declines in confidence were significantly ($P<0.05$) associated with the absence of an NDS mark and smaller nodules (odds ratio=71.0, 95% confidence interval =14.8–339.7). Among nodules with pre-NDS confidence less than 100%, increases in confidence were significantly ($P<0.05$) associated with the presence of an NDS mark (odds ratio=6.0, 95% confidence interval =2.7–13.6) and larger nodules. Secondary findings showed that NDS did not improve reader diagnostic accuracy.

Conclusion—Although in this study NDS does not seem to enhance reader accuracy, the confidence of the radiologists in identifying small pulmonary nodules with CT is greatly influenced by NDS.

Keywords

clinical decision making; computed tomography scan; diagnostic imaging; lung neoplasm; diagnostic errors

Due to improvements in computed tomography (CT) imaging technology, detection of small pulmonary nodules is increasing.¹ The detection of small nodules is also of paramount importance in finding early-stage lung cancer. However, nodules less than 1 cm in diameter often pose a dilemma to clinicians and patients, as these nodules may be difficult to biopsy and may be easily confused with normal anatomic structures within the lung. In addition, alternative imaging techniques may not improve diagnostic yield. These uncertainties may have important clinical implications, as decision-making about patient management is highly dependent on the anticipated course of disease.

To assist radiologists with pulmonary nodule detection, a CT nodule detection software (NDS) that uses computer algorithms in conjunction with CT imaging to identify potential nodules has been developed. With increasing CT imaging being performed for diagnostic and/or screening purposes, the use of NDS to assist radiologists in identifying potentially malignant nodules is also likely to rise. Although many studies have shown that additional nodules will be detected by NDS systems, these same systems will occasionally mark normal anatomic structures as a nodule (ie, false positive).² Conversely, the NDS software could also potentially miss a nodule that is present (ie, false negative). The overall effect of an NDS system may, therefore, be beneficial if it leads to more accurate detection. However, the overall effect may be detrimental if radiologists accept false-positive marks as true-positive, ultimately leading to unnecessary follow-up imaging, or if the lack of NDS marks causes readers to ignore truly positive nodules. We designed a study to examine the extent to which NDS influences the confidence of the radiologists in identifying small pulmonary nodules. Secondary aims focused on determining whether the influence of NDS on reader confidence was significantly associated with nodule characteristics (ie, size, location, density) or reader experience and whether the use of NDS improved reader accuracy.

MATERIALS AND METHODS

This study was approved with a waiver of consent from our local institutional review board, the Human Research Protections Office of the U.S. Army Medical Research and Materiel Command, and was performed in compliance with the Health Insurance Portability and Accountability Act regulations.

A total of 131 patients with thoracic CT scans were selected by a dedicated thoracic radiologist (blind) with 10 years of clinical experience with patients who had been referred for a CT of the thorax for various clinical reasons. The patients were selected such that roughly half would have at least 1 definite nodule (nodule defined prospectively at the clinical interpretation). All studies were performed on a 16-multidetector CT scanner (GE, Fairfield, CT) with a 16×1.25 -mm detector array, 0.5-second tube rotation, 120 kEv and variable mA using z-axis tube current modulation with a maximum of 440 mA. Images were reconstructed into 2.5-mm section thickness at 2.5-mm intervals using both mediastinal and detail algorithms. Eight radiologists (readers) with different levels of experience (3 radiology residents, 3 general radiologists, and 2 senior chest faculty) were then asked to independently examine the CT scans and identify all pulmonary nodules that they deemed were "clinically relevant" (ie, any nodules that the reader believed would warrant additional followup). The expert had 10 years experience and had completed a thoracic imaging fellowship, whereas the 2 senior chest faculty members had 10 and 30 years of practice, but

no formal thoracic imaging fellowship. For each of these suspected nodules, the reader recorded the case and slice number on which the suspected nodule was identified, their manual computer-assisted measurement of nodule diameter, and their subjective assessment of density (solid/part solid/ground glass), laterality (left/right), lobe (lower/middle/upper), and confidence (1% to 33%, 34% to 66%, 67% to 99%, 100%) associated with their belief that the suspected nodule was a true nodule. Once the scans were read by all readers, a single expert thoracic radiologist unblinded to all reader-detected nodules reviewed the scans (with any accompanying NDS marks) to identify all nodules ≥ 4 mm in diameter. No additional nodules ≥ 4 mm were identified by the expert. Nodules prospectively detected by 2 or more readers were considered true positive. In addition, the expert used his judgment to classify all identified nodules ≥ 4 mm in diameter identified by a single reader as “true positives” or “false positives,” and this adjudication became the reference standard.

Approximately 2 weeks after the initial reading, each reader reevaluated the same scans unblinded to the nodules they detected during the original reading; however, this time they were provided marks, when present, as indicated by an NDS program, GE Digital Contrast Agent (GE Healthcare, Waukesha, WI), a commercially available software program to assist in nodule detection. Thus, during the second reading, NDS was used as a second reader. The NDS software was used in conjunction with the same CT images that the readers had assessed earlier. Readers were allowed to view their initial confidence level assignment, and during the second reading, readers were asked to reassess their level of confidence, which included the ability to raise, lower, or leave unchanged their initial level of confidence for suspected nodules. Each reader was blinded to the findings from other readers. Their first-set readings (without assistance from NDS) were referred to as pre-NDS assessments, and the second readings were referred to as post-NDS assessments. During the second reading, readers were given the option of assigning a confidence level of 0%, indicating that they now had no confidence that a nodule identified earlier was a true nodule. During the second reading, readers also assessed any nodules detected by NDS but not detected during their pre-NDS readings. Although readers were not blinded to their initial assessments, they were blinded to the underlying nodule prevalence among the selected cases.

Statistical Analyses

Initially, statistical comparisons were made with suspected nodule assessments as the unit of analysis to elicit differences between suspected nodules with and without NDS marks. These 2 groups of nodules were compared with respect to nodule characteristics and pre/post-NDS reader confidence levels. The analyses were conducted using generalized linear mixed regression models to control reader variability and clustering of nodules within cases.

The primary analyses examining the effect of NDS on changes in reader confidence also incorporated generalized linear mixed regression models. Two models were developed using PROC GLIMMIX within SAS (Cary, NC) v9.2. The first model included all identified nodules and was created to determine factors associated with a reader decreasing his/her confidence level. The second model, limited to suspected nodules that readers initially assigned a pre-NDS confidence level less than 100%, was used to examine factors associated with a reader increasing his/her confidence level. In both models, only nodules that were originally identified by the readers were included, and the key independent variable of interest was an indicator for whether an NDS mark was present on/around that nodule image; covariates included reader's experience and nodule characteristics (ie, size, location, density). Both models contained random effects for cases and readers with a variance component covariance structure, to control for reader variability and clustering of nodules within cases. The primary analyses were restricted to nodules ≥ 4 mm in diameter; however, secondary analyses also considered the effects of NDS on nodules of all sizes.

Additional secondary analyses focused on determining whether the NDS markings improved reader accuracy in identifying nodules ≥ 4 mm in diameter. For this purpose, all nodules identified by readers but not by the expert were labeled as false positives, and all nodules identified by readers and by the expert were labeled as true positives. Nodules identified by the expert but not by the readers were labeled as false negatives. In this setting in which readers were asked to identify nodules, the term true negative is not practically measurable; thus, the measures of accuracy included nodule sensitivity and positive predictive value, and both measures were compared pre-NDS and post-NDS. For a given reader's accuracy calculations, the analyses included nodules originally identified by the reader (which may have had an NDS mark) and nodules identified by NDS but not by the reader.

RESULTS

From the 131 cases, 2912 suspected nodules of any size were identified and assessed by a group of 8 readers, with individual readers identifying an average of 364 suspected nodules (range, 127 to 971 nodules per reader) of any size, or an average of 2.8 nodules per reader per case (range, 1.0 to 7.4). A total of 327 unique nodules ≥ 4 mm in diameter were identified by the readers, 180 of which were confirmed as true-positive nodules by the expert. Of the 131 cases, 73 (55.7%) were considered by the expert to have at least 1 pulmonary nodule ≥ 4 mm in diameter. A total of 196 (59.9%) of the 327 unique nodules ≥ 4 mm in diameter were identified by multiple readers, whereas 131 (40.1%) of these 327 nodules were identified by only 1 reader (but not necessarily the same reader). The NDS identified 143 additional suspected nodules, none of which was deemed by the expert to be a true nodule ≥ 4 mm. Among these 143 nodules, post-NDS confidence levels of 100% were reported 11.2% of the time, whereas 82.7% of the time readers reported a 0% confidence that the suspected nodules were true nodules.

Descriptive statistics for reader-suspected nodules and comparisons between nodules with and without NDS marks are summarized in Table 1. After adjustment for within-reader and within-case clustering, nodules identified in the presence of NDS marks were significantly larger in diameter, more likely to be identified in the left lung, more likely to be identified in the lower portion of the lung, less likely to be identified in the middle portion of the lung, and more likely to be labeled as solid than nodules with NDS marks absent.

Table 2 displays the raw (unadjusted) data summarizing the influence of NDS on reader confidence, stratified by nodule size, pre-NDS confidence level, and presence/absence of NDS mark. NDS marks were not present for 2563 of 2912 (88%) suspected nodule assessments, and NDS marks were present for 349 (12%) suspected nodule assessments. On average, confidence levels tended to increase when readers were allowed to review their nodule assessments, whether or not NDS marks were present. However, when a reader's confidence declined, it was almost always accompanied by the absence of an NDS mark. Among the assessments for nodules < 4 mm, reader confidence declined in a similar manner, regardless of whether a NDS mark was absent (declining in 4.9% of 1315 assessments) or present (declining in 6.5% of 31 assessments). However, among the suspected nodules ≥ 4 mm, reader confidence declined much more often when NDS marks were absent (declining in 5.9% of 1248 assessments) than when NDS marks were present (declining in 0.3% of 318 assessments). Reader confidence increased much more often when NDS marks were present than when they were absent, both in nodules < 4 mm (increasing in 75.0% of 8 assessments versus 21.2% of 407 assessments, respectively) and in nodules ≥ 4 mm (increasing in 46.5% of 43 assessments versus 10.3% of 466 assessments, respectively).

Details of the primary multivariable models for nodules ≥ 4 mm are presented in Table 3. In the multivariable model examining factors associated with confidence decline ($n=1566$

relevant assessments), the absence of an NDS mark was highly and significantly associated with the odds of a decline in confidence [odds ratio (OR)=71.0, 95% confidence interval (CI)=14.8–339.7]. Confidence decline was not significantly associated with location, or density, or with the reader's level of experience; however, confidence decline was significantly more likely in nodules 4.0 to 5.9mm (OR=4.7, 95% CI=1.0–21.6) and 6.0 to 7.9mm (OR=5.3, 95% CI=1.1–25.7) than in nodules ≥ 10 mm in diameter. In the multivariable model examining factors associated with confidence increase (n=509 relevant assessments), the presence of an NDS mark was associated with a 6-fold increase in the odds of recording a higher post-NDS confidence (OR=6.0, 95% CI=2.7–13.6). Increase in confidence was not significantly associated with location, or density, or with the reader's level of experience; however, confidence increases were more likely in nodules 6.0 to 7.9mm (OR= 2.4, 95% CI=1.3–4.4) and nodules ≥ 10 mm (OR=5.0, 95% CI=1.6–15.6) compared with nodules 4.0 to 5.9mm in diameter.

Despite the NDS marks having a significant influence on reader confidence, secondary findings indicated that the use of NDS did not improve reader accuracy. The mean sensitivity for identifying nodules ≥ 4 mm in diameter across all readers was relatively unchanged pre-NDS and post-NDS [pre-NDS mean (SD): 52.3% (13.0%); post-NDS means (SD): 53.8% (13.0%); mean (SD) increase: 1.53% (0.8%)]; as was the mean positive predictive value [pre-NDS mean (SD): 71.8% (11.0%); post-NDS mean (SD): 72.7% (11.2%); mean (SD) increase: 0.9% (0.7%)]; The mean false-positive rate did not change significantly [pre-NDS mean (SD): 28.2% (11.0%); post-NDS mean (SD): 27.3% (11.2%); mean (SD) decrease: 0.9% (0.7%)]. In situations in which NDS marks were absent for truly positive nodules, readers decreased their confidence levels an average (SD) of 9.8% (15.6%) of the time. No readers ever accepted a false-positive NDS mark for a nodule that he/she had not initially identified. However, in situations in which false-positive NDS marks were present for nodules that initially had been identified, reader confidence increased an average (SD) of 10.3% (11.6%) of the time.

DISCUSSION

This project has several important findings regarding NDS and radiologist confidence level. First, although confidence levels tended overall to increase when readers were allowed to review their nodule assessments, the absence of an NDS mark was strongly associated with a radiologist being less confident in his/her assessment of a suspected nodule. Second, the presence of an NDS mark was strongly associated with a radiologist being more confident in his/her assessment of a suspected nodule. Declines in reader confidence were significantly more likely to occur among smaller nodules, whereas increases in reader confidence were significantly more likely to occur among larger nodules. Other secondary findings include that neither reader experience nor any of the other nodule characteristics we examined (ie, all except nodule size) were significantly associated with declines or increases in reader confidence levels. In addition, the use of NDS does not result in improved accuracy in detecting pulmonary nodules ≥ 4 mm in diameter.

NDS has been shown in numerous studies to increase radiologist sensitivity in the detection of pulmonary nodules, often with a concomitant increase in the number of false-positive results.^{2–6} Across these studies, improvement in reader sensitivity ranges from 2% to 40%, with the greatest improvement showed by Hirose et al.² In their study, more than 50% of the nodules were less than 4mm in size. As nodule size increases, the relative contribution of NDS to sensitivity diminishes. For example, a recent study by White et al.⁴ included nodules across a wider range of diameters (4 to 30mm) and showed a range in the change in the area under the receiver operating characteristic curve from 0.0% to 5.2%. The magnitude of this overall modest improvement in individual reader sensitivity for nodules >4 mm with NDS is

similar to that found in our study. In our study, it is important to note that NDS failed to identify a single additional true-positive nodule ≥ 4 mm beyond those identified by all the readers in aggregate, which may explain why we observed minimal improvement in reader accuracy when NDS was used.

Other studies have examined NDS and reader confidence for identifying pulmonary nodules. In a study by Rubin et al,⁵ comparing readers to NDS, they asked readers to assess their initial confidence levels; however, they did not assess post-NDS confidence. Our study seems to be the first that examines the influence of NDS on reader confidence in the assessment of small pulmonary nodules. We found that confidence was most likely to be increased by NDS detection and that our readers were more apt to accept their findings even when dismissed by the NDS. In this regard, the influence of NDS may have less influence on the final impression, similar to findings of Li et al,⁷ who showed that NDS only resulted in radiologists changing their recommendations on 18% of 896 observations, with NDS having a beneficial effect for 68% of the cases with a changed recommendation. However, the nodules assessed were much larger than ours, as 28 nodules examined in their study were all larger than 6 mm, with 19 being greater than 10mm in diameter. Studies of the effects of NDS prompts on reader confidence have also been conducted in breast cancer imaging. For example, in one study, NDS prompts were shown to do a better job of improving reader accuracy in the context of nodules initially assigned low confidence scores than in the context of nodules initially assigned high confidence scores.⁸

Our findings have significant implications regarding the use of NDS in detecting small pulmonary nodules with CT. Readers are clearly influenced by NDS when assessing these types of nodules. If the lack of NDS marks introduces more uncertainty, radiologists may find it even more difficult to diagnose patients and follow/manage suspected small pulmonary nodules over time. If the presence of NDS marks incorrectly results in radiologists being “overconfident,” patients may ultimately be subjected to unnecessary follow-up, procedures/surgeries, and increased healthcare costs. This can be seen by the acceptance of 11% of false-positive detections by the readers. For NDS systems to be more readily accepted in clinical practice, improvement is needed in their accuracy (eg, sensitivity, specificity, and positive/negative predictive value), so that they do not unnecessarily influence the radiologist’s judgment. These sentiments have also been argued in earlier studies of radiologist confidence in computer-aided detection in mammography.^{9,10} It should also be noted that NDS may require additional computational resources, including thin collimation, and using it with small nodules may slow image processing and reading time and require additional data storage capabilities; however, such limitations are likely to be lessened over time as computer technologies continue to improve. Finally, the full medical-legal implications of using NDS in light of its effects on reader confidence and accuracy remain unknown.

There are several limitations of our study. Although we had initially hoped that including 8 radiologists with different levels of experience would allow us to determine whether NDS had a greater influence on less-experienced readers, high correlations among the within-reader nodules decreased our effective statistical power to do so. Future studies examining such a phenomenon would need to include more readers with different levels of experience. The secondary analyses examining reader accuracy used a reference standard based on the assessments of one dedicated thoracic radiologist. Even experienced thoracic radiologists exhibit substantial variability in lung nodule identification.^{2,7,11–13} Having a different reference standard might have resulted in different findings; however, given the variability across the other readers in this study, it is likely that any changes in findings in these secondary analyses would have been minimal. Like most other studies of NDS, the prevalence of nodules was higher than would be expected in the general population. Finally,

because we used a single NDS program, our findings are not necessarily generalizable to other similar programs.

A key strength of our study is the fact that we included a large number of cases (n=131), substantially more than have been traditionally examined in multireader imaging studies.^{2,4,7,11–13} In addition, by including multiple readers with different levels of experience, we have bolstered the external validity of our findings. In other words, if all our readers had been residents or all had been senior chest faculty, the generalizability would certainly have been lower. Another important characteristic of this study is that by using multivariable statistical models, we were able to examine the independent effects of NDS software on confidence and reader accuracy, while adjusting for a variety of potential confounding factors.

This study shows the need for radiologists to be cautious when using NDS in conjunction with CT to evaluate small pulmonary nodules. Further research is needed to identify objective measures and nodule characteristics that can help predict which nodules will become malignant. Such research could help standardize small pulmonary nodule assessment, minimizing interreader and intrareader variability, increasing radiologists' confidence in their assessments, and increasing diagnostic accuracy.

Acknowledgments

Supported by a grant from the Department of Defense (W81XWH-05-1-0378; PI: Silvestri).

References

1. Shiau MC, Bonavita J, Naidich DP. Adenocarcinoma of the lung: current concepts in radiologic diagnosis and management. *Curr Opin Pulm Med*. 2007; 13:261–266. [PubMed: 17534170]
2. Hirose T, Nitta N, Shiraishi J, et al. Evaluation of computer-aided diagnosis (CAD) software for the detection of lung nodules on multidetector row computed tomography (MDCT): JAFROC study for the improvement in radiologists' diagnostic accuracy. *Acad Radiol*. 2008; 15:1505–1512. [PubMed: 19000867]
3. Awai K, Murao K, Ozawa A, et al. Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. *Radiology*. 2004; 230:347–352. [PubMed: 14752180]
4. White CS, Pugatch R, Koonce T, et al. Lung nodule CAD software as a second reader: a multicenter study. *Acad Radiol*. 2008; 15:326–333. [PubMed: 18280930]
5. 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; 234:274–283. [PubMed: 15537839]
6. Das M, Muhlenbruch G, Mahnken AH, et al. Small pulmonary nodules: effect of two computer-aided detection systems on radiologist performance. *Radiology*. 2006; 241:564–571. [PubMed: 17057074]
7. Li F, Li Q, Engelmann R, et al. Improving radiologists' recommendations with computer-aided diagnosis for management of small nodules detected by CT. *Acad Radiol*. 2006; 13:943–950. [PubMed: 16843846]
8. Alberdi E, Povykalo A, Strigini L, et al. Effects of incorrect computer-aided detection (CAD) output on human decisionmaking in mammography. *Acad Radiol*. 2004; 11:909–918. [PubMed: 15354301]
9. Chan HP, Sahiner B, Helvie MA, et al. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study. *Radiology*. 1999; 212:817–827. [PubMed: 10478252]
10. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med*. 2007; 356:1399–1409. [PubMed: 17409321]

11. Armato SG 3rd, Roberts RY, Kocherginsky M, et al. Assessment of radiologist performance in the detection of lung nodules: dependence on the definition of “truth”. *Acad Radiol.* 2009; 16:28–38. [PubMed: 19064209]
12. Freedman MT. Computer-aided detection of lung cancer on chest radiographs: effect of machine CAD true positive/false negative detections on radiologists’ confidence level. *Proc SPIE.* 2004; 537210.1117/1112.536157
13. Taylor SA, Halligan S, Slater A, et al. Comparison of radiologists’ confidence in excluding significant colorectal neoplasia with multidetector-row CT colonography compared with double contrast barium enema. *Br J Radiol.* 2006; 79:208–215. [PubMed: 16498032]

TABLE 1

Comparison Between Characteristics of Suspected Nodules With and Without NDS Marks

Reader-reported Nodule Characteristics	NDS Mark Absent (n=2563)	NDS Mark Present (n=349)
Diameter (mm):mean (SD)	3.77 (2.37) [*]	6.71 (2.89) [*]
Density		
Solid (%)	70.6 [*]	87.1 [*]
Part solid (%)	15.5 [*]	8.6 [*]
Ground glass (%)	13.9 [*]	4.3 [*]
Laterality		
Left (%)	42.1 [*]	55.6 [*]
Right (%)	57.9 [*]	44.4 [*]
Lobe		
Lower (%)	44.3 [*]	54.2 [*]
Middle (%)	21.2 [*]	8.9 [*]
Upper (%)	34.5	37.0

^{*} $P < 0.05$ after adjusting for within-reader and within-case clustering.

TABLE 2
Post-NDS Change in Confidence, Stratified by Nodule Size, Pre-NDS Confidence Level, and Presence/Absence of NDS Mark

Pre-NDS Reader-assigned Confidence Level	Post-NDS Change in Confidence							
	NDS Mark Absent (n=2563)				NDS Mark Present (n=349)			
	N	Confidence Declined (%) [*]	Confidence Unchanged (%) [*]	Confidence Increased (%) [*]	N	Confidence Declined (%) [*]	Confidence Unchanged (%) [*]	Confidence Increased (%) [*]
Nodules <4 mm								
1% to 33%	78	2.6	83.3	14.1	2	0.0	0.0	100.0
34% to 66%	155	11.0	69.0	20.0	3	33.3	0.0	66.7
67% to 99%	174	9.8	64.9	25.3	3	0.0	33.3	66.7
100%	908	3.2	96.8	†	23	4.4	95.7	†
Total of nodules<4 mm	1315	4.9	88.5	21.2‡	31	6.5	74.2	75.0‡
Nodules ≥4 mm								
1% to 33%	80	3.8	83.8	12.5	3	0.0	66.7	33.3
34% to 66%	168	6.0	85.1	8.9	15	0.0	53.3	46.7
67% to 99%	218	6.4	83.0	10.6	25	0.0	52.0	48.0
100%	782	5.9	94.1	†	275	0.4	99.6	†
Total of nodules ≥4 mm	1248	5.9	90.3	10.3‡	318	0.3	93.4	46.5‡
All nodules								
1% to 33%	158	3.2	83.5	13.3	5	0.0	40.0	60.0
34% to 66%	323	8.4	77.4	14.2	18	5.6	44.4	50.0
67% to 99%	392	7.9	75.0	17.1	28	0.0	50.0	50.0
100%	1690	4.4	95.6	†	298	0.7	99.3	†
Total (all nodules)	2563	5.4	89.4	15.4‡	349	0.9	91.7	51.0‡

^{*} Percentages should sum to 100% across rows for each of the strata, although some do not because of rounding.

[†] By design, readers were not allowed to increase their confidence level assessments for nodules they initially assigned (i.e. Pre-NDS) a 100% confidence level.

[‡] Increase in reader confidence was only assessed in suspected nodules with Pre-NDS confidence <100%.

TABLE 3
Summary of Multivariable Models Predicting Declines and Increases in Reader Confidence for Nodules ≥ 4 mm in Diameter

Variable	Model Predicting Confidence Decline			Model Predicting Confidence Increase		
	Eligible Nodule Assessments (n)	Number (%) With Confidence Decline	Adjusted* Odds Ratio (95% CI)	Eligible Nodule Assessments (n)	Number (%) With Confidence Increase	Adjusted* Odds Ratio (95% CI)
NDS mark						
Present	318	1 (0.3)	Reference	43	20 (46.5)	6.0 (2.7–13.6)
Absent	1248	73 (5.9)	71.0 (14.8–339.7)	466	48 (10.3)	Reference
Nodule diameter (mm)						
4.0–5.9	964	49 (5.1)	4.7 (1.0–21.6)	359	37 (10.3)	Reference
6.0–7.9	347	21 (6.1)	5.3 (1.1–25.7)	93	18 (19.4)	2.4 (1.3–4.4)
8.0–9.9	138	2 (1.5)	1.8 (0.3–11.7)	28	7 (25.0)	2.5 (0.9–7.0)
≥ 10	117	2 (1.7)	Reference	29	6 (20.7)	5.0 (1.6–15.6)
Nodule density						
Ground glass	201	9 (4.5)	0.8 (0.4–1.9)	134	18 (13.4)	1.1 (0.6–2.3)
Part solid	215	12 (5.6)	0.5 (0.3–1.0)	123	29 (23.4)	1.6 (0.8–3.2)
Solid	1149	53 (4.6)	Reference	251	21 (8.4)	Reference
Laterality						
Left	696	28 (4.0)	Reference	206	32 (15.5)	Reference
Right	870	46 (5.3)	1.3 (0.8–2.2)	303	36 (11.9)	0.9 (0.5–1.5)
Lobe						
Lower	779	33 (4.2)	1.1 (0.6–2.1)	229	37 (16.2)	1.6 (0.8–3.0)
Middle	305	20 (6.6)	1.9 (1.0–3.9)	99	8 (8.1)	1.1 (0.5–2.5)
Upper	482	21 (4.4)	Reference	181	23 (12.7)	Reference
Reader experience						
Radiology residents	547	42 (7.7)	0.3 (0.0–31.7)	280	20 (7.1)	0.2 (0.0–4.5)
General radiologists	743	16 (2.2)	0.5 (0.0–56.3)	201	37 (18.4)	0.1 (0.0–3.2)
Senior chest faculty	276	16 (5.8)	Reference	28	11 (39.3)	Reference

* Both models adjusted for each of the variables listed in this table and accounted for clustering of responses within readers and within cases.