



# Variable radiological lung nodule evaluation leads to divergent management recommendations

Arjun Nair<sup>1,16</sup>, Emily C. Bartlett<sup>2,16</sup>, Simon L.F. Walsh<sup>2</sup>, Athol U. Wells<sup>3</sup>, Neal Navani<sup>4</sup>, Georgia Hardavella<sup>5</sup>, Sanjeev Bhalla<sup>6</sup>, Lucio Calandriello<sup>7</sup>, Anand Devaraj<sup>8</sup>, Jin Mo Goo<sup>9</sup>, Jeffrey S. Klein<sup>10</sup>, Heber MacMahon<sup>11</sup>, C.M. Schaefer-Prokop<sup>12</sup>, Joon-Beom Seo<sup>13</sup>, Nicola Sverzellati<sup>14</sup>, Sujal R. Desai<sup>2,8,15</sup> and the Lung Nodule Evaluation Group<sup>17</sup>

 @ERSpublications

**Radiological variations in pulmonary nodule evaluation translate into good overall agreement for guideline-modelled management, but poor agreement for radiologists' actual recommendations: 21% of radiologists can disagree with majority decisions** <http://ow.ly/bPOu30mcRYA>

**Cite this article as:** Nair A, Bartlett EC, Walsh SLF, *et al.* Variable radiological lung nodule evaluation leads to divergent management recommendations. *Eur Respir J* 2018; 52: 1801359 [<https://doi.org/10.1183/13993003.01359-2018>].

**ABSTRACT** Radiological evaluation of incidentally detected lung nodules on computed tomography (CT) influences management. We assessed international radiological variation in 1) pulmonary nodule characterisation; 2) hypothetical guideline-derived management; and 3) radiologists' management recommendations.

107 radiologists from 25 countries evaluated 69 CT-detected nodules, recording: 1) first-choice composition (solid, part-solid or ground-glass, with percentage confidence); 2) morphological features; 3) dimensions; 4) recommended management; and 5) decision-influencing factors. We modelled hypothetical management decisions on the 2005 and updated 2017 Fleischner Society, and both liberal and parsimonious interpretations of the British Thoracic Society 2015 guidelines.

Overall agreement for first-choice nodule composition was good (Fleiss'  $\kappa=0.65$ ), but poorest for part-solid nodules (weighted  $\kappa$  0.62, interquartile range 0.50–0.71). Morphological variables, including spiculation ( $\kappa=0.35$ ), showed poor-to-moderate agreement ( $\kappa=0.23$ –0.53). Variation in diameter was greatest at key thresholds (5 mm and 6 mm). Agreement for radiologists' recommendations was poor ( $\kappa=0.30$ ); 21% disagreed with the majority. Although agreement within the four guideline-modelled management strategies was good ( $\kappa=0.63$ –0.73), 5–10% of radiologists would disagree with majority decisions if they applied guidelines strictly.

Agreement was lowest for part-solid nodules, while significant measurement variation exists at important size thresholds. These variations resulted in generally good agreement for guideline-modelled management, but poor agreement for radiologists' actual recommendations.

---

Published online Dec 20, 2018; republished Dec 06, 2019 with a correction to the name of one collaborator.

This article has supplementary material available from [erj.ersjournals.com](http://erj.ersjournals.com)

Received: July 18 2018 | Accepted after revision: Oct 07 2018

Copyright ©ERS 2018

**Affiliations:** <sup>1</sup>Dept of Radiology, University College London Hospitals NHS Foundation Trust, London, UK. <sup>2</sup>Dept of Radiology, King's College Hospital NHS Foundation Trust, London, UK. <sup>3</sup>Dept of Respiratory Medicine, The Royal Brompton Hospital and Harefield NHS Foundation Trust, London, UK. <sup>4</sup>Dept of Thoracic Medicine, UCLH and Lungs for Living Centre, UCL Respiratory, University College London, London, UK. <sup>5</sup>Dept of Respiratory Medicine, King's College Hospital, London, UK. <sup>6</sup>Mallinckrodt Institute of Radiology, St. Louis, MO, USA. <sup>7</sup>Radiologia Diagnostica e Interventistica Generale - Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy. <sup>8</sup>Dept of Radiology, The Royal Brompton Hospital and Harefield NHS Foundation Trust, London, UK. <sup>9</sup>Seoul National University Hospital, Seoul, South Korea. <sup>10</sup>The University of Vermont Medical Center, Burlington, VT, USA. <sup>11</sup>Dept of Radiology, University of Chicago Medical Center, Chicago, IL, USA. <sup>12</sup>Dept of Radiology, Meander Medical Centre, Amersfoort, The Netherlands. <sup>13</sup>Dept of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea. <sup>14</sup>Dept of Clinical Sciences, Division of Radiology, University of Parma, Parma, Italy. <sup>15</sup>National Heart and Lung Institute, Imperial College London, London, UK. <sup>16</sup>Both authors contributed equally. <sup>17</sup>A list of the Lung Nodule Evaluation Group collaborators can be found at the end of this article.

**Correspondence:** Arjun Nair, Dept of Radiology, University College London Hospitals NHS Foundation Trust, 235 Euston Rd, Fitzrovia, London NW1 2BU, UK. E-mail: arjun7764@gmail.com

## Introduction

Pulmonary nodules detected on thoracic computed tomography (CT), either incidentally [1] or at screening [2, 3], are common and the majority are benign. However, subsolid nodules (encompassing part-solid (PSN) and pure ground-glass (GGN) lesions), are associated with a higher likelihood of malignancy [4, 5]. Published guidelines aim to standardise imaging follow-up and management [6–9] with initial decisions governed by classification of nodule morphology (*i.e.* solid *versus* part-solid *versus* pure ground-glass) and size (either the maximum diameter for a solid nodules or the solid component of part-solid lesions) [6–10].

Decisions about the management of nodules, especially larger nodules (typically those  $\geq 8$  mm in maximum diameter) usually follow multidisciplinary team discussion; this is especially true in tertiary referral or academic centres where such expertise is available. However, in nonspecialist centres, physicians do not adhere rigorously to existing guidelines [11], and the follow-up of nodules is likely, to a greater or lesser degree, to be influenced by the morphological description and recommendations in the radiological report. Previous studies indicate that, while the overall agreement for the characterisation of pulmonary nodules is fair to good, there is not-inconsiderable variation [12–14]. However, these studies have been limited either by the small numbers of radiologists or because of cohort enrichment, in which the original morphological categorisations (as provided by participating radiologists or from pre-annotated lung cancer screening cases), have been used to define a reference standard. Moreover, earlier studies have not formally investigated variation in radiologist's recommendation for further nodule management.

In recognising the importance of international collaboration for enhancing the relevance of research [15], the aim of the present study was to quantify interobserver agreement and the magnitude of variation among a large international cohort of thoracic radiologists for 1) nodule morphology and dimension; 2) hypothetical guideline-derived management decisions; and 3) the recommended management options chosen by radiologists.

## Materials and methods

### Case selection

The baseline CT studies of 64 consecutive patients from routine clinical practice (female  $n=36$ ; median (range) age 66.5 (26–90) years), with 70 nodules reviewed by the pulmonary nodule management service at King's College Hospital (London, UK), were retrospectively identified for the study. All patients with incidentally detected lung nodules on CT are referred to the pulmonary nodule management service for decisions regarding discharge and further follow-up (based on risk factors, size and composition). Two patients had three nodules each, another two patients had two nodules each, and the remaining 60 patients each had a single nodule under follow-up. For the purposes of this study, each nodule was considered as a separate case; therefore, radiologists were instructed that there was only one nodule per case and that there were 70 nodules in 70 cases to be evaluated. While the upper maximum diameter for a pulmonary nodule is arbitrarily set at 3 cm [16], for the purposes of simplicity, there was no upper diameter threshold in this study. For the retrospective review of clinically indicated data in this fashion, the institutional review board waived the need for ethical approval and informed consent.

### Image acquisition, reconstruction and presentation

For each case, anonymised targeted contiguous CT slices, extracted from the original study and displaying the whole nodule on both lung and soft tissue reconstructions and conforming to current practice were saved as Digital Imaging and Communication in Medicine (DICOM) datasets. To avoid confusion, each nodule was marked on image sections using a red arrow (figure 1).

In each case, the clinical details comprising age, sex, pack-year smoking status, history of lung cancer (or other malignancy), and the presence/absence of emphysema on CT were provided (online supplementary material).

#### **Observer invitation and registration**

Members of the Fleischner Society, Society of Thoracic Radiology, European Society of Thoracic Imaging, British Society of Thoracic Imaging and the Korean Society of Thoracic Radiology were invited to participate *via* email. A bespoke secure website was created for registration (online supplementary material) and image evaluation. Following registration, radiologists securely downloaded a zipped file containing the DICOM datasets of all 70 cases.

#### **Nodule characterisation: categorisation and measurement of dimensions**

To mirror “real life” practice, radiologists were free to review cases on any viewer or picture archiving and communications system, alter window settings and orientation, and use electronic callipers/magnification tools. For each nodule, radiologists characterised nodule composition, size and morphology (table 1). Radiologists assigned a percentage level of confidence to each category of nodule composition (GGN, PSN and solid). For a given nodule, the category with the highest confidence level was taken as a radiologist’s first-choice categorisation.

#### **Nodule management recommendations**

##### *Management based on extant guidelines*

To assess the impact of nodule characterisation and measurement on interobserver agreement for initial and subsequent management of nodules, assuming adherence to existing pulmonary nodule management guidelines, we modelled the management decisions for each nodule and observer based on characterisation of that nodule in four scenarios (table 2), as follows. 1) The Fleischner Society 2005 [6] and 2013 [7] recommendations for the management of pulmonary nodules (FG1); 2) the updated Fleischner Society 2017 [9] recommendations for the management of pulmonary nodules (FG2); 3) a liberal interpretation of the British Thoracic Society (BTS) 2015 guidelines for the management of pulmonary nodules [8], where any suspicious morphological features warrant further investigation (BTS1); and 4) a parsimonious interpretation of the BTS 2015 guidelines where suspicious morphological features were not considered as warranting further investigation (BTS2) (table 2).

We modelled both liberal and parsimonious interpretations of the BTS 2015 guidelines because the designation of morphological features that are suspicious (lobulation, a concave notch in the nodule margin, bubble-like lucencies, air bronchograms and pleural indentation or tags) can be subjective, and as such we wished to assess the effect of regarding any such feature as suspicious (in BTS1) or disregarding any of them entirely (in BTS2).

We assigned one of three management strategies to each nodule/observer combination as follows. 1) No further action; 2) CT surveillance; or 3) further investigation (including recommendations for positron emission tomography (PET)-CT and invasive workup).

Further details of our modelling are provided in the online supplementary material.

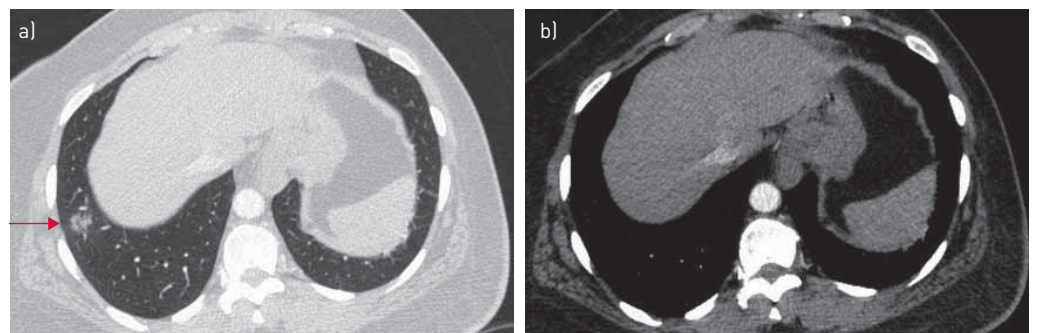


FIGURE 1 a) Pulmonary nodule in the right lower lobe demarcated by a red arrow on the lung reconstruction, viewed on lung windows; b) soft tissue reconstructions on mediastinal windows. The first-choice classification for this nodule was ground-glass by 33 radiologists (all with  $\geq 60\%$  confidence), part-solid (PSN) by 71 radiologists (all with  $\geq 55\%$  confidence), and solid by three radiologists (all with  $\geq 75\%$  confidence). As 66 radiologists chose PSN as the first-choice categorisation with  $\geq 70\%$  confidence, this nodule was assigned a majority classification of PSN.

TABLE 1 Nodule characterisation and management recommendations

<b>Nodule characterisation<sup>#</sup></b>	
Composition	Categorise as pure GGN, PSN or solid nodule Assign a percentage level of confidence, in 5% increments, for each categorisation (figure 1)
Diameter	For whole nodule, record longest axial diameter (nearest whole mm); and If first choice-categorisation was PSN, record the maximum diameter (whole mm) of the solid component
Solid proportion	Visually estimate (to the nearest 5%), the proportion of the nodule deemed solid, as a percentage of the total nodule volume (note: for GGNs and solid nodules, radiologists recorded the solid proportion as 0% and 100%, respectively)
Morphological features	Record the presence/absence of the following morphological features regarded as more often indicative of a malignant lesion [17–20]. 1) Spiculation; 2) lobulation; 3) a concave notch in the nodule margin; 4) “bubble-like” lucencies; 5) air bronchograms; 6) vessel touching or passing through the nodule; and 7) pleural indentation or tags (sample images illustrating each feature were provided on the website)
<b>Questions on management recommendations<sup>#</sup></b>	
1) Does this case require follow-up or invasive management?	“No, I would discharge the patient” OR “yes” If yes: CT follow-up OR proceed directly to invasive management If “CT follow-up”, recommend interval for first and second low-dose follow-up CT scan (assuming persistent but stable nodule on the first follow-up CT) If “proceed directly to invasive management”, choose one of the following: 1) biopsy followed by surgical management if deemed appropriate following MDT discussion 2) direct to surgical resection (no biopsy recommended) 3) thermal ablation 4) other management <sup>#</sup>
2) Would you also request a FDG PET-CT?	“No” OR “yes”
3) List, in order of descending importance (to the nearest 5%), factors influencing your management recommendations as follows: nodule composition, overall nodule dimension, dimension of the solid component (for PSN), nodule features other than size, smoking status and other clinical features	
FDG: 2-[ <sup>18</sup> F]-fluoro-2-deoxy-D-glucose; PET: positron emission tomography; CT: computed tomography; GGN: ground-glass nodule; PSN: part-solid nodule; MDT: multidisciplinary team. <sup>#</sup> : all responses were restricted to predetermined dropdown options on the website, except for the “other management” option if selecting “proceed directly to invasive management”.	

#### *Management recommendations by radiologists*

Following nodule characterisation, radiologists were asked to make recommendations on management (table 1) as they would in usual practice. Specifically, radiologists could select no further action, CT surveillance, proceeding directly to invasive management or a hybrid strategy comprising CT surveillance followed by invasive management (as indicated by free-text responses).

Radiologists were not specifically required to follow any published guideline in their decision-making.

#### *Statistical analysis*

We measured overall unweighted interobserver agreement for nodule composition and binary morphological variables with multirater Fleiss’  $\kappa$  statistics, and pairwise weighted interobserver agreement for nodule composition with Cohen’s weighted  $\kappa$  coefficient values ( $\kappa_w$ ). Weighting was assigned by converting the percentage confidence level for nodule composition into a four-point (0–3) scale as follows. 0=not confident; 1=low confidence (5–25%); 2=intermediate confidence (30–65%); and 3=high confidence ( $\geq 70\%$ ).  $\kappa_w$  is expressed as median and interquartile range (IQR) values for all 5671 unique observer pair combinations per nodule. We categorised Fleiss’  $\kappa$  and Cohen’s  $\kappa_w$  as poor ( $0 < \kappa / \kappa_w \leq 0.20$ ), fair ( $0.20 < \kappa / \kappa_w \leq 0.40$ ), moderate ( $0.40 < \kappa / \kappa_w \leq 0.60$ ), good ( $0.60 < \kappa / \kappa_w \leq 0.80$ ) and excellent ( $0.80 < \kappa / \kappa_w \leq 1.00$ ).

TABLE 2 Modelling of guideline-based management decisions

	No further action/not applicable	Recommendation	
		CT surveillance	Further investigation
<b>FG1</b>	Not applicable due to age/ extrathoracic cancer/history of lung cancer; GGN or PSN $\leq$ 5 mm	Solid nodules up to 8 mm; GGN>5 mm PSN with solid component <5 mm	Solid nodules >8 mm; PSN with solid component $\geq$ 5 mm
<b>FG2</b>	Solid nodules <6 mm, not in upper lobes and with no spiculation; GGN and PSN <6 mm	Solid nodules $\geq$ 6 mm OR solid nodules <6 mm AND either in upper lobes or spiculated; GGN $\geq$ 6 mm, OR PSN $\geq$ 6 mm with a solid component <6 mm	Solid nodules >8 mm; PSN with solid component $\geq$ 6 mm
<b>BTS1</b>	All nodules <5 mm	Solid nodules $\geq$ 5 mm and <8 mm OR solid nodules $\geq$ 8 mm but with Brock score <10%; GGN or PSN with no suspicious features <sup>#</sup> , solid component <6 mm AND Brock score <10%	Solid nodules $\geq$ 8 mm with Brock score $\geq$ 10%; GGN or PSN $\geq$ 5 mm with suspicious morphological features <sup>#</sup> OR solid component $\geq$ 6 mm OR Brock score $\geq$ 10%
<b>BTS2</b>	All nodules <5 mm	Solid nodules $\geq$ 5 mm and <8 mm OR solid nodules $\geq$ 8 mm but with Brock score <10%; GGN or PSN with solid component <6 mm AND Brock score <10%	Solid nodules $\geq$ 8 mm with Brock score $\geq$ 10%; GGN or PSN $\geq$ 5 mm with solid component $\geq$ 6 mm OR Brock score $\geq$ 10%

CT: computed tomography; FG1: Fleischner Society 2005 and 2013 guidelines; FG2: Fleischner Society 2017 guidelines; BTS1: British Thoracic Society 2015 guidelines with suspicious morphological features warranting further investigation; BTS2: British Thoracic Society 2015 guidelines where suspicious morphological features were not considered as warranting further investigation; GGN: ground-glass nodule; PSN: part-solid nodule. <sup>#</sup>: suspicious morphological features were lobulation, a concave notch in the nodule margin, bubble-like lucencies, air bronchograms and pleural indentation or tags.

For each nodule, we calculated multiple classification and measurement metrics, and assessed relevant relationships between these measurements. Details are provided in the online supplementary material.

We analysed interobserver agreement on management strategies as chosen by the radiologists and within each modelled guideline using multirater Fleiss'  $\kappa$ . For the modelled guideline evaluation, we determined the majority decision for each nodule, and calculated the number and percentage of nodules where there would be disagreement on management even when radiologists agreed on the nodule classification.

Statistical analysis was performed using MedCalc for Windows (version 14.12.0; MedCalc Software, Ostend, Belgium) and Stata (StataCorp, College Station, TX, USA). Multirater Fleiss'  $\kappa$  was generated using the Real Statistics Resource Pack (version 5.1; Charles Zaiontz, available at [www.real-statistics.com](http://www.real-statistics.com)). A p-value of <0.05 was considered statistically significant.

## Results

### Baseline data

Out of 70 nodules, one was excluded from analysis, as some radiologists wished to record a maximum diameter that exceeded the available dropdown options provided on the website. 107 radiologists (91 attending/consultants; median (range) experience 9 (1–50) years) from 25 countries (UK/Europe n=75; North and South America/Canada n=24; Asia and Africa n=8) completed the study.

### Agreement and diagnostic confidence for nodule characteristics

There was good overall agreement for first-choice categorisation ( $\kappa=0.652$ , 95% CI 0.650–0.655) (table 3). However, the weighted agreement for first-choice categorisation was best for solid nodules, followed by GGNs and worst for PSNs. Interobserver agreement for nodule composition was neither affected by the duration of thoracic observer experience nor by attendance at a lung cancer or a dedicated pulmonary nodule multidisciplinary team meeting (online supplementary tables S2 and S3). The overall median percentage of radiologists disagreeing with the predominant classification was 12.1% (IQR 2.8–30.8%), and was highest among nodules predominantly classified as PSNs, followed by GGNs and lowest for solid nodules.

For all nodules, >50% of the 107 radiologists designated their diagnostic confidence for their first-choice categorisation as high ( $\geq$ 70%), with a median of 94 (87.9%) out of 107 radiologists. However, the number of radiologists with this high level of confidence varied widely, ranging from 55 (51.4%) to 107 (100%)



TABLE 3 Characterisation of nodule features by radiologists

	Nodules classified by majority of radiologists	Weighted agreement for first-choice categorisation $\kappa_w$	Radiologists disagreeing with majority classification %	Fleiss' $\kappa$ (95% CI)
<b>Nodule type</b>				
Ground-glass	23 (33.3)	0.73 (0.65–0.80)	15.0 (4.9–32.4)	
Part-solid	18 (26.1)	0.62 (0.50–0.71)	20.1 (9.3–38.3)	
Solid	28 (40.6)	0.80 (0.72–0.85)	3.7 (1.9–17.8)	
<b>Morphological variables</b>				
Spiculation				0.353 (0.350–0.356)
Lobulated margin				0.233 (0.230–0.236)
Concave notch along margin				0.243 (0.240–0.246)
Bubble-like lucencies				0.492 (0.489–0.496)
Air bronchograms				0.483 (0.480–0.486)
Vessel touching or passing through				0.333 (0.330–0.336)
Pleural indentation/pleural tag				0.532 (0.529–0.535)

Data are presented as n (%) or median (interquartile range), unless otherwise stated.  $\kappa_w$ : weighted  $\kappa$ .

radiologists. The variation in first-choice diagnostic confidence was narrow, with a median coefficient of variation (CoV) of 9.7% (IQR 5.5–13.2%).

The multirater Fleiss'  $\kappa$  for the presence of the other morphological variables was poor to moderate (table 3). In particular, the agreement on spiculation was only fair ( $\kappa=0.365$ ). In 13 (18.8%) out of 69 nodules, all 107 radiologists agreed that spiculation was absent, but there was no nodule in which all 107 radiologists agreed that spiculation was present.

#### Agreement for nodule size

The median (IQR) longest nodule diameter measurement was 7.8 (5.9–13.4) mm. There were 15 (21.7%), 30 (43.5%), 18 (26.1%) and six (8.7%) nodules with median longest diameters of  $\leq 5$  mm,  $>5$  mm to  $\leq 10$  mm,  $>10$  mm to  $\leq 20$  mm and  $>20$ , respectively. The variation in diameter measurement for each nodule, in these categories, is shown in figure 2.

The median (IQR) CoV for longest diameter was 15.7 (12.9–18.7)%. When analysed at clinically relevant median diameter thresholds of 5 mm, 6 mm, 8 mm, 10 mm and 20 mm, there were greater magnitudes of variation for smaller (*i.e.* below the threshold) compared to larger (*i.e.* above the threshold) nodules at the 5 mm and 6 mm thresholds, but not at higher thresholds (table 4). Paradoxically, the six nodules with median diameters of  $>20$  mm demonstrated (statistically nonsignificant) greater CoVs than the 63 nodules  $\leq 20$  mm (18.1% *versus* 15.6%, respectively;  $p=0.45$ ).

Across 2091 nodule–observer combinations, where radiologists classified a nodule as PSN, the median (IQR) of the diameter and proportions of the solid component were 14.0 (9.0–19.0) mm and 40.0 (20.0–60.0)%, respectively, with wide variation (median (IQR) CoVs of 35.6 (21.7–44.3)% and 42.5 (27.8–59.9)%, respectively). In the 18 nodules predominantly classified as PSN, the median (IQR) of the diameter and proportions of the solid component were 6.5 (4.0–11.0) mm and 32.5 (20.0–50.0)%, respectively.

Relationships between nodule size, diagnostic confidence and magnitude of variation are presented in table 5. In brief, there was 1) a moderate negative correlation between the median and CoV of longest diameter ( $r_s=-0.52$ ,  $p=0.005$ ) for the 28 majority-classified solid nodules; and 2) a strong negative correlation between the median and CoV of the solid component diameter ( $r_s=-0.69$ ,  $p=0.0015$ ) and between the median and CoV of the solid proportion ( $r_s=-0.88$ ,  $p<0.0001$ ) for the 18 majority-classified PSNs. In other words, interobserver variation increased with decreasing size for both solid nodules and the solid component of PSNs.

#### Nodule management recommendations

##### Management based on extant guidelines

The overall agreement, majority decisions and number and proportions that would have hypothetically disagreed with the majority of nodule management decisions as defined by the four modelled guidelines are shown in table 6. All showed good agreement. Applying the latest Fleischner Society recommendations uniformly to this nodule cohort would have resulted in the lowest proportion of nodules requiring further investigation. Almost one in 20 radiologists would disagree with the majority management decision

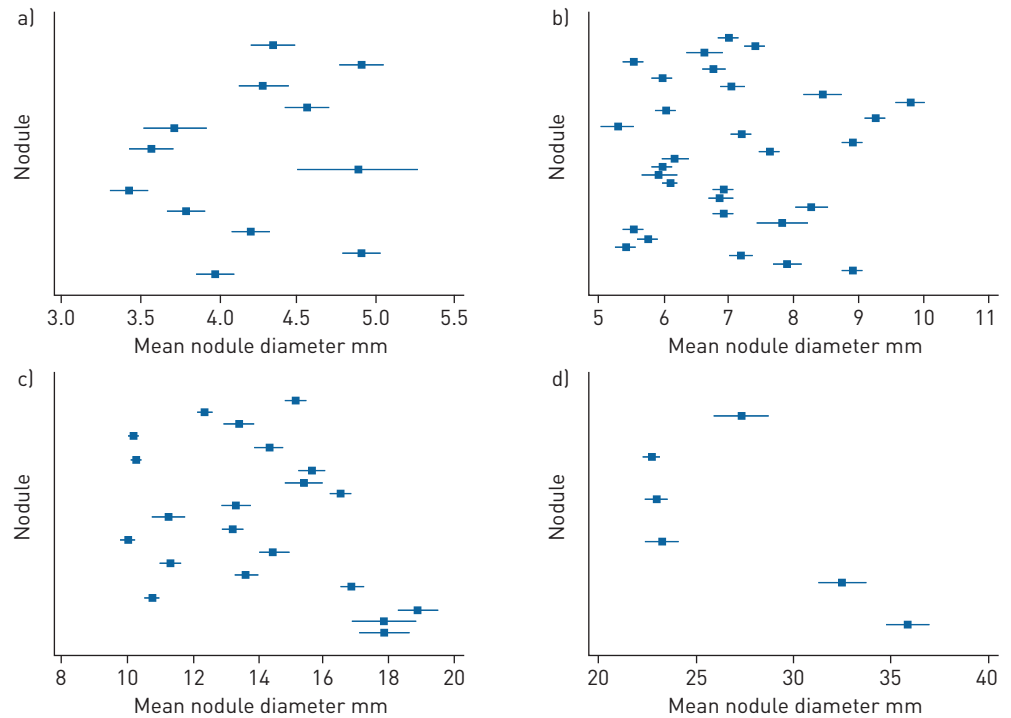


FIGURE 2 Forest plots depicting the variation in nodule diameters for nodules with mean diameters of a)  $\leq 5$  mm, b)  $>5$  mm to  $\leq 10$  mm, c)  $>10$  mm to  $\leq 20$  mm and d)  $>20$  mm. Each nodule is represented by a horizontal line, with the central square along the line representing the mean diameter, and the two ends of the line representing the lower and upper limits of the 95% confidence interval for nodule mean, respectively.

according to the more conservative model of the BTS guideline (BTS2), and this rises to almost one in 10 radiologists with the latest Fleischner Society recommendations.

Even when radiologists agreed with the majority classification of the nodule, there would have been disagreement on the need for further action in 12 (17.4%) out of 69, 26 (37.7%) out of 69, 31 (44.9%) out of 69 and 31 (44.9%) out of 69 nodules using FG1, FG2, BTS1 and BTS2, respectively. Of those nodules where there was disagreement, 0 (0.0%) out of 12, eight (30.8%) out of 26, 15 (48.4%) out of 31 and eight (25.8%) out of 31 would have had variation in the type of management (CT surveillance *versus* further investigation) using FG1, FG2, BTS1 and BTS2, respectively.

#### Management recommendations by radiologists

The overall agreement for the choice between the four management strategies that radiologists could select was only fair ( $\kappa=0.296$ , 95% CI 0.294–0.298). In all cases, further investigation was recommended by at least one radiologist. The majority chose no further action in six (8.7%) nodules, CT surveillance in 57

TABLE 4 Magnitude of variation of longest nodule diameter at various thresholds.

Diameter threshold mm	Coefficient of variation for longest diameter measurement %		p-value <sup>#</sup>
	At or below threshold	Above threshold	
5	17.2 [16.5–20.7]	14.6 [12.1–17.9]	0.0035
6	17.1 [15.0–20.6]	14.7 [12.1–17.8]	0.0107
8	16.5 [14.3–18.7]	13.9 [11.8–19.0]	0.1599
10	15.4 [12.7–17.8]	16.7 [13.2–20.0]	0.3479
20	15.6 [12.9–18.3]	18.1 [13.4–20.6]	0.4496

Data are presented as median [interquartile range], unless otherwise stated. <sup>#</sup>: Mann–Whitney U-test.

TABLE 5 Relationships between nodule size, diagnostic confidence and magnitude of variation (measured by coefficient of variation [CoV])

	Nodules n	Spearman's $\rho$ (95% CI)	p-value
<b>All nodules</b>	69		
Median longest diameter <i>versus</i> CoV of longest diameter		-0.146 [-0.370-0.942]	0.23
CoV of longest diameter <i>versus</i> CoV of first-choice diagnostic confidence		0.531 [0.337-0.682]	<0.001
<b>Nodules classified as solid by majority</b>	28		
Median longest diameter <i>versus</i> CoV of longest diameter		-0.516 [-0.745--0.177]	0.005
<b>Nodules classified as part-solid by majority</b>	18		
Median solid diameter <i>versus</i> CoV of solid diameter		-0.690 [-0.875--0.329]	0.002
Median solid proportion <i>versus</i> CoV of solid proportion		-0.875 [-0.953--0.690]	<0.001

(82.6%) nodules and proceeding directly to invasive management in six (8.7%) nodules. A median (IQR) 85 (72–102) radiologists (79.4%, 67.1–95.3%) agreed on the majority decision, and a median (IQR) of 22 (9–35) radiologists (20.6%; 8.4–32.9%) disagreed with the majority decision. Among radiologists disagreeing with the majority decision, the overall median (IQR) percentages of these observers recommending under- or over-management, relative to the majority decision, was 50 (0–100)% for both; therefore, there was no overall trend towards under- or over-management among disagreeing radiologists. However, there were 34 (49.2%) nodules in which the number of radiologists agreeing with the majority decision was below the median of 85; in other words, the majority decision in these nodules could be considered less compelling. In this subgroup of nodules, there was a tendency for the disagreeing radiologists to over-manage (median 77.6%, IQR 0–100%) rather than under-manage (median 22.4%, IQR 0–100%) such nodules, but this difference was not statistically significant (Wilcoxon test for paired samples  $p=0.86$ ).

For 16 (23.2%) out of 69 nodules, all 107 radiologists agreed on the need for further action. For these nodules, the median (and percentage of total) observers choosing CT follow-up, proceeding directly to invasive management or a hybrid strategy comprising CT and invasive management (as indicated by free-text responses) were 77 (72.0%), 31 (29.0%) and three (2.8%) radiologists, respectively. In addition, overall agreement for management, and for PET-CT in these 16 nodules, was only fair ( $\kappa=0.33$  and  $\kappa=0.32$ , respectively).

Nodule composition was considered the dominant factor in determining management by the majority of observers in the majority of nodules (65 (94.2%) out of 69). However, for a given nodule, there was a wide variation in the proportion of radiologists citing nodule composition as the most influential factor, with a median of 64 (59.8%) out of 107 observers (range 35 (32.7%) to 92 (86.0%) radiologists). In four (5.8%) out of 69 nodules, overall diameter was considered most important.

TABLE 6 Variation in management decisions based on hypothetical decision-making according to the Fleischner Society 2005 and 2013 (FG1) recommendations, the Fleischner Society 2017 (FG2) recommendations, the British Thoracic Society 2015 guidelines with suspicious morphological features warranting further investigation (BTS1) and without such features warranting further investigation (BTS2).

	Modelled guideline			
	FG1	FG2	BTS1	BTS2
<b>Overall agreement on management Fleiss' <math>\kappa</math> (95% CI)</b>	0.658 [0.655–0.660]	0.643 [0.641–0.645]	0.672 [0.670–0.675]	0.731 [0.729–0.734]
<b>Majority decision</b>				
No further action/guideline not applicable <sup>#</sup>	14 [20.3]	19 [27.5]	18 [26.1]	16 [23.2]
CT surveillance	40 [58.0]	39 [56.5]	21 [30.4]	27 [39.1]
Further investigation	15 [21.7]	11 [15.9]	30 [43.5]	26 [37.7]
<b>Radiologists disagreeing with majority decision n</b>	7 [2–23]	10 [2–28]	8 [2–27]	5 [2–21]
<b>Radiologists disagreeing with majority decision %</b>	6.5 [1.8–21.5]	9.3 [1.9–26.2]	7.5 [1.9–25.0]	4.7 [1.6–19.4]

Data are presented as n (%) or median (interquartile range), unless otherwise stated. CT: computed tomography. <sup>#</sup>: guidelines were considered not applicable in FG1 and FG2 for solid nodules in patients aged <35 years, with a known extrathoracic malignancy or previous history of lung cancer.



## Discussion

The present study is the largest international evaluation of observer agreement for nodule characterisation and, to the best of our knowledge, the only study to have specifically evaluated variation for self-reported management recommendations. Our results show that agreement was lowest for PSNs, but that variation is not influenced by observer experience. We have also shown that around one in five radiologists disagreed with the part-solid classification assigned by the majority of radiologists. Moreover, there was an inverse relationship between the magnitude of variation and median diameters as well as proportions of the solid component in PSNs. With respect to radiologists' management recommendations, the overall agreement was only fair, and ~20% of radiologists disagreed with the majority decision. These findings are important not only for highlighting difficulties in nodule characterisation, but also given current issues regarding the management of adenocarcinomas with a small solid component (measuring  $\leq 5$  mm) [21, 22], which may mandate more conservative management.

The overall variation for nodule classification in our study mirrors that in the study by RIDGE *et al.* [14] for differentiating solid from subsolid nodules, but is greater than for screen-detected nodules reviewed by VAN RIEL *et al.* [13]. The differences between studies may, in part, be attributable to variations in the relative prevalence of nodule types included in the studies. However, by requiring radiologists to assign a level of diagnostic confidence to their chosen categorisation, we were able to weight categorisations to emphasise differences in diagnostic confidence and, thus, arguably obtain a more accurate representation of interobserver agreement. Furthermore, we analysed variation at the important diameter thresholds which dictate either different nodule management recommendations (*i.e.* in the case of 5 mm, 6 mm and 8 mm) or tumour stage (in the case of 10 mm and 20 mm), should a nodule prove to be a primary lung cancer [23]. We found that variation in maximum nodule diameter was greatest for nodules at or below 5 mm and 6 mm, the key diameter thresholds (based on current guidelines [8, 9]), governing whether discharge or further CT surveillance should be recommended. The problem of interobserver variation for diameter measurement of small pulmonary nodules using electronic callipers is not new [24, 25]. Variation in measurement at the specified diameter thresholds for follow-up potentially translates into variation in recommendations for further follow-up, even though radiologists may agree on the nodule composition. Indeed, in our study over one-third of nodules would have had a variation in recommended management strategy, irrespective of the guideline used. Such variation in measurement and recommendations may be reduced by applying volumetric analysis, computer-aided detection tools [26, 27] and, potentially, artificial intelligence (AI) techniques. However, with respect to the latter, the interobserver variation in our study sounds an equally important note of caution for AI nodule characterisation tools developed using machine-learning techniques; such tools require datasets annotated by radiologists for their development, and may thus mimic the biases and interobserver variation of their contributing radiologists.

To assess how management recommendations might vary if guidelines were adhered to, we modelled hypothetical management decisions based on four guideline scenarios. Modelling of this type inherently foregoes interobserver variation in decision-making by assuming that guidelines will be applied consistently based on nodule composition and diameter, and is thus artificial. However, our findings illustrate that even in such artificial situations, agreement is at best good, and up to almost one in 10 radiologists would deviate from the majority management decision.

An important ancillary finding of the present study was the poor-to-moderate agreement for other morphological variables. Our results, derived from a much larger observer pool, contrast with the high agreement for assessing nodule margins, air bronchograms, bubble-like lucencies and pleural tags in the study by LIM *et al.* [17]. In part, this may be because different morphological features are more difficult to identify in smaller lesions: in our study, around two-thirds of nodules were  $\leq 10$  mm in diameter. This aside, a salutary finding was that observer variation for the presence/absence of spiculation (Fleiss'  $\kappa=0.35$ ) was below clinically-acceptable limits. Plainly, there are implications for future modifications to existing guidelines, not only because the presence of spiculation is regarded as a more indicative sign of cancer, but also because of its importance as a variable in risk calculators [5, 28]. In addition, it is noteworthy that pleural indentation (Fleiss'  $\kappa=0.53$ ) and "bubble-like" lucencies (Fleiss'  $\kappa=0.49$ ), features regarded as "suspicious" in the BTS guidelines [8], were associated, at best, with fair observer agreement. Subjectivity in the interpretation of these features will further add to the variation in management strategies. As we were interested in the effect that these other morphological variables could have on the decisions taken in the BTS guideline, we modelled two versions of this guideline, one mandating further investigation when these suspicious features were considered present, and the other with a conservative approach irrespective of their presence. The conservative approach resulted in only four fewer nodules requiring further investigation according to the majority of radiologists, suggesting that the effect of these morphological features on recommended nodule management was limited, at least in our cohort.

Since modelling of management strategies based on guidelines is artificial, it was essential that we also capture the real-life management decisions preferred by radiologists for each nodule. The interobserver agreement for these management strategies in practice was only fair, and much lower than that for any of the modelled guidelines, despite all radiologists listing nodule composition or nodule size as the most influential factors in their decision-making. This lower level of agreement reinforces the role of radiologists' subjective judgement in deciding nodule management, despite existing recommendations. A recent study demonstrated that such subjective upgrading of management recommendations using the American College of Radiology Lung CT Screening Reporting and Data System (Lung-RADS) assessment 4X category resulted in higher malignancy detection rates in lung cancer screening [29], underscoring this point. In addition, it is worth noting that, although our nodule dataset mirrored real-life clinical practice, the radiologists in our study still performed their evaluations under "test" conditions, and the majority were of higher grade (consultant/attending). As such, it is conceivable that real-world management variation may be considerably higher.

Our results also have potential ramifications for respiratory physicians interpreting radiological reports and recommendations. In particular, physicians practising in nonspecialist centres may wish to obtain a tertiary centre multidisciplinary opinion on the classification and management of nodules 1) that have been classified as PSNs (given the lower agreement for this classification); 2) that are at the potential diameter thresholds for discharge (*i.e.* 5 mm and 6 mm); 3) in which management has been influenced by morphological variables (given the poor-to-moderate agreement for some CT features (*e.g.* spiculation)); and 4) where the radiological recommendation is discordant with that suggested by guidelines.

The current study has limitations: firstly, for the purposes of management strategies, the nodules were assumed to be stable. In practice, growth of nodules often leads to further investigation; however, our prime intention was to model the management of nodules as first discovered. Also, both the prior and current versions of the Fleischner recommendations advocate the use of the average of short- and long-axis diameter for measurement [6, 30], whereas we have modelled management on the longest diameter only. In support of our approach, a recent survey suggests that, at least for subsolid nodules, only 9% of radiologists manage nodules based on bidimensional measurements [31]. While no such data are available for solid nodules, we suggest that our modelling reflects everyday practice. Finally, we did not instruct radiologists as to the type of reconstruction (lung or soft tissue) or viewing window (lung or mediastinal) on which measurements should be made. Although recommendations in this regard have been made [7, 30], variation between radiologists in this measurement practice could have influenced measurement variation.

In conclusion, in a large international cohort of radiologists, we found that the overall agreement for nodule composition was good, but was lowest for PSNs, and significant magnitudes of variation in diameter measurement exist at important diameter thresholds. As such, radiologists are not consistent in characterising or measuring pulmonary nodules. These variations resulted in generally good agreement for management modelled on current recommendations, but agreement on self-reported management decisions taken by radiologists is poor.

The Lung Nodule Evaluation Group collaborators: Fredrik Ahlfors, European Telemedicine Clinic, Sweden; Andréa Bastos, Universidade Federal de Minas Gerais – Faculdade de Medicina, Brazil; Julien Behr, Hospital Jean Minjoc Besançon, France; Jiri Benes Jr, General University Hospital, Charles University, Czech Republic; Gracijela Bozovic, Radiology Department at University Hospital Lund, Sweden; Maria Teodora Buzan, Papworth Hospital NHS Foundation Trust, UK; Eva Castañer, Institut Universitari del Parc Tauli, UDIAT – Centre Diagnòstic, Spain; Marcelo Chaves, DASA, Brazil; Yo Won Choi, Hanyang University Seoul Hospital, Republic of Korea; Giorgia Dalpiaz, Bellaria Hospital, Italy; Sarojini David, University Hospital Lewisham, UK; Pim A. De Jong, UMC Utrecht, the Netherlands; Gael Dournes, Centre Hospitalier Universitaire de Bordeaux, France; Matthias Eberhard, University Hospital Zurich, Switzerland; Dante Escuissato, Federal University of Paraná, Brazil; Alessandra Farchione, A. Gemelli University Hospital, Università Cattolica del Sacro Cuore, Italy; Lucia Flors, University of Virginia, USA; Paola Franchi, AUSL Teramo, Italy; Thomas Frauenfelder, Radiology, University Hospital Zurich, Switzerland; Kiminori Fujimoto, Department of Radiology, Kurume University School of Medicine, Japan; Marco Gatti, University of Turin, Italy; Giampaolo Gavelli, Istituto Tumori Romagna, Italy; Aleksandar Grgic, Radiological Practice, University Hospital, Germany; Louise Haine, Royal Derby Hospital, UK; Linda Haramati, Albert Einstein College of Medicine, Montefiore Medical Center, USA; Thomas Hartman, Mayo Clinic, USA; Ieneke Hartmann, Maastricht Hospital Rotterdam, the Netherlands; Jan Philipp Hering, Institute of Clinical Radiology University Muenster, Germany; Vincent Herpels, AZ Groeninge, Belgium; Inmaculada Herráez, Hospital Universitario de León, Spain; Susan Hobbs, University of Rochester Medical Center, USA; Bruno Hochhegger, Pontifical Catholic University of Rio Grande do Sul, Brazil; Nur Hursoy, Department of Radiology, Ankara University School of Medicine, Turkey; Joana Ip, Fundação Champalimaud, Portugal; Srikanth Iyengar, Frimley Health NHSFT, UK; Juan Arenas Jiménez, Hospital General Universitario de Alicante, Spain; Katharine Johnson, Salisbury NHS Trust, UK; Jeffrey Kanne, University of Wisconsin – Madison, USA; Sangita Kapur, University of Cincinnati, USA; Can Zafer Karaman, Adnan Menderes University, Turkey; Robert Karl, LSU, USA; Galina Kirova, Tokuda Hospital Sofia, Bulgaria; Jane Ko, NYU Langone Medical Center, USA; Gerdien Kramer, CHRU, France; Melahat Kul, Department of Radiology, Ankara University School of Medicine, Turkey; Anna Rita Larici, Catholic University of the Sacred Heart of Rome, Italy; Youkyung Lee, Hanyang University Guri Hospital, Republic of

Korea; Leonard Leifels, Universitat Leipzig, Germany; Sebastian Ley, University Teaching Hospital, Germany; Stephen Machnicki, Lenox Hill Hospital, USA; Panagiotis Maidas, University Hospital of Aalborg, Denmark; Sze Mun Mak, Guy's and St Thomas' Hospital, UK; Francesca Mariani, Sant'Orsola Hospital, Italy; Gustavo Meirelles, Fleury Group, Brazil; Firdaus A. Mohamed Hoesein, University Medical Center Utrecht, the Netherlands; Xavier Montet, Geneva University Hospital, Switzerland; John Murchison, Edinburgh Royal Infirmary, UK; Hrudaya Nath, UAB, USA; Tatyana Nenkina, private hospital, Bulgaria; Edward Nganga, Aga Khan University Hospital, Kenya; Thi Dan Linh Nguyen-Kim, University Hospital Zurich, Switzerland; Tilo Niemann, Kantonsspital Baden, Switzerland; Carlos Francisco Muñoz Núñez, Hospital Universitari i Politècnic La Fe, Spain; Mariaelena Occhipinti, University of Florence, Italy; Anastasia Oikonomou, Sunny Brook Health Sciences Centre, Canada; Yoshiyuki Ozawa, Nagoya City University Graduate School of Medical Sciences, Japan; Tej Pandher, Plymouth Hospital NHS Trust, UK; Anagha P. Parkar, Haralds plass, Deaconess Hospital, Norway; Sara Piciocchi, GB Morgagni-L Pierantoni Hospital, Forlì, Italy; Diogo Pinheiro, Hospital Vita, Brazil; Antonio Poerio, University of Bologna, Italy; Roberta Polverosi, Affidea-Padova, Italy; Kristen Pope, University of Kansas Medical Centre, USA; Helmut Prosch, Medical University of Vienna, Austria; Ilaria Pulzato, University of Genoa, Italy; Cristiano Rampinelli, European Institute of Oncology, Italy; Jonathan Rodrigues, Bristol Royal Infirmary, UK; Odysseus Romanos, private practice, Greece; Chiara Romei, Pisa University Hospital, Italy; Justus Roos, Cantonal Hospital Lucerne, Department of Radiology and Nuclear Medicine, Switzerland; Santiago Rossi, ROSSI, Argentina; Roman Rubtsov, Heidelberg University Hospital, Germany; Margherita Ruggirello, University of Parma, Italy; Marcelo Sánchez, Hospital Clinic de Barcelona, Spain; Marcel Koenigkamt Santos, University of Sao Paulo – School of Medicine of Ribeirao Preto, Brazil; Michael Schubert, University Hospital, Norway; Taiwo Senbanjo, Epsom and St Helier NHS Trust, UK; Claudio Silva, Clinica Alemana de Santiago, Chile; Mario Silva, University of Parma, Italy; Annemie Snoeckx, Antwerp University Hospital, Belgium; Sushilkumar Sonavane, University of Alabama at Birmingham, USA; Mona Sriharan, King's College Hospital NHS Foundation Trust, UK; Konstantinos Stefanidis, Kings College Hospital NHS Foundation Trust, UK; Onur Taydas, Hacettepe University, Turkey; Noriyuki Tomiyama, Diagnostic and Interventional Radiology, Graduate School Of Medicine, Osaka University, Japan; Sa Tran, King's College Hospital, UK; Alberto Villanueva, Hinchingsbrooke Hospital and Peterborough City Hospital, UK; Michelle Williams, University of Edinburgh, UK.

Conflict of interest: A. Nair: this work was partly undertaken at University College London Hospitals (UCLH)/University College London (UCL), by A. Nair, who received a proportion of funding from the Department of Health National Institute for Health Research (NIHR) Biomedical Research Centres funding scheme. A. Nair is a member of the advisory board for Aidence Artificial Intelligence, but has not received any fees from this entity to date, and the work with this entity is not in any way related to the current submission. E.C. Bartlett has nothing to disclose. S.L.F. Walsh has nothing to disclose. A.U. Wells has nothing to disclose. N. Navani has nothing to disclose. G. Hardavella has nothing to disclose. S. Bhalla has nothing to disclose. L. Calandriello has nothing to disclose. A. Devaraj has nothing to disclose. J.M. Goo reports grants from Lunit Inc., outside the submitted work. J.S. Klein has nothing to disclose. H. MacMahon reports personal fees for advisory board work from Riverain Technologies, personal fees for consultancy from GE Healthcare, grants from Philips Healthcare, honoraria from Konica-Minolta, stock options for research collaborations from Hologic, and payments for patents and licences from University of Chicago, outside the submitted work. C.M. Schaefer-Prokop has nothing to disclose. J-B. Seo has nothing to disclose. N. Sverzellati reports personal fees from Roche and Boehringer Ingelheim, outside the submitted work. S.R. Desai has nothing to disclose.

Support statement: This work was partly undertaken at University College London Hospitals (UCLH)/University College London (UCL), by A. Nair and N. Navani, who received a proportion of funding from the Department of Health National Institute for Health Research (NIHR) Biomedical Research Centres funding scheme. Funding information for this article has been deposited with the Crossref Funder Registry.

## References

- Gould MK, Tang T, Liu IL, *et al.* Recent trends in the identification of incidental pulmonary nodules. *Am J Respir Crit Care Med* 2015; 192: 1208–1214.
- Henschke CI, McCauley DI, Yankelevitz DF, *et al.* Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999; 354: 99–105.
- Diederich S, Wormanns D, Semik M, *et al.* Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers. *Radiology* 2002; 222: 773–781.
- Henschke CI, Yankelevitz DF, Mirtcheva R, *et al.* CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. *AJR Am J Roentgenol* 2002; 178: 1053–1057.
- McWilliams A, Tammemagi MC, Mayo JR, *et al.* Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013; 369: 910–919.
- MacMahon H, Austin JH, Gamsu G, *et al.* Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* 2005; 237: 395–400.
- Naidich DP, Bankier AA, MacMahon H, *et al.* Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 2013; 266: 304–317.
- Callister ME, Baldwin DR, Akram AR, *et al.* British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015; 70: Suppl. 2, ii1–ii54.
- MacMahon H, Naidich DP, Goo JM, *et al.* Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology* 2017; 284: 228–243.
- American College of Radiology. Lung CT Screening Reporting and Data System (Lung-RADS). [www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads](http://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads) Date last accessed: October 31, 2018. Date last updated: April 2014.
- Tanner NT, Aggarwal J, Gould MK, *et al.* Management of pulmonary nodules by community pulmonologists: a multicenter observational study. *Chest* 2015; 148: 1405–1414.
- Penn A, Ma M, Chou BB, *et al.* Interreader variability when applying the 2013 Fleischner guidelines for potential solitary subsolid lung nodules. *Acta Radiol* 2015; 56: 1180–1186.

- 13 van Riel SJ, Sánchez CI, Bankier AA, *et al.* Observer variability for classification of pulmonary nodules on low-dose CT images and its effect on nodule management. *Radiology* 2015; 277: 863–871.
- 14 Ridge CA, Yildirim A, Boiselle PM, *et al.* Differentiating between subsolid and solid pulmonary nodules at CT: inter- and intraobserver agreement between experienced thoracic radiologists. *Radiology* 2015; 278: 888–896.
- 15 Rosenkrantz AB, Parikh U, Duszak R Jr. Citation impact of collaboration in radiology research. *J Am Coll Radiol* 2018; 15: 258–261.
- 16 Hansell DM, Bankier AA, MacMahon H, *et al.* Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008; 246: 697–722.
- 17 Lim HJ, Ahn S, Lee KS, *et al.* Persistent pure ground-glass opacity lung nodules  $\geq 10$  mm in diameter at CT scan: histopathologic comparisons and prognostic implications. *Chest* 2013; 144: 1291–1299.
- 18 Lee SM, Park CM, Goo JM, *et al.* Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using CT features. *Radiology* 2013; 268: 265–273.
- 19 Miyata N, Endo M, Nakajima T, *et al.* High-resolution computed tomography findings of early mucinous adenocarcinomas and their pathologic characteristics in 22 surgically resected cases. *Eur J Radiol* 2015; 84: 993–997.
- 20 Zhang Y, Qiang JW, Ye JD, *et al.* High resolution CT in differentiating minimally invasive component in early lung adenocarcinoma. *Lung Cancer* 2014; 84: 236–241.
- 21 Lee JH, Park CM, Lee SM, *et al.* Persistent pulmonary subsolid nodules with solid portions of 5 mm or smaller: Their natural course and predictors of interval growth. *Eur Radiol* 2016; 26: 1529–1537.
- 22 Cohen JG, Reymond E, Lederlin M, *et al.* Differentiating pre- and minimally invasive from invasive adenocarcinoma using CT-features in persistent pulmonary part-solid nodules in Caucasian patients. *Eur J Radiol* 2015; 84: 738–744.
- 23 Rami-Porta R, Bolejack V, Crowley J, *et al.* The IASLC Lung Cancer Staging Project: proposals for the revisions of the T descriptors in the forthcoming eighth edition of the TNM classification for lung cancer. *J Thorac Oncol* 2015; 10: 990–1003.
- 24 Reeves AP, Biancardi AM, Apanasovich TV, *et al.* The Lung Image Database Consortium (LIDC): a comparison of different size metrics for pulmonary nodule measurements. *Acad Radiol* 2007; 14: 1475–1485.
- 25 Revel MP, Bissery A, Bienvenu M, *et al.* Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? *Radiology* 2004; 231: 453–458.
- 26 Devaraj A, van Ginneken B, Nair A, *et al.* Use of volumetry for lung nodule management: theory and practice. *Radiology* 2017; 284: 630–644.
- 27 Jeon KN, Goo JM, Lee CH, *et al.* Computer-aided nodule detection and volumetry to reduce variability between radiologists in the interpretation of lung nodules at low-dose screening computed tomography. *Invest Radiol* 2012; 47: 457–461.
- 28 Swensen SJ, Silverstein MD, Ilstrup DM, *et al.* The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med* 1997; 157: 849–855.
- 29 Chung K, Jacobs C, Scholten ET, *et al.* Lung-RADS category 4X: does it improve prediction of malignancy in subsolid nodules? *Radiology* 2017; 284: 264–271.
- 30 Bankier AA, MacMahon H, Goo JM, *et al.* Recommendations for measuring pulmonary nodules at CT: a statement from the Fleischner Society. *Radiology* 2017; 285: 584–600.
- 31 Mets OM, de Jong PA, Chung K, *et al.* Fleischner recommendations for the management of subsolid pulmonary nodules: high awareness but limited conformance – a survey study. *Eur Radiol* 2016; 26: 3840–3849.