Introduction

The use of low-dose computed tomography (LDCT) for screening of pulmonary nodules in at-risk population was shown to reduce lung cancer mortality by 20% as compared to screening with chest radiography (1). An estimated 4.8 million Americans are believed to have undergone at least one chest LDCT scan and 1.57 million were found to have nodules; of these, 63,000 were diagnosed with lung cancer within a period of 2 years (2). Lung cancer accounted for 17.1% of all newly diagnosed cancer cases and 21.7% of all cancer-related mortality in China (3). Therefore, early screening and treatment became a top priority for lung cancer (4). The experience of radiologists plays a major role in the detection and correct categorization of the pulmonary nodules on screening LDCT (5).

Although, there have many guidelines for reporting of pulmonary nodules, radiologists still have a huge difference...
in judging specific pulmonary nodules. Objectively, there are some nodules that are easily overlooked, especially for hilar nodules and ground-glass opacity nodules (6,7). In a study with LDCT, more than one third of nodules were missed at the baseline scan, and were detected at the follow-up scan 1 year later (8). A consensus double reading did increase the pulmonary nodules detectability of 19%, compared with a single reading (9).

There were a few studies dealing with the inter-reader variability of pulmonary nodules on CT screening (10-12). Inter-observer variability performance for the detection of pulmonary nodules has been found to be relatively high with both standard-dose and LDCT (13). However, the article on the analysis of causes of inconsistencies between radiologists has not yet been published. Our aim was to evaluate the consistency of radiologists in judging pulmonary nodules and its correlation with image features of the pulmonary nodules.

Methods

Imaging data

This retrospective study was approved by the Ethics Committee of Zhejiang Provincial People’s Hospital. The requirement for written informed consent of patients was waived off because of retrospective use of imaging data.

A total of 730 chest LDCT cases were randomly collected from three different medical centers (Zhejiang Provincial People’s Hospital, Zhejiang Hospital and Tongde Hospital) from the PACS system between June to July 2017. All patient sensitive information was eliminated from the data. Two radiology residents screened the 730 LDCT cases and eliminated 75 cases based on the following exclusion criteria: incomplete images; presence of motion or metal artifacts; history of chest surgery; diffuse pulmonary disease; presence of more than 10 pulmonary nodules. Then, 30 cases were randomly selected from the remaining 655 cases and reviewed together by three senior radiologists (each having more than 20 years’ experience) to familiarize with the operating system and reach a common sense of criteria.

Equipment and examination parameters

Chest CT images were acquired with a low-dose CT protocol on 64-slice multi-detector CT scanners (Somatom Definition AS, AS+ & Flash, Siemens, Germany; Optima 680, Discovery 750 HD, GE Healthcare, USA). The following CT protocol settings were used: 120 kV; 10–40 mA; gantry rotation speed: 0.5 seconds; helical scan mode (cranial-caudal direction) with a pitch of 1.2. The cumulative radiation dose was <1 mGy for each patient. Breath-holding lasted for 15 seconds to prevent image motion artifacts during the scan. Lung LDCT images were then reconstructed using a kernel of B75f at a slice thickness of 2 mm. All image readings were performed at a lung window width of 1,500 HU and a window level of -400 HU.

A dedicated system for lung nodule assessment was used. Professional reading monitors (Giant Shark, Nanjing Giant Shark Medical) were used to display LDCT images.

Evaluation of pulmonary nodules

Three experienced radiologists evaluated the 625 lung LDCT cases independently with no restrictions of time. Pulmonary nodules detected by each radiologist were recorded automatically. All detected nodules were categorized into three groups: nodules detected by all three radiologists (group I); nodules detected by two of the three radiologists (group II); and nodules detected by only one radiologist (group III). The flowchart was shown in Figure 1.

Two chest radiologists reviewed all nodules one by one, and recorded imaging features of each nodule presented by former three radiologists. Another expert made a final judgment in case of any controversy.

The following image features of the nodules were analyzed: size, density and location. The nodules were divided into two groups based on the diameter: less than 4 mm and greater than or equal to 4 mm (14). The location of the nodules was classified as follows: a peripheral nodule was defined as a nodule located within 2 cm of the pleura. A hilar nodule was located within 2 cm of the hilus. Central nodule was located between the peripheral and hilar zones (15). The nodules were classified as solid, subsolid, or calcified according to the nodule attenuation (16) (Figures 2-4).

Statistical analysis

Normally distributed continuous variables are expressed as mean ± standard deviation or as median and interquartile range (IQR), as appropriate. Categorical variables are expressed as percentage. Chi-square test was used to
to analyze the correlation between radiologist’s consistency and nodule characteristics. All analyses were performed using SPSS (version 18.0; Chicago, III). Statistical significance level was set at $P<0.05$.

Results

Descriptive statistics

A total of 625 chest LDCT cases were included in this study. The mean age of patients was $50.67\pm13.52$ years [range, 20–84 years; median: 63 years (interquartile range, 43–78 years)].

A total of 1,206 nodules were detected with an average of 1.9 (1,206/625) nodules per case. There were 234 (19.4%) nodules in group I, 377 (31.3%) nodules in group II, and 595 (49.3%) nodules in group III (Figure 5). Chi-square analysis revealed a significant correlation between groups and nodular parameters ($P<0.001$) (Table 1).

Nodule size

Among the 1,206 nodules, 1,070 (88.7%) were less than 4 mm, and only 136 (11.3%) were greater than or equal to 4 mm. There was a significant difference with respect to nodule size between groups I and II ($P<0.001$) and between groups I and III ($P<0.001$). No significant different was observed between groups II and III in this respect ($P=0.765$).

Nodule location

There were 37 (3.1%) nodules at the hilar zones, 188 (15.6%) at the central zone, 981 (81.3%) peripheral nodules. There was significant difference between group II and III ($P<0.001$) and between groups I and III ($P<0.001$) in terms of location. No significant different was observed between groups I and II in this respect ($P=0.765$).

Nodule density

A total of 266 (22.1%), 658 (54.6%) and 282 (23.4%) nodules were classified as subsolid, solid and calcified, respectively. There were significant differences with respect to nodule density type between groups I and II, groups I and III, groups II and III ($P<0.001$).

Figure 1 Flowchart for detecting pulmonary nodules. LDCT, low-dose computed tomography.

Figure 2 Nodules sized >4 mm.
Multiple logistic regression analysis

We selected group I, which was detected by all radiologists, as the positive group and group III was the negative group. The two groups were analyzed by multivariate logistic regression analysis. Logistic regression analysis of nodule size yielded an odds ratio of 0.053 [95% confidence interval (CI): 0.031–0.093; \( P < 0.001 \)]. There was a significant difference between categorization of central and peripheral nodules (\( P < 0.001 \)). There was a significant difference with respect to categorization of solid and calcified nodules (\( P < 0.001 \)) (Table 2).

A sub-analysis was performed for the nodules greater
Table 1 The correlation between consistency and imaging features of pulmonary nodules

<table>
<thead>
<tr>
<th>Features</th>
<th>Group I</th>
<th>Group II</th>
<th>Group III</th>
<th>Total</th>
<th>$\chi^2$</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size, mm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>194.737</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>&lt;4</td>
<td>147</td>
<td>357</td>
<td>566</td>
<td>1,070</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥4</td>
<td>87</td>
<td>20</td>
<td>29</td>
<td>136</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>57.336</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hilar</td>
<td>1</td>
<td>6</td>
<td>30</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>19</td>
<td>40</td>
<td>129</td>
<td>188</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peripheral</td>
<td>214</td>
<td>331</td>
<td>436</td>
<td>981</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>139.039</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Subsolid</td>
<td>40</td>
<td>79</td>
<td>147</td>
<td>266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solid</td>
<td>118</td>
<td>105</td>
<td>347</td>
<td>658</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcification</td>
<td>76</td>
<td>193</td>
<td>101</td>
<td>282</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>234</td>
<td>377</td>
<td>595</td>
<td>1,206</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Group I represents agreements of three, group II represents agreements of two, group III represents agreements of only one.

Table 2 A multivariate logistic regression between group I and group III

<table>
<thead>
<tr>
<th>Factors</th>
<th>B</th>
<th>SE</th>
<th>Wals</th>
<th>df</th>
<th>Sig</th>
<th>Exp(B)</th>
<th>95% CI for EXP (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>-2.930</td>
<td>0.283</td>
<td>106.914</td>
<td>1</td>
<td>0.000</td>
<td>0.053</td>
<td>0.031–0.093</td>
</tr>
<tr>
<td>Hilar</td>
<td>36.937</td>
<td>2</td>
<td>0.000</td>
<td>2</td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>-3.918</td>
<td>1.078</td>
<td>13.218</td>
<td>1</td>
<td>0.000</td>
<td>0.020</td>
<td>0.002–0.164</td>
</tr>
<tr>
<td>Peripheral</td>
<td>-1.582</td>
<td>0.308</td>
<td>26.425</td>
<td>1</td>
<td>0.000</td>
<td>0.206</td>
<td>0.112–0.376</td>
</tr>
<tr>
<td>Subsolid</td>
<td>22.598</td>
<td>2</td>
<td>0.000</td>
<td>2</td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Solid</td>
<td>-1.085</td>
<td>0.269</td>
<td>16.283</td>
<td>1</td>
<td>0.000</td>
<td>0.338</td>
<td>0.200–0.572</td>
</tr>
<tr>
<td>Calcification</td>
<td>-0.919</td>
<td>0.214</td>
<td>18.414</td>
<td>1</td>
<td>0.000</td>
<td>0.399</td>
<td>0.262–0.607</td>
</tr>
<tr>
<td>Constants</td>
<td>2.546</td>
<td>0.326</td>
<td>61.123</td>
<td>1</td>
<td>0.000</td>
<td>12.760</td>
<td></td>
</tr>
</tbody>
</table>

The subsolid nodule and hilar nodule as the reference categories.

than or equal to 4 mm. Group I was selected as positive group and group III as negative group. The two groups were analyzed by multivariate logistic regression analysis. There was a significant difference between categorization of central and peripheral nodules ($P<0.001$). There was no significant difference between categorization of solid nodules and calcified nodules ($P=0.292$) (Table 3).

**Discussion**

Pulmonary nodules are commonly detected on chest CT examination, with an incidence varies from 13% to 58% (7-10). The sensitivity of radiologists in detecting pulmonary nodules ranges from 64% to 82% based on LDCT (9,10). Due to the lack of gold standard, pulmonary nodules are typically judged by two or more doctors as a reference standard. In addition to the actual incidence of pulmonary nodules, doctors’ different detection and reporting of pulmonary nodules may be one of the possible reasons for the great variation. Samim et al. (17) reported the agreement between two radiologists with respect to identification of pulmonary nodules was 0.78 kappa value in patient-based analysis and 0.40 in lobe-based analysis. In the present study, only 19.4% of 1,206 nodules were
recognized by all three raters, although all three radiologists were experienced and underwent a common training before the study.

A previous study by Wormanns et al. (13) showed that the consistency of pulmonary nodules assessed by three radiologists was 47%. Our result presented a lower consistency of 19%. There were two differences between the two studies. First, we used clinical data, which include positive and negative cases, and Wormanns’ entire case was clearly positive in advance to readers; second, we use the number of more cases, radiologists need to pay more workload in the assessment.

This study explored the relationship between the consistency of pulmonary nodules detection and its image features. The results showed that the size, location and density of pulmonary nodules significantly affected the consistency of radiologists large diameter nodules, solid, calcified nodules and peripherally distributed pulmonary nodules have higher consistency, because these nodules are more easily detected by radiologists than small nodules, sub-solid nodules and inner distributed nodules, which is consistent with previous research and our work experience.

On sub-analysis for larger than or equal to 4 mm nodules, our results showed no significant difference in consistency among radiologists about the density of nodules. It means the consistency is less affected by the density for the larger nodules. However, the location was still the factor for the consistency.

We use 4 mm diameter as a threshold according the guidelines (18-24). We referred to the 2016 National Comprehensive Cancer Network guidelines (NCCN), 2017 Fleischer Society guidelines, the 2013 American College of Chest Physicians guidelines (ACCP) and the 2016 Clinical practice consensus guidelines for Asia. Four millimeters was the minimum threshold for nodules do not require routine follow-up.

In the present study, most nodules smaller than 4 mm were detected by only one radiologist. We speculate that radiologists tend to ignore small nodules which appear clinically insignificant. This selectivity exhibits greater randomness, which may be the main reason for the difference of detecting results. However, there were still some pulmonary nodules larger than 4 mm missing by radiologist. Most of these nodules were in the hilar location or sub-solid nodules. This shows that even if experienced radiologists, the possibility of missed detection still exists. Check by a second or third radiologist may help improve the accuracy of the diagnostic report. The artificial intelligence, which has developed rapidly in recent years, shows high sensitivity and accuracy to pulmonary nodules (25-27). AI’s pre-processing of images may become an assistant of radiologists in pulmonary detection.

There were several limitations in this study. First, none of the detected nodules were confirmed by surgery and pathology. Therefore, the total number of true nodules remains unknown. Second, the three radiologists belonged to different medical centers; although consistent training was carried out, there may still be differences in diagnostic habits. Third, the LDCT images in our study population were obtained with 2 mm section thickness, not the 1.25 mm from the guideline.

**Conclusions**

There was considerable inter-reader variability with respect to characterization of pulmonary nodules in LDCT. Larger nodules, solid or calcified nodules, and nodules in the outer zone were more likely to be consistently evaluated.
Acknowledgments

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This retrospective study was approved by the Ethics Committee of Zhejiang Provincial People’s Hospital. The requirement for written informed consent of patients was waived off because of retrospective use of imaging data.

References