Visual discrimination of screen-detected persistent from transient subsolid nodules: An observer study

Kaman Chung1*, Francesco Ciompi1, Ernst T. Scholten1, Jin Mo Goo2, Mathias Prokop1, Colin Jacobs3, Bram van Ginneken1, Cornelia M. Schaefer-Prokop1,3

1 Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, the Netherlands, 2 Department of Radiology, Seoul National University College of Medicine, and Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Korea, 3 Department of Radiology, Meander Medical Center, Amersfoort, the Netherlands

* kaman.chung10@gmail.com

Abstract

Purpose
To evaluate whether, and to which extent, experienced radiologists are able to visually correctly differentiate transient from persistent subsolid nodules from a single CT examination alone and to determine CT morphological features to make this differentiation.

Materials and methods
We selected 86 transient and 135 persistent subsolid nodules from the National Lung Screening Trial (NLST) database. Four experienced radiologists visually assessed a predefined list of morphological features and gave a final judgment on a continuous scale (0–100). To assess observer performance, area under the receiver operating characteristic (ROC) curve was calculated. Statistical differences of morphological features between transient and persistent lesions were calculated using Chi-square. Inter-observer agreement of morphological features was evaluated by percentage agreement.

Results
Forty-nine lesions were excluded by at least 2 observers, leaving 172 lesions for analysis. On average observers were able to differentiate transient from persistent subsolid nodules ≥ 10 mm with an area under the curve of 0.75 (95% CI 0.67–0.82). Nodule type, lesion margin, presence of a well-defined border, and pleural retraction showed significant differences between transient and persistent lesions in two observers. Average pair-wise percentage agreement for these features was 81%, 64%, 47% and 89% respectively. Agreement for other morphological features varied from 53% to 95%.

Conclusion
The visual capacity of experienced radiologists to differentiate persistent and transient subsolid nodules is moderate in subsolid nodules larger than 10 mm. Performance of the visual...
assessment of CT morphology alone is not sufficient to generally abandon a short-term follow-up for subsolid nodules.

Introduction

Results of lung cancer screening Computed Tomography (CT) studies revealed the importance of subsolid nodules as potential early adenocarcinomas. In the Early Lung Cancer Action Project (ELCAP) study the prevalence of subsolid nodules was found to be lower compared to solid nodules. However, subsolid nodules demonstrated a higher malignancy rate in the detected subsolid nodules of 34% (15/44) compared to 7% (14/189) for solid nodules [1]. Another study evaluating a group of clinically and screen-detected lesions even reported 81% (43/53) of resected subsolid nodules to be (pre)malignant [2].

The most frequent benign disease causing subsolid nodules is a focal infection [3, 4]. Other more rare underlying benign diseases are a focal organizing pneumonia or focal fibrosis [5, 6]. Subsolid nodules caused by infection will eventually disappear. Differentiation of transience versus persistence of subsolid nodules thus represents the first diagnostic task to discriminate between benign and potentially malignant lesions, and a short-term three months follow-up has been recommended by the Fleischner Society and the British Thoracic Society [7, 8]. The percentage of subsolid nodules detected in screening studies varied from 2% to 20% of all baseline screen-detected non-calcified nodules [1, 9, 10]. Prospective discrimination of transient from persistent lesions would therefore contribute to the reduction of follow-up CTs. Previous studies on this subject evaluated the contribution of texture analysis and clinical features, but did not assess human observer performance [11–13].

The only other morphological feature used for risk prediction of subsolid nodules besides persistence and lesion growth, is nodule size and the presence/size of a solid component [10, 14]. For solid nodules spiculation is an important predictor of malignancy in a recently published (screening) risk model [10]. However, for subsolid nodules no additional morphological features have been established. Defining morphological features for transient and persistent subsolid nodules would be a valuable first step.

The purpose of this study was therefore to evaluate whether and to which extent experienced radiologists would be able to differentiate transient from persistent subsolid nodules from a single CT examination by visual analysis alone. Secondly, we aimed to identify which morphological features are used by the radiologists to make this differentiation.

Materials and methods

Study population

We recruited subsolid nodules from the National Lung Screening Trial (NLST). The NLST was approved by the institutional board at each participating medical institution and participants provided written informed consent before randomization [15]. In total the NLST had 26,722 participants. Of those, 3194 participants had at least one subsolid nodule annotated by the NLST screening radiologist in any of the 3 screening rounds. Nine participants did not have any scans available, leaving 3185 participants for further analysis.

For this observer study, we used baseline (year 0) subsolid nodules only. The NLST annotations did not contain year-to-year linking between the same lesions, therefore we re-annotated all lesions by using information from the NLST database (slice number, nodule type, lobe location, size). Annotations were done by two medical students and one medical researcher using
in-house software (CIRRUS Lung Screening, Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, the Netherlands). A subsolid nodule was defined as transient if the nodule had disappeared on follow-up CT. A subsolid nodule was defined as persistent if the nodule remained visible on follow-up CT.

Subsequently we only selected CTs with a slice thickness of ≤ 2 mm, to ensure the quality of the coronal and sagittal projections of the lesions. As morphology is more difficult to assess in smaller lesions and thicker slices, we only selected lesions ≥ 10 mm (rounded average diameter) in this observer study. In total 232 subsolid lesions were eligible for our study. Eleven lesions (11/232, 4.7%) could not be located on the scans. Thus, our final dataset for the observer study contained 221 subsolid lesions.

**Observer study**

All study lesions were independently evaluated by four experienced radiologists (ETS, CSP, MP, and JMG). All of them had > 15 years of experience in reading chest CTs and had extensive experience with evaluating screen-detected nodules. Nodules were presented in a random order to each observer. Observers were asked to score the morphological nodule features using a predefined list. The list of morphological features as well as the definitions can be found in Table 1. In addition, they were asked to estimate the probability that the lesion was persistent on a scale between 0 and 100, with 0 representing certainly transient and 100 representing certainly persistent.

**Reading methodology**

A reading workstation designed to optimize workflow and to document the scoring data was used (CIRRUS Observer, Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, the Netherlands). After opening a case, a magnified axial view of the nodule under evaluation was centered in the middle of the display. Coronal/sagittal projections were available on the right side of the screen (Fig 1). The position of the magnified view was indicated by center lines on the smaller views. Using this set-up, observers did not have to search for the lesion. For all cases, the full 3D CT dataset was available for evaluation. All views could be (de)magnified. A standard lung window with a width of 1500 HU and a center of -650 HU was used as a default, but could be adjusted if necessary. Nodule diameters could be measured manually using electronic calipers.

The morphological features to be scored were listed on the left side of the monitor display. Scoring had to be completed before the next lesion could be displayed. Observers were allowed to place any comments if needed. No specifications with respect to comments were made prospectively. Lesions indicated in the comments for exclusion by 2 or more observers were omitted for further analysis. No information about follow-up appearance, persistence or any other outcome was provided.

**Statistical analysis**

Receiver Operating Characteristic (ROC) analysis was performed for each observer. Areas under the curve (AUC) and 95% confidence intervals (CI) were calculated to determine the ability to differentiate between transient and persistent lesions. We did not take into account within-participant correlation of participants with multiple nodules, because multiple nodules in a participant were considered as multiple independent nodules [16, 17]. Univariate analysis (Chi-square) per observer was used to assess whether a certain morphological feature was scored significantly different in transient or persistent nodules. P-values < 0.05 were considered significant. Inter-observer agreement for each CT morphological feature was investigated.
by calculating percentage agreement for each pair of observers. Statistical analyses were performed using SPSS, version 20.0 (SPSS, Chicago, Ill).

**Results**

**Study group**

Of all lesions 61% (135/221) were persistent. The median average diameter of persistent lesions was 12.0 mm (IQR 10.0–15.0 mm). Accordingly, 39% (86/221) were transient and had a median average diameter of 12.0 mm (IQR 11.0–16.5 mm). Forty-nine lesions were excluded from further data analysis because at least 2 of the 4 observers had made the comment that the opacification under review, which had been marked as subsolid nodule in the NLST database, would in fact not represent a nodular (subsolid) opacification when taking all three planes into consideration. Comments leading to exclusion were non-nodular (N = 11), solid lesion (N = 8), wall of emphysema (N = 3), apical scarring, (N = 7), fibrosis (N = 5) and (plate-like)
atelectasis (N = 15). Thus the final study group consisted of 172 subsolid lesions (101 persistent, 71 transient).

Discrimination of persistent from transient nodules

Observers 1 to 4 separately achieved an $A_z$ for discriminating persistent from transient subsolid nodules of 0.75 (95% CI 0.68–0.82), 0.75 (95% CI 0.67–0.82), 0.62 (95% CI 0.53–0.70) and 0.69 (95% CI 0.60–0.77), respectively (Fig 2).

Considering the score of 50 as a threshold for discriminating between transience (scores 0–50) and persistence (scores 51–100), the four observers correctly identified 58/71 (82%), 63/71 (89%), 51/71 (72%) and 55/71 (77%) transient lesions. The observers correctly identified 52/101 (51%), 37/101 (37%), 47/101 (47%) and 61/101 (60%) persistent nodules, respectively.

Taking the same thresholds for transience (0–50) and persistence (51–100), all four observers agreed on the same classification in 105 of the 172 nodules (61%). 68 of these 105 nodules (65%) were correctly classified, 37 of the 105 nodules (35%) were misclassified by all four observers. Thirty of the correctly classified nodules were persistent and 38 were transient. Figs 3 and 4 show examples of correctly and incorrectly identified lesions for which all or the majority of observers agreed on the classification.
Averaging the scores of the four observers resulted in an $A_z$ of 0.75 (95% CI 0.68–0.82) (Fig 5). Using the average scores a sensitivity of $> 90\%$ for persistent lesions was only achieved at the expense of a specificity of $< 30\%$ (e.g., sensitivity/specificity is 91% / 28%).

Morphology assessment: Univariate analysis

Morphological features that showed significant difference between transient and persistent in at least 2 observers are listed in Table 2. At a significance level of $p < 0.05$, nodule type and lesion margin were scored significantly different by 2 observers ($p = 0.016$ and $p = 0.025$, $p = 0.001$ and $p = 0.044$ respectively). Part-solid nodules were more often seen in persistent lesions compared to transient lesions in all observers, reaching statistical significance in two of them ($p = 0.016$ and $p = 0.025$). The subcategory of a well-defined border yielded significant difference in 2 observers ($p < 0.001$ and $p = 0.001$). Linear demarcation following the lobular border was the only feature in this category to be seen more often in transient lesions in three observers. Lobulated, spiculated and smooth borders were scored more often in persistent lesions. Pleural retraction was observed more frequently in persistent than transient lesions reaching significance in two observers ($p = 0.006$, $p = 0.037$).

Inter-reader variability of morphology

Average pair-wise percentage agreement was highest in external retraction, pleural retraction and bubble lucency (95%, 89%, and 86% respectively). Nodule type had an agreement of 81%

Fig 2. Receiver Operating Characteristic (ROC) curves for observer 1, 2, 3 and 4 to predict the persistence of the subsolid lesions $\geq 10 \text{ mm}$. $A_z$ (Areas Under the Curve) and 95% confidence interval in parenthesis, obs = observer.
followed by nodule multiplicity (73%), solid core margin (71%) and presence of an air-bronchogram (70%). Lower agreement was found in density of ground-glass component (67%), lesion margin (64%) and the subcategory of a well-defined solid core margin (62%). Lowest agreement was found for aspect of ground-glass component (53%) and the subcategory of a well-defined lesion margin (47%). The average pair-wise agreement and the 95% confidence intervals can be found in Table 3.
Fig 5. Receiver Operating Characteristic (ROC) curves for the average of all four observers. A_\text{z} (Area Under the Curve) and 95% confidence interval in parenthesis.

**Table 2. Univariate analyses.** Table describes morphological features with at least 2 observers in which the feature is seen significantly different between transient (T) and persistent (P) subsolid nodules using Chi-square. The total number of included nodules after exclusion is 172.

<table>
<thead>
<tr>
<th>Nodule type</th>
<th>T</th>
<th>P</th>
<th>P-value Obs. 1</th>
<th>T</th>
<th>P</th>
<th>P-value Obs. 2</th>
<th>T</th>
<th>P</th>
<th>P-value Obs. 3</th>
<th>T</th>
<th>P</th>
<th>P-value Obs. 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-solid</td>
<td>46</td>
<td>49</td>
<td></td>
<td>45</td>
<td>47</td>
<td></td>
<td>43</td>
<td>43</td>
<td></td>
<td>49</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>part-solid</td>
<td>23</td>
<td>52</td>
<td><strong>P = 0.016</strong></td>
<td>25</td>
<td>53</td>
<td><strong>P = 0.081</strong></td>
<td>27</td>
<td>55</td>
<td><strong>P = 0.064</strong></td>
<td>21</td>
<td>48</td>
<td><strong>P = 0.025</strong></td>
</tr>
<tr>
<td>other</td>
<td>2</td>
<td>0</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td>3</td>
<td></td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td><strong>Lesion Margin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ill-defined</td>
<td>29</td>
<td>30</td>
<td><strong>P = 0.130</strong></td>
<td>55</td>
<td>51</td>
<td><strong>P &lt; 0.001</strong></td>
<td>37</td>
<td>68</td>
<td><strong>P = 0.044</strong></td>
<td>38</td>
<td>39</td>
<td><strong>P = 0.053</strong></td>
</tr>
<tr>
<td>well-defined</td>
<td>30</td>
<td>71</td>
<td></td>
<td>16</td>
<td>50</td>
<td></td>
<td>34</td>
<td>33</td>
<td></td>
<td>33</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>If well-defined</td>
<td>N = 112</td>
<td>N = 66</td>
<td>N = 67</td>
<td>N = 96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>linearly demarcated</td>
<td>27</td>
<td>13</td>
<td></td>
<td>1</td>
<td>3</td>
<td></td>
<td>18</td>
<td>9</td>
<td></td>
<td>23</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>lobulated</td>
<td>1</td>
<td>15</td>
<td><strong>P &lt; 0.001</strong></td>
<td>8</td>
<td>28</td>
<td><strong>P = 0.813</strong></td>
<td>0</td>
<td>1</td>
<td><strong>P = 0.063</strong></td>
<td>7</td>
<td>20</td>
<td><strong>P = 0.001</strong></td>
</tr>
<tr>
<td>spiculated</td>
<td>5</td>
<td>14</td>
<td></td>
<td>1</td>
<td>4</td>
<td></td>
<td>0</td>
<td>2</td>
<td></td>
<td>3</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>smooth</td>
<td>8</td>
<td>29</td>
<td></td>
<td>7</td>
<td>14</td>
<td></td>
<td>15</td>
<td>22</td>
<td></td>
<td>1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Pleural Retraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>66</td>
<td>78</td>
<td><strong>P = 0.006</strong></td>
<td>71</td>
<td>96</td>
<td><strong>P = 0.057</strong></td>
<td>67</td>
<td>95</td>
<td><strong>P = 0.933</strong></td>
<td>68</td>
<td>87</td>
<td><strong>P = 0.037</strong></td>
</tr>
<tr>
<td>yes</td>
<td>5</td>
<td>23</td>
<td></td>
<td>0</td>
<td>5</td>
<td></td>
<td>4</td>
<td>6</td>
<td></td>
<td>3</td>
<td>14</td>
<td></td>
</tr>
</tbody>
</table>
The most frequent cause of transient subsolid nodules is focal infection. A persistent subsolid nodule, however, is potentially malignant and requires follow-up or alternative diagnostic work-up. A prospective estimation of whether the lesion would be persistent or transient would aid in reducing unnecessary follow-ups. This is the first study assessing the performance of human visual analysis for predicting the likelihood of persistence in subsolid nodules.

Results of our study indicate that experienced radiologists are at best only moderately able (average $A_z$ of all readings 0.75) to visually differentiate transient from persistent character in subsolid nodules $\geq 10$ mm. In addition the individual performance among the observers varied substantially with $A_z$ values ranging from 0.62 to 0.75. Given the variability among the observers, the moderate agreement and the imperfect performance of experienced radiologists, human visual analysis alone has to be considered insufficient to reproducibly predict if a subsolid nodule is persistent or transient. In that respect our results confirm published management strategies [7, 8] that recommend a 3-month follow-up CT for clarification of persistency.

A study by Lee HJ et al. [18] evaluated the performance of radiologists predicting benign and malignant subsolid nodules, a differentiation that might be less complex, since persistent lesions can be both benign and malignant and malignant lesions may expose more suggestive features. However, even with the availability of several clinical parameters (age, sex, pack years, history of lung cancer) and knowledge of predefined predictive CT information, an average $A_z$ value of 0.77 for non-solid and of 0.76 for part-solid nodules were achieved, thus in fact comparable to our results.

Secondly, we found that none of the morphology features yielded significant discrimination in all four observers. Most promising features were nodule type, lesion margin, presence of a well-defined lesion margin and pleural traction. The average pair-wise percentage agreement was relatively high in nodule type and pleural retraction (81% and 89% respectively). A considerably lower agreement, however, was found for features that had to be rated qualitatively such as lesion margin in general or the subcategory of a well-defined lesion margin (63% and 47%, respectively), indicating that these features do not appear to be sufficiently definable by visual analysis to serve as a broadly applicable criterion within a screening process.

### Table 3. Average pair-wise percentage agreement of the morphological features.

<table>
<thead>
<tr>
<th>Morphological feature</th>
<th>Average percentage agreement</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodule type</td>
<td>81%</td>
<td>79–84</td>
</tr>
<tr>
<td>Nodule multiplicity</td>
<td>73%</td>
<td>61–84</td>
</tr>
<tr>
<td>Lesion margin</td>
<td>64%</td>
<td>56–71</td>
</tr>
<tr>
<td>Subcategory of well-defined margins</td>
<td>47%</td>
<td>44–50</td>
</tr>
<tr>
<td>Solid core margin</td>
<td>71%</td>
<td>67–75</td>
</tr>
<tr>
<td>Subcategory of well-defined solid core margins</td>
<td>62%</td>
<td>59–65</td>
</tr>
<tr>
<td>Density of ground-glass component</td>
<td>67%</td>
<td>64–69</td>
</tr>
<tr>
<td>Aspect of ground-glass component</td>
<td>53%</td>
<td>45–61</td>
</tr>
<tr>
<td>Air bronchogram</td>
<td>70%</td>
<td>59–82</td>
</tr>
<tr>
<td>Bubble lucency</td>
<td>86%</td>
<td>82–89</td>
</tr>
<tr>
<td>Pleural retraction</td>
<td>89%</td>
<td>86–92</td>
</tr>
<tr>
<td>External retraction</td>
<td>95%</td>
<td>92–98</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0191874.t003
Interestingly however, when looking at the subcategory of a well-defined border, three observers scored linearly demarcated border more frequently in transient lesions (27/40, 18/27, and 23/39) compared to persistent lesions. We did not prospectively define whether the linear demarcation following the lobular border had to be present in several projections, which most likely contributed to the fact that one observer scored the feature only 4 times. The finding of linear demarcation shows similarity with a finding reported by Felix et al. [13]. Their study described a polygonal shape (defined “as a lesion with linear or concave margins at every corner”) as indicative for a transient lesion. Furthermore, they found that transient subsolid nodules were more frequently lobulated than persistent nodules. The finding of lobulation being predictive for transience reported by Felix et al. [13] is in contradiction to the other study by Lee SM et al. [19], who reported lobulation as indicative for malignancy. Similarly we found that 74% to 100% of the lobulated lesions were found to be persistent (15/16, 28/36, 1/1, 20/27 respectively).

In this study we selected the subsolid nodules following the nodule type annotations of the NLST database. Previous studies have shown that the agreement among radiologists is only moderate with regards to the differentiation of part-solid, non-solid and solid nodules [20, 21]. Therefore we decided to exclude all lesions that were considered not a subsolid nodule by at least 2 of the 4 experienced observers in our study, as indicated in their comments. We did so, to increase accuracy and reliability of the observer data.

Our study has some limitations. First, our study did not include any elaborate texture or quantitative analysis. Visual CT features in combination with elaborate objectively quantifiable measures might not only improve performance but also achieve a higher reproducibility. Second, we selected lesions \( \geq 10 \) mm only, taking into account the fact that the majority of the NLST CTs has not been reconstructed with 1 mm slice thickness, thus not providing isotropic high resolution image quality in all three projections. The level of performance and reader agreement we found, confirms the notion that visual assessment of morphological features in lesions \( < 10 \) mm will be even more difficult and less reliable. Last, the CT examinations of the NLST trial have been obtained with different scanners and variable slice thickness. Though only scans with a slice thickness of \( \leq 2 \) mm were included, the diverging image quality might have influenced the visual assessment of the nodules.

In conclusion, experienced radiologists are moderately able to determine persistent and transient nodule character in lesions \( \geq 10 \) mm visually. There are morphological features indicative for the discrimination of persistent and transient nodules, but none of them yielded significant discrimination in all four observers. Our results show that performance of the visual assessment of CT morphology alone is not sufficient to generally abandon a short-term follow-up and inter-reader variability plays a substantial role even among highly experienced observers.

Supporting information

S1 File. Main data file. Data file containing observers’ scores (probability that the lesion was persistent on a scale between 0–100 and morphology) and the ground truth for each case. (SAV)

Acknowledgments

The authors thank the National Cancer Institute for access to NCI’s data collected by the National Lung Screening Trial (project no. NLST-74). The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.
Author Contributions


Data curation: Kaman Chung, Francesco Ciompi, Ernst T. Scholten, Colin Jacobs.

Formal analysis: Kaman Chung, Francesco Ciompi, Colin Jacobs, Cornelia M. Schaefer-Prokop.

Funding acquisition: Cornelia M. Schaefer-Prokop.

Investigation: Kaman Chung, Francesco Ciompi, Bram van Ginneken.


Project administration: Kaman Chung.


Software: Francesco Ciompi, Colin Jacobs.

Supervision: Bram van Ginneken, Cornelia M. Schaefer-Prokop.

Visualization: Kaman Chung, Francesco Ciompi.

Writing – original draft: Kaman Chung, Cornelia M. Schaefer-Prokop.

Writing – review & editing: Kaman Chung, Francesco Ciompi, Ernst T. Scholten, Jin Mo Goo, Mathias Prokop, Colin Jacobs, Bram van Ginneken, Cornelia M. Schaefer-Prokop.

References


