

# Predicting Lung Cancer Risk of Incidental Solid and Subsolid Pulmonary Nodules in Different Sizes

This article was published in the following Dove Press journal:  
*Cancer Management and Research*

Rui Zhang<sup>1,\*</sup>  
Panwen Tian<sup>1,2,\*</sup>  
Bojiang Chen<sup>1</sup>  
Yongzhao Zhou<sup>1</sup>  
Weimin Li<sup>1</sup>

<sup>1</sup>Department of Pulmonary and Critical Care Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan, People's Republic of China; <sup>2</sup>Department of Lung Cancer Treatment Center, West China Hospital, Sichuan University, Chengdu, Sichuan, People's Republic of China

\*These authors contributed equally to this work

**Objective:** Malignancy prediction models for pulmonary nodules are most accurate when used within nodules similar to those in which they were developed. This study was to establish models that respectively predict malignancy risk of incidental solid and subsolid pulmonary nodules of different size.

**Materials and Methods:** This retrospective study enrolled patients with 5–30 mm pulmonary nodules who had a histopathologic diagnosis of benign or malignant. The median time to lung cancer diagnosis was 25 days. Four training/validation datasets were assembled based on nodule texture and size: subsolid nodules (SSNs)  $\leq 15$  mm, SSNs between 15 and 30 mm, solid nodules  $\leq 15$  mm and those between 15 and 30 mm. Univariate logistic regression was used to identify potential predictors, and multivariate analysis was used to build four models.

**Results:** The study identified 1008 benign and 1813 malignant nodules from a single hospital, and by random selection 1008 malignant nodules were enrolled for further analysis. There was a much higher malignancy rate among SSNs than solid nodules (rate, 75% vs 39%,  $P < 0.001$ ). Four distinguishing models were respectively developed and the areas under the curve (AUC) in training sets and validation sets were 0.83 (0.78–0.88) and 0.70 (0.61–0.80) for SSNs  $\leq 15$  mm, 0.84 (0.74–0.93) and 0.72 (0.57–0.87) for SSNs between 15 and 30 mm, 0.82 (0.77–0.87) and 0.71 (0.61–0.80) for solid nodules  $\leq 15$  mm, 0.82 (0.79–0.85) and 0.81 (0.76–0.86) for solid nodules between 15 and 30 mm. Each model showed good calibration and potential clinical applications. Different independent predictors were identified for solid nodules and SSNs of different size.

**Conclusion:** We developed four models to help characterize subsolid and solid pulmonary nodules of different sizes. The established models may provide decision-making information for thoracic radiologists and clinicians.

**Keywords:** lung cancer, subsolid nodule, solid nodule, prediction model

## Introduction

Globally, lung cancer continues to be the leading cause of cancer-related deaths in men and women.<sup>1,2</sup> Great progress has been made in knowing tumor biology. For example, as the incidence of adenocarcinoma rose to be the most common histologic subtype of lung cancer, new concepts were introduced, such as adenocarcinoma in situ (AIS) and minimally invasive adenocarcinoma (MIA), while invasive adenocarcinomas are classified by predominant patterns like lepidic, acinar, papillary, micropapillary, and solid patterns.<sup>3</sup> Moreover, molecular characteristics of lung

Correspondence: Weimin Li  
Department of Respiratory and Critical Care Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan Province 610041, People's Republic of China  
Email weimi003@yahoo.com

cancer were also identified, such as epidermal growth factor receptor insertions and deletions, KRAS mutations, ALK gene rearrangements, ROS1 translocations, PD-L1 expression, and so on.<sup>4</sup> It is confirmed that the updated grouping correlates with clinical outcomes, of which AIS and MIA showing an indolent clinical course with almost 100% curability.<sup>3</sup> Besides, the availability of targeted therapies based on these molecular markers also provides a more favorable prognosis.<sup>4</sup> However, despite knowledge gains in recognizing the disease, mortality from lung cancer remains high for most patients around the world. Both improvements in early detection and technological advances in genomics and genetics are necessary to ultimately improve lung cancer survival.<sup>2</sup>

At the early stage, lung cancers usually present as a solitary pulmonary nodule (round or oval opacity smaller than 30 mm in diameter that is completely surrounded by pulmonary parenchyma), which can be identified through thoracic computed tomography (CT) scans or radiographs.<sup>5</sup> The nodules are further classified into two categories, solid and subsolid nodules [SSNs, including pure ground-glass nodule (GGN) and part-solid GGN], respectively.<sup>6</sup> Both the US-based National Lung Screening Trial and Dutch–Belgian lung-cancer screening trial have shown that screening with the low-dose CT can reduce mortality from lung cancer.<sup>7,8</sup> Incidental pulmonary nodules, which are incidentally detected on a chest CT made for purposes other than lung cancer screening are increasing with an incidence that is much greater than recognized previously.<sup>9,10</sup> However, for either screen-detected or incidentally identified nodules, the major challenging is the definition of a positive result and the appropriate management of detected lung nodules.<sup>11,12</sup> A practical and accurate model that can predict the malignancy of a pulmonary nodule and that can be used to guide clinicians in clinic will be essential to reduce costs, radiation dose, and the risk of mortality in medical care.

Previous studies did establish clinical prediction models to estimate nodule malignancy. However, most models were developed based on all solitary pulmonary nodules without considering nodule texture.<sup>13–19</sup> Multiple sources of evidence have demonstrated that subsolid lung cancers are a fundamentally different disease than traditional solid lung cancers, with different cause, genetic pattern, and clinical behavior.<sup>20</sup> In addition, small-sized ( $\leq 15$  mm) and large-sized (15–30 mm) nodules also exhibit different features and need separate management.<sup>21</sup> Moreover, some models were based on screening nodules, which were different from incidental nodules encountered in routine

clinic.<sup>19,22,23</sup> Therefore, the current study intended to predict the lung cancer risk of incidental solid nodules and SSNs of different sizes ( $\leq 15$  mm and 15–30 mm) separately.

## Materials and Methods

### Patients Selection

Approval of the study was obtained from the institutional review board of the West China Hospital of Sichuan University. Informed consent was waived as this was a retrospective study and the privacy and identity information of the subjects were guaranteed. By reviewing electronic medical records, patients with suspicious pulmonary lesions were identified in West China Hospital from January 2010 to July 2017. Then, two thoracic specialists who were blinded to the pathological results of lesions evaluated all CT scans in consensus. The patient was included if there was an untreated, 5–30 mm, single, noncalcified solid nodule, or SSN on thin-section CT. Otherwise, patients were excluded based on the following criteria: a) the nodule was less than 5 mm (low risk) or there were multiple pulmonary nodules; b) pleural effusion, atelectasis, or lymph node enlargement was observed; or c) the pathological diagnosis was not clear or it was a metastatic tumor. Definitive diagnoses of all malignant and benign nodules were based on sputum cytology or pathologic examination of lung tissues obtained from thoracic surgery, bronchoscopy, or CT-guided percutaneous lung biopsy. The median time was 25 days (range=0–2287 days) from identification of the nodule to diagnosis of lung cancer.

### Clinical Variables

To explore the potential risk factors of solid nodules and SSNs, we collected clinical variables for further analysis. The demographic characteristics included age (years), sex (male or female), smoking (yes or no), history of malignancy (yes or no), and family history of lung cancer (yes or no). Besides, several laboratory tests on peripheral blood were collected, including red blood cells ( $10^{12}/L$ ), white blood cells ( $10^9/L$ ), blood platelets ( $10^9/L$ ), lymphocytes ( $10^9/L$ ), neutrophil to lymphocyte ratio, prothrombin time (PT, s), activated partial thromboplastin time (APTT, s), carcinoembryonic antigen (CEA, ng/mL), cytokeratin 19 fragment (CYFRA21-1, ng/mL), and neuron specific enolase (NSE, ng/mL). Moreover, results of the pulmonary function test were also recorded, including forced expiratory volume in 1 second (FEV1), the ratio of FEV1 to forced vital capacity (FEV1/FVC), maximal mid-

expiratory flow (MMEF), 50% of FVC (V50), 25% of FVC (V25), vital capacity (VC), the ratio of residual volume to total lung capacity (RV/TLC), diffusing capacity for carbon monoxide (DLCO), and body mass index (BMI). A few missing data of laboratory test and pulmonary function test were populated by a median value.

## Interpretation of CT Images

Two thoracic specialists reviewed the CT images independently and resolved discrepancies by consensus. All images were analyzed at both lung (width=1500 HU; level=-700 HU) and mediastinal (width=350 HU; level=40 HU) window settings. The following radiological features of each nodule were evaluated, including nodule diameter (maximum, mm), texture (subsolid or solid), location (upper right, middle right, lower right, upper left or lower left), shape (regular or irregular), spiculation (yes or no), and lobulation (yes or no).

## Statistical Analysis

The numerical data sets were presented with median (minimum, maximum; or 25th, 75th) and compared with Mann-Whitney *U*-test as the parameters did not coincide with normal distribution. The categorical data were described in a number of cases (proportion) and compared by contingency-table analyses. The study involved 2621 histopathology-confirmed 5–30 mm solitary pulmonary nodules (1008 benign and 1813 malignant). For further analysis, 1008 malignant nodules were randomly selected, equal to the benign ones. Then, all nodules were randomly divided into two groups at a ratio of 7:3. The larger subset (n=1412) was applied to develop risk prediction models, and the small subset (n=604) was used to validate models. Moreover, four training/validation datasets were assembled to build different models based on nodule texture and size: SSNs  $\leq 15$  mm, SSNs between 15–30 mm, solid nodules  $\leq 15$  mm, and solid nodules between 15–30 mm (Figure 1).

During the model-building process, univariate logistic regression was used to identify potential predictors and the variables that were significant at  $P < 0.05$  were selected. Then, multivariate logistic regression models were built for SSNs and solid nodules in different sizes. Moreover, independent predictors of each type of nodules were also identified during multivariate analysis.

The study evaluated the predictive performance of four models by assessing its discrimination (ability to classify correctly) and its calibration (whether probabilities

predicted by the model match observed probabilities). Discrimination was measured by using the area under the receiver-operating-characteristic curve (AUC), sensitivity and specificity, while the calibration was presented by the calibration curve. In addition, the decision-curve analysis was performed to quantify the net benefits of different threshold probabilities, thereby determining the clinical applicability of the models.<sup>24</sup> Except for the validation set, the AUC of established models were also evaluated by resampling. The statistical tests were all two-sided and differences with  $P < 0.05$  were considered statistically significant. All statistical analyses were conducted using SPSS 20.0 and R version 3.6.0.

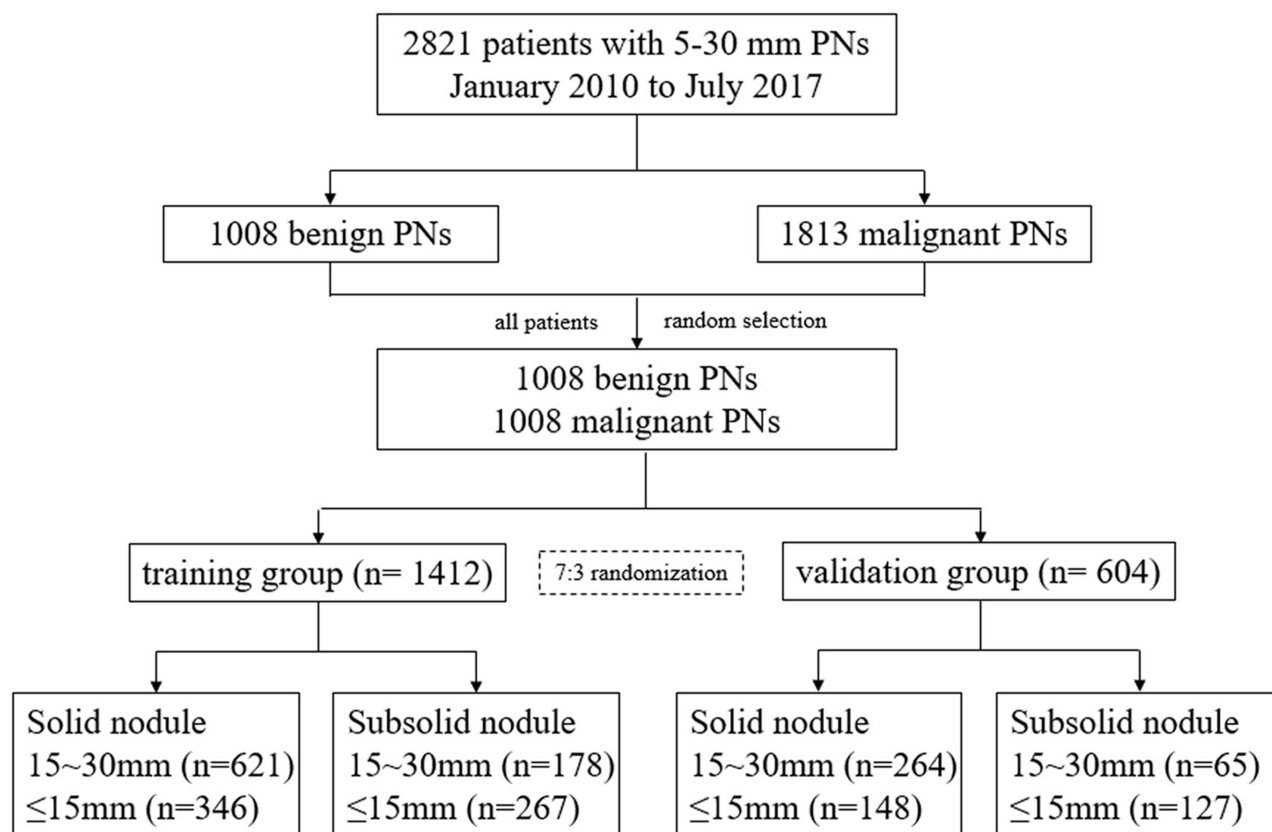
## Results

### Baseline Characteristics

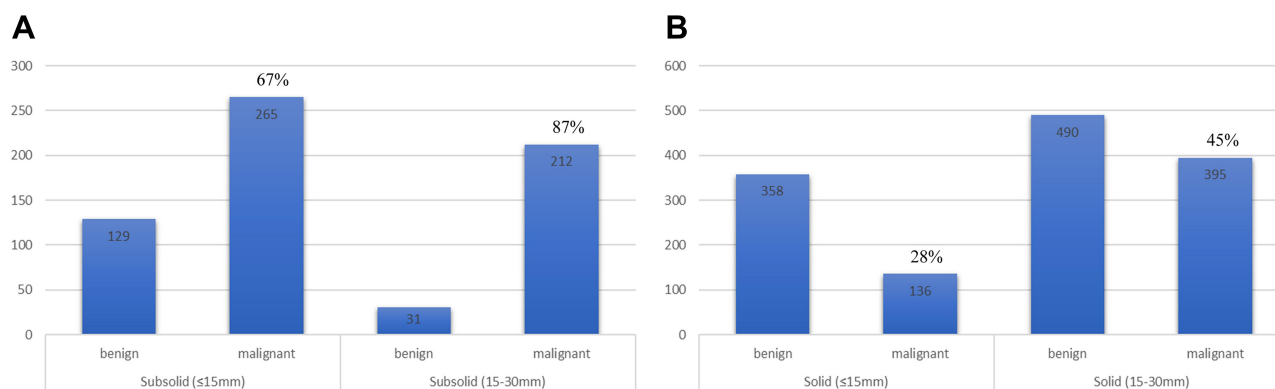
The study finally enrolled 2016 patients for further analysis, 55.5% female, 69.5% non-smokers, with a median age of 54 (quartile, 47–63). The median diameter of all nodules was 17 mm (quartile, 12–22), and 55.6% were located in the upper lobe. Besides, 31.6% of nodules were SSNs, while 68.4% were solid nodules. For SSNs  $\leq 15$  mm and those between 15–30 mm, the proportion of malignancy was 67% and 87%, respectively (total, 75%). In terms of solid nodules, the malignancy rate was 28% and 45%, respectively (total, 39%). Hence, there was a higher rate of malignancy for either smaller or larger SSNs ( $P < 0.001$ , Figure 2). The detailed comparative analyses of the benign and malignant pulmonary nodules for the training set and the validation set are summarized in Table 1. Histology and staging of the malignant nodules are shown in Supplementary Table 1.

### Four Models

As malignant lesions were predominant among SSNs, the logistic analysis for SSNs was to predict benign outcome, whereas for solid nodules the analysis predicted malignant outcome. Based on predictors selected from univariate analysis, four multivariable logistic-regression models were established and validated internally for SSNs and solid nodules of different sizes (Figure 3). The AUC of four models in the training set and the validation set was 0.83 (0.78–0.88) and 0.70 (0.61–0.80) for SSNs  $\leq 15$  mm, 0.84 (0.74–0.93) and 0.72 (0.57–0.87) for SSNs between 15–30 mm, 0.82 (0.77–0.87) and 0.71 (0.61–0.80) for solid nodules  $\leq 15$  mm, 0.82 (0.79–0.85) and 0.81 (0.76–0.86) for solid nodules between 15–30 mm, respectively



**Figure 1** Flow chart of enrolled patients.  
**Abbreviation:** PNs, pulmonary nodules.



**Figure 2** Proportion of malignant nodules in each group. (A) SSNs; (B) solid nodules.

(Table 2). In addition, the models also demonstrated stable performance in the resampled dataset (Supplementary Table 2). The calibration curves of the training set demonstrated that the predictions by each model showed good agreement with the actual observations. Besides, the decision curve analysis of the training set suggested that each model had a good overall net benefit.

## Independent Predictors

Independent predictors identified from multivariate analysis were different for SSNs and solid nodules in different size. For SSNs ≤15 mm, the predictors for a benign outcome were sex (OR=3.29, 95% CI=1.66–6.50), diameter (OR=0.76, 95% CI=0.67–0.85), location (OR=0.37, 95% CI=0.20–0.70), PT (OR=0.54, 95% CI=0.35–0.83), and NSE (OR=0.85, 95% CI=0.78–0.94; [Supplementary](#)

**Table I** Comparative Analyses of the Benign and Malignant Pulmonary Nodules in the Training and Validation Set

Variables	Training Set			Validation Set		
	Benign (N=706)	Malignant (N=706)	P	Benign (N=302)	Malignant (N=302)	P
Age	50 (17–80)	60 (30–87)	0.000*	51 (16–78)	59 (28–82)	0.000*
Sex			0.000*			0.006*
Female	322 (45.6)	440 (62.3)		162 (53.6)	195 (64.6)	
Male	384 (54.4)	266 (37.7)		140 (46.4)	107 (35.4)	
Smoking			0.006*			0.004*
Yes	245 (34.7)	197 (27.9)		102 (33.8)	70 (23.2)	
No	461 (65.3)	509 (72.1)		200 (66.2)	232 (76.8)	
History of malignancy			0.000*			0.069
Yes	31 (4.4)	64 (9.1)		11 (3.6)	21 (7.0)	
No	675 (95.6)	642 (90.9)		291 (96.4)	281 (93.0)	
Family history of lung cancer			0.065			0.158
Yes	37 (5.2)	54 (7.6)		13 (4.3)	21 (7.0)	
No	669 (94.8)	652 (92.4)		289 (95.7)	281 (93.0)	
Diameter	16 (5–30)	18 (5–30)	0.000*	16 (5–30)	17 (5–30)	0.065
Texture			0.000*			0.000*
Subsolid	114 (16.1)	331 (46.9)		46 (15.2)	146 (48.3)	
Solid	592 (83.9)	375 (53.1)		256 (84.8)	156 (51.7)	
Location			0.000*			0.397
Upper right lobe	199 (28.2)	255 (36.1)		96 (31.8)	100 (33.1)	
Middle right lobe	71 (10.1)	46 (6.5)		25 (8.3)	22 (7.3)	
Lower right lobe	155 (22.0)	112 (15.9)		59 (19.5)	56 (18.5)	
Upper left lobe	146 (20.7)	179 (25.4)		65 (21.5)	81 (26.8)	
Lower left lobe	135 (19.1)	114 (16.1)		57 (18.9)	43 (14.2)	
Shape			0.000*			0.000*
Irregular	422 (59.8)	564 (79.9)		171 (56.6)	239 (79.1)	
Regular	284 (40.2)	142 (20.1)		131 (43.4)	63 (20.9)	
Spiculation			0.000*			0.001*
Yes	258 (36.5)	401 (56.8)		115 (38.1)	154 (51.0)	
No	448 (63.5)	305 (43.2)		187 (61.9)	148 (49.0)	
Lobulation			0.000*			0.000*
Yes	286 (40.5)	392 (55.5)		118 (39.1)	170 (56.3)	
No	420 (59.5)	314 (44.5)		184 (60.9)	132 (43.7)	
BMI, kg/m <sup>2</sup>	23.0 (22.2–23.8)	23.1 (21.3–25.2)	0.624	23.0 (22.2–23.8)	23.1 (21.1–25.0)	0.982
FEV1%	102.6 (98.0–106.1)	101.5 (89.8–113.4)	0.322	102.6 (100.4–106.9)	102.0 (91.4–112.5)	0.088
FEV1/FVC	80.6 (78.6–82.0)	79.4 (74.2–83.7)	0.000*	80.6 (79.9–82.7)	80.3 (74.7–83.7)	0.025*
MMEF%	76.0 (68.0–83.0)	71.1 (52.6–89.2)	0.000*	76.0 (72.8–85.1)	73.5 (51.4–90.7)	0.004*
V50%	82.4 (73.6–89.8)	74.8 (56.5–76.8)	0.000*	82.4 (78.7–91.4)	78.5 (56.5–97.3)	0.007*
V25%	63.0 (54.5–71.1)	58.2 (41.0–76.8)	0.000*	63.0 (59.9–75.8)	57.8 (40.4–77.1)	0.000*
VC%	103.2 (100.5–106.8)	103.8 (94.3–113.6)	0.602	103.2 (100.8–108.0)	104.1 (95.5–113.0)	0.939
RV/TLC	36.2 (34.5–37.7)	40.6 (36.2–46.2)	0.000*	36.2 (35.0–38.4)	40.0 (36.0–44.2)	0.000*
DLC0%	99.2 (95.5–103.2)	96.9 (87.1–107.8)	0.002*	99.2 (94.6–101.6)	97.6 (87.9–108.5)	0.199
Red blood cell, 10 <sup>12</sup> /L	4.6 (2.8–7.6)	4.5 (2.4–6.7)	0.000*	4.6 (3.3–7.3)	4.5 (3.0–6.5)	0.017*
White blood cell, 10 <sup>9</sup> /L	5.6 (2.4–16.4)	5.7 (2.3–19.2)	0.376	5.6 (2.1–12.4)	5.7 (2.0–21.5)	0.744

(Continued)

**Table I** (Continued).

Variables	Training Set			Validation Set		
	Benign (N=706)	Malignant (N=706)	P	Benign (N=302)	Malignant (N=302)	P
Blood platelet, 10 <sup>9</sup> /L	173 (52–615)	172 (41–582)	0.912	178 (36–484)	172 (57–362)	0.227
Lymphocyte, 10 <sup>9</sup> /L	1.7 (0.4–4.3)	1.7 (0.5–10.3)	0.110	1.7 (0.8–4.3)	1.7 (0.4–3.4)	0.301
NLR	1.9 (0.6–21.8)	2.0 (0.3–14.9)	0.455	1.9 (0.5–12.8)	2.0 (0.1–21.4)	0.302
PT, s	11.1 (8.6–16.8)	11.2 (8.7–14.7)	0.017*	11.1 (9.0–13.6)	11.2 (9.0–23.9)	0.082
APTT, s	27.4 (18.8–59.6)	26.8 (17.8–45.9)	0.000*	27.4 (19.4–57.2)	26.8 (18.5–49.9)	0.039*
CEA, ng/mL	1.7 (0.2–30.9)	2.0 (0.2–164.5)	0.000*	1.7 (0.2–13.4)	2.0 (0.2–234.5)	0.000*
CYFRA21-1, ng/mL	1.9 (0.5–8.1)	2.0 (0.6–8.5)	0.000*	1.9 (0.5–6.6)	2.0 (0.6–10.4)	0.000*
NSE, ng/mL	12.2 (6.1–58.0)	13.4 (5.6–56.9)	0.000*	12.2 (7.2–53.9)	13.4 (6.8–48.5)	0.000*

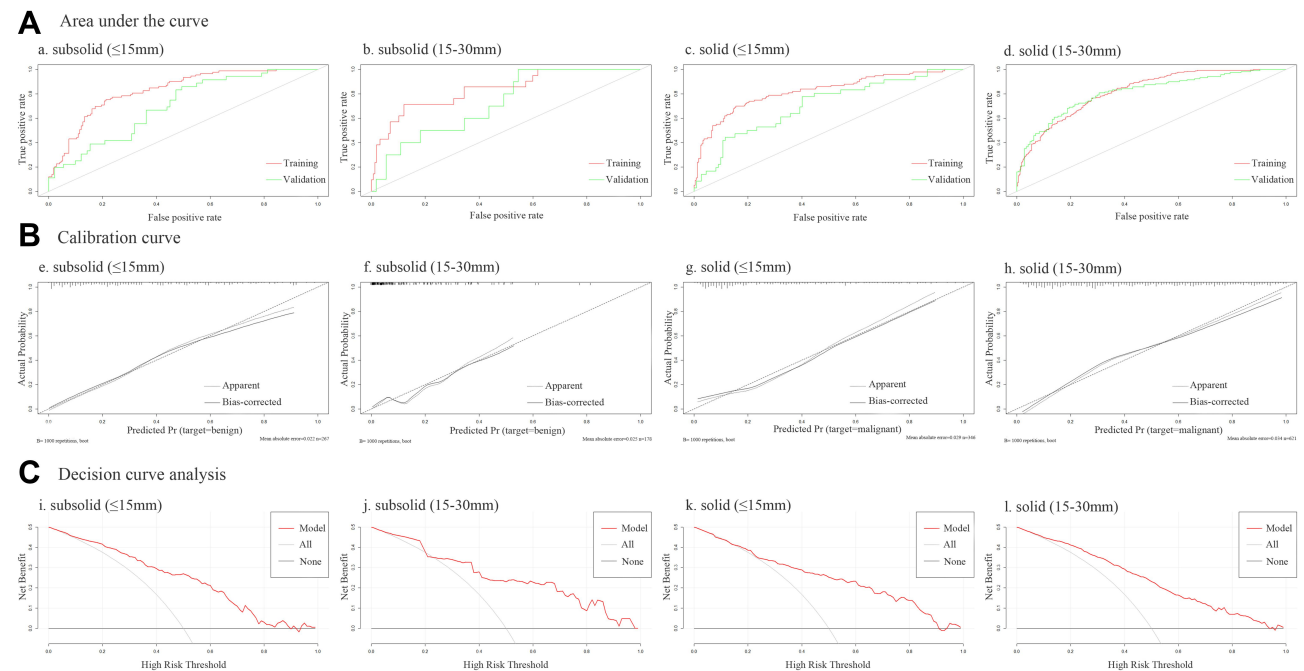
**Notes:** \*Significant variables ( $P < 0.05$ ); As there were outliers, variables of pulmonary function test were presented with median (25th, 75th); Other numerical data were presented with median (minimum, maximum).

**Abbreviations:** BMI, body mass index; FEV1, forced expiratory volume in 1 second; FEV1/FVC, the ratio of FEV1 to forced vital capacity (FVC); MMEF, maximal mid-expiratory flow; V50, the 50% of FVC; V25, the 25% of FVC; VC, the vital capacity; RV/TLC, the ratio of residual volume to total lung capacity; DLCO, diffusing capacity for carbon monoxide; NLR, neutrophil to lymphocyte ratio; PT, prothrombin time; APTT, activated partial thromboplastin time; CEA, carcinoembryonic antigen; CYFRA21-1, cytokeratin 19 fragment; NSE, neuron specific enolase.

**Figure 1).** Regarding SSNs between 15~30 mm, the independent predictor for a benign outcome was age alone (OR=0.92, 95% CI=0.87–0.98; [Supplementary Figure 2](#)).

For solid nodules ≤15 mm, the independent predictors for a malignant outcome were age (OR=1.03, 95% CI=1.00–1.07), shape (OR=2.50, 95% CI=1.21–5.18), spiculation (OR=3.03, 95% CI=1.61–5.69), V25 (OR=0.96, 95% CI=0.93–0.99), lymphocytes (OR=1.81, 95% CI=1.07–3.04), and CEA (OR=1.35, 95% CI=1.05–1.73; [Supplementary Figure 3](#)). With

respect to solid nodules between 15~30 mm, the independent predictors for a malignant outcome were age (OR=1.04, 95% CI=1.02–1.06), sex (OR=0.47, 95% CI=0.29–0.75), history of malignancy (OR=2.74, 95% CI=1.04–7.19), shape (OR=2.87, 95% CI=1.63–5.06), spiculation (OR=2.08, 95% CI=1.34–3.21), RV/TLC (OR=1.06, 95% CI=1.02–1.09), CEA (OR=1.14, 95% CI=1.03–1.26), CYFRA21-1 (OR=1.31, 95% CI=1.03–1.68), and NSE (OR=1.05, 95% CI=1.01–1.09; [Supplementary Figure 4](#)).



**Figure 3** Predictive performance of four models. **(A)** The receiver operating characteristic curve of models in the training set and the validation set; **(B)** The calibration curve of models in the training set; **(C)** The decision curve analysis of models in the training set.

**Table 2** Performance of Established Four Models

<b>A. For subsolid nodules</b>				
	<b>Subsolid Nodule <math>\leq 15</math> mm</b>		<b>Subsolid Nodule Between 15~30 mm</b>	
	<b>Training Group (N=267)</b>	<b>Validation Group (N=127)</b>	<b>Training Group (N=178)</b>	<b>Validation Group (N=65)</b>
N benign/malignant	93/174	36/91	21/157	10/55
Cutoff	0.40	0.40	0.22	0.22
AUC	0.83 (0.78–0.88)	0.70 (0.61–0.80)	0.84 (0.74–0.93)	0.72 (0.57–0.87)
Sensitivity	75.3%	61.1%	71.4%	40.0%
Specificity	78.7%	63.7%	87.9%	83.6%
Accuracy	77.2%	61.4%	85.4%	78.5%
<b>B. For solid nodules</b>				
	<b>Solid Nodule <math>\leq 15</math> mm</b>		<b>Solid Nodule Between 15~30 mm</b>	
	<b>Training Group (N=346)</b>	<b>Validation Group (N=148)</b>	<b>Training Group (N=621)</b>	<b>Validation Group (N=264)</b>
N benign/malignant	246/100	112/36	346/275	144/120
Cutoff	0.37	0.37	0.42	0.42
AUC	0.82 (0.77–0.87)	0.71 (0.61–0.80)	0.82 (0.79–0.85)	0.81 (0.76–0.86)
Sensitivity	70.0%	52.8%	76.7%	72.5%
Specificity	84.6%	73.2%	71.7%	75.0%
Accuracy	79.8%	68.2%	73.9%	73.9%

## Discussion

It is sure that subsolid lung cancers are a fundamentally different disease from solid lung cancers. By studying solitary pulmonary nodules, we confirmed that there was a much higher malignancy rate among SSNs than solid nodules. Therefore, four prediction models were respectively established for each type of nodules, which have shown high diagnostic yields. Different independent predictors were also identified in each subgroup.

A higher likelihood of malignancy for SSNs was also observed in previous studies.<sup>20,25,26</sup> For example, Henschke et al<sup>25</sup> found malignancy was diagnosed in 34% of SSNs as opposed to a 7% for solid nodules ( $P < 0.05$ ) and the malignancy rate for part-solid nodules was even higher (63%). Similarly, Nakata et al<sup>26</sup> demonstrated a dramatic malignancy rate of 93.3% for part-solid nodules. While transient SSNs have been mainly caused by inflammation, most persistent SSNs have been preinvasive lesions (AAH, AIS), MIA, or invasive adenocarcinomas.<sup>27</sup> Compared to solid lung cancers, malignant SSNs usually present an indolent natural course, some of which could stay stable for several years.<sup>20</sup> Histologically, solid components of SSNs were usually related to histologic subtypes like papillary, acinar, solid, or invasive mucinous adenocarcinoma.<sup>3</sup> Radiogenomically, Li et al<sup>28</sup>

found that higher ratios of solid components in SSNs were accompanied by significantly higher mutation frequencies in EGFR, TP53, RBM10, and ARID1B, suggesting that these genes play roles in the progression of lung adenocarcinomas. Anyway, proper management of SSNs is critical.

Currently available clinical lung cancer risk prediction models were basically established from solitary pulmonary nodules, including both SSNs and solid nodules. The well-known models are the Gurney model,<sup>16</sup> Mayo model,<sup>15</sup> Herder model,<sup>14</sup> VA model,<sup>18</sup> PKUPH model,<sup>13</sup> Brock model, as well as the TREAT model.<sup>17,19</sup> Nevertheless, the high malignancy rate and indolent course of SSNs (different from solid nodules) call for different risk prediction models for SSNs and solid nodules. Zheng et al<sup>29</sup> have established models for incidental SSNs with a proportion of ground glass opacity less than 50% (predictors: age, presence of symptoms, total protein; diameter, lobulation, calcified nodes; AUC, 0.86) or  $\geq 50\%$  (predictors: sex, FEV1%; diameter, calcified nodes; AUC=0.84). However, as an acknowledged marker for benign lesions, calcification contributed a lot to the performance of their models, which can be reflected from the variable coefficient. Differently, the current study has developed models for non-calcified SSNs less than

15 mm (AUC=0.83) or between 15~30 mm (AUC=0.84), which showed comparative performance and can help clinicians from another aspect.

With respect to solid nodules, the acknowledged BIMC model is to assess probability of malignancy in solid solitary pulmonary nodules and indicates that nodule size, enhancement, morphology, and VDT are the best predictors of malignancy (AUC=0.89).<sup>30</sup> Besides, another three models were also established for solid nodules based on a Chinese population (AUC=0.85~0.87).<sup>31-33</sup> Compared to these four models, the current solid-nodule models shared some similar risk factors, such as age, sex, history of malignancy, morphology, and serum CEA, but some novel markers like V25, RV/TLC, serum lymphocyte, CYFRA21-1, and NSE were also identified. As for model performance, these four were a little better than our models (AUC=0.82), which can be associated with some valuable predictors they enrolled, such as enhancement, VDT, as well as maximum uptake value of nodules. Besides, one study also enrolled calcification in the model.<sup>33</sup> Anyway, all these solid-nodule models demonstrated good performance, and clinicians can choose the most appropriate one according to the variable availability and nodule similarity where models were derived.

It is worthy noticing that a computational approach such as DL (machine learning method based on artificial neural networks) and radiomics (high-throughput mining of quantitative image features from standard-of-care medical imaging) has shown promise in pulmonary nodule detection and classification.<sup>34</sup> Mao et al<sup>21</sup> established a quantitative radiomic model for predicting malignancy of screen-detected small solid pulmonary nodules (6–15 mm) and found that the accuracy was high (AUC=0.97). Similarly, when differentiating focal organizing pneumonia and solitary granulomas from solid adenocarcinomas, the radiomics model also demonstrated great performance with an AUC of 0.96 and 0.94, respectively.<sup>35,36</sup> Besides, Gong et al<sup>37</sup> have investigated radiomic features between benign and malignant SSNs and found the average AUCs of classifiers were 0.55~0.93. Hence, the radiomics models were of higher accuracy than current clinical models, showing exciting and potential future applications to lung cancer.

Estimating the risk of lung cancer is important in three common clinical scenarios: the management of pulmonary nodules, the selection of people for lung cancer screening, and in the early identification of symptomatic disease.<sup>38</sup> The models in the current study were to predict nodule risk

of malignancy based on a complete representative hospital series, thus to provide evidence for pulmonary nodule management in routine clinical practice. Hence, there is a need to understand the potential population in which the models can be used and its' preferred to perform validation studies in the population to ensure that models reliably predict the chance of malignancy.

The present study exhibited some limitations. Firstly, this was a retrospective, single-center study with inevitable selection bias. Besides, to establish the risk models, several missing data were populated by a median value, which could be another potential reason for bias. Secondly, the models were established and internally validated on patients with single SSNs or solid nodules, thus the use of our model excludes the opportunity to predict lung cancer risk of multiple pulmonary nodules. Similarly, the models were not suitable for metastatic lung tumors. Thirdly, some valuable predictors were not evaluated due to limited data, such as VO<sub>2</sub> max, VDT, and positron emission tomography/computed tomography.<sup>39,40</sup>

In conclusion, incidental pulmonary nodules are increasing and a growing body of evidence is emerging to support distinguishing management of SSNs and solid pulmonary nodules. Therefore, the current study has developed logistic models to respectively predict malignancy risk of incidental solid and subsolid pulmonary nodules of different sizes (less than 15mm or 15mm or greater). Different independent predictors were also identified for each subgroup. Hopefully, the non-invasive and convenient models can provide decision-making information for thoracic radiologists and clinicians to properly manage pulmonary nodules in routine clinical practice.

## Ethical Statement

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy and confidentiality of any part of the work are appropriately investigated and resolved. The work was conducted in accordance with the Declaration of Helsinki.

## Author Contributions

All authors made substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work.



## Funding

This work was supported by the National Key Development Plan for Precision Medicine Research (2017YFC0910004) and Chengdu Science and Technology Program Projects (2017-CY02-00030-GX).

## Disclosure

The authors report no conflicts of interest in this work.

## References

- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7–34.
- Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. *Ann Glob Health*. 2019;85(1).
- Travis WD, Brambilla E, Noguchi M, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol*. 2011;6(2):244–285.
- Reck M, Rabe KF. Precision diagnosis and treatment for advanced non-small-cell lung cancer. *N Engl J Med*. 2017;377(9):849–861.
- Winer-Muram HT. The solitary pulmonary nodule. *Radiology*. 2006;239(1):34–49.
- 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(1):304–317.
- National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395–409.
- de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med*. 2020;382(6):503–513.
- 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(10):1208–1214.
- Van't Westeinde SC, de Koning HJ, Xu DM, et al. How to deal with incidentally detected pulmonary nodules less than 10mm in size on CT in a healthy person. *Lung Cancer*. 2008;60(2):151–159.
- Tanoue LT, Tanner NT, Gould MK, et al. Lung cancer screening. *Am J Respir Crit Care Med*. 2015;191(1):19–33.
- 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(1):228–243.
- Li Y, Wang J. A mathematical model for predicting malignancy of solitary pulmonary nodules. *World J Surg*. 2012;36(4):830–835.
- Herder GJ, van Tinteren H, Golding RP, et al. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. *Chest*. 2005;128(4):2490–2496.
- 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(8):849–855.
- Gurney JW, Lyddon DM, McKay JA. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part II. Application. *Radiology*. 1993;186(2):415–422.
- Deppen SA, Blume JD, Aldrich MC, et al. Predicting lung cancer prior to surgical resection in patients with lung nodules. *J Thorac Oncol*. 2014;9(10):1477–1484.
- Gould MK, Ananth L, Barnett PG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest*. 2007;131(2):383–388.
- 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(10):910–919.
- Mase VJ, Dettlerbeck FC. Approach to the subsolid nodule. *Clin Chest Med*. 2020;41(1):99–113.
- Mao L, Chen H, Liang M, et al. Quantitative radiomic model for predicting malignancy of small solid pulmonary nodules detected by low-dose CT screening. *Quant Imag Med Surg*. 2019;9(2):263–272.
- Tammemagi M, Ritchie AJ, Atkar-Khattra S, et al. Predicting malignancy risk of screen-detected lung nodules—mean diameter or volume. *J Thorac Oncol*. 2019;14(2):203–211.
- Tammemagi MC, Freedman MT, Pinsky PF, et al. Prediction of true positive lung cancers in individuals with abnormal suspicious chest radiographs: a prostate, lung, colorectal, and ovarian cancer screening trial study. *J Thorac Oncol*. 2009;4(6):710–721.
- Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak*. 2008;8:53.
- 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(5):1053–1057.
- Nakata M, Saeki H, Takata I, et al. Focal ground-glass opacity detected by low-dose helical CT. *Chest*. 2002;121(5):1464–1467.
- Kakinuma R, Noguchi M, Ashizawa K, et al. Natural history of pulmonary subsolid nodules: a prospective multicenter study. *J Thorac Oncol*. 2016;11(7):1012–1028.
- Li Y, Li X, Li H, et al. Genomic characterisation of pulmonary subsolid nodules: mutational landscape and radiological features. *Eur Respir J*. 2020;55(2).
- Zheng B, Zhou X, Chen J, et al. A modified model for preoperatively predicting malignancy of solitary pulmonary nodules: an Asia cohort study. *Ann Thorac Surg*. 2015;100(1):288–294.
- Soardi GA, Perandini S, Motton M, et al. Assessing probability of malignancy in solid solitary pulmonary nodules with a new Bayesian calculator: improving diagnostic accuracy by means of expanded and updated features. *Eur Radiol*. 2015;25(1):155–162.
- Chen X-B, Yan R-Y, Zhao K, et al. Nomogram for the prediction of malignancy in small (8–20 mm) indeterminate solid solitary pulmonary nodules in Chinese populations. *Cancer Manag Res*. 2019;11:9439–9448.
- Guo H-Y, Lin J-T, Huang H-H, et al. Development and validation of a 18F-FDG PET/CT-based clinical prediction model for estimating malignancy in solid pulmonary nodules based on a population with high prevalence of malignancy. *Clin Lung Cancer*. 2020;21(1):47–55.
- She Y, Zhao L, Dai C, et al. Development and validation of a nomogram to estimate the pretest probability of cancer in Chinese patients with solid solitary pulmonary nodules: a multi-institutional study. *J Surg Oncol*. 2017;116(6):756–762.
- Kim TJ, Kim CH, Lee HY, et al. Management of incidental pulmonary nodules: current strategies and future perspectives. *Expert Rev Respir Med*. 2019;14(2):173–194.
- Yang X, He J, Wang J, et al. CT-based radiomics signature for differentiating solitary granulomatous nodules from solid lung adenocarcinoma. *Lung Cancer*. 2018;125:109–114.
- Zhang T, Yuan M, Zhong Y, et al. Differentiation of focal organising pneumonia and peripheral adenocarcinoma in solid lung lesions using thin-section CT-based radiomics. *Clin Radiol*. 2019;74(1):78.e23–78.e30.
- Gong J, Liu J, Hao W, et al. Computer-aided diagnosis of ground-glass opacity pulmonary nodules using radiomic features analysis. *Phys Med Biol*. 2019;64(13):135015.
- Baldwin DR. Prediction of risk of lung cancer in populations and in pulmonary nodules: significant progress to drive changes in paradigms. *Lung Cancer*. 2015;89(1):1–3.

39. Divisi D, Di Francesco C, Di Leonardo G, et al. Preoperative pulmonary rehabilitation in patients with lung cancer and chronic obstructive pulmonary disease. *Eur J Cardio-Thorac*. 2012;43(2):293–296.
40. Divisi D, Barone M, Zaccagna G, et al. Fluorine-18 fluorodeoxyglucose positron emission tomography in the management of solitary pulmonary nodule: a review. *Ann Med*. 2017;49(7):626–635.

### Cancer Management and Research

Dovepress

### Publish your work in this journal

Cancer Management and Research is an international, peer-reviewed open access journal focusing on cancer research and the optimal use of preventative and integrated treatment interventions to achieve improved outcomes, enhanced survival and quality of life for the cancer patient.

The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/cancer-management-and-research-journal>