

# Development and validation of a prediction model for malignant pulmonary nodules

## A cohort study

Zhen Ren, MD<sup>a,b</sup>, Hongmei Ding, MD<sup>a,b</sup>, Zhenzhen Cai, MD<sup>a,b</sup>, Yuan Mu, PhD<sup>a,b</sup>, Lin Wang, MD<sup>a,b</sup>, Shiyang Pan, MD, PhD<sup>a,b,\*</sup>

### Abstract

This study is to develop and validate a preoperative prediction model for malignancy of solitary pulmonary nodules. Data from 409 patients who underwent solitary pulmonary nodule resection at the First Affiliated Hospital of Nanjing Medical University, China between June 2018 and December 2020 were retrospectively collected. Then, the patients were nonrandomly split into a training cohort and a validation cohort. Clinical features, imaging parameters and laboratory data were then collected. Logistic regression analysis was used to develop a prediction model to identify variables significantly associated with malignant pulmonary nodules (MPNs) that were then included in the nomogram. We evaluated the discrimination and calibration ability of the nomogram by concordance index and calibration plot, respectively. MPNs were confirmed in 215 (52.6%) patients by a pathological examination. Multivariate logistic regression analysis identified 6 risk factors independently associated with MPN: gender (female, odds ratio [OR] = 2.487; 95% confidence interval [CI]: 1.313–4.711;  $P = .005$ ), location of nodule (upper lobe of lung, OR = 1.126; 95%CI: 1.054–1.204;  $P < .001$ ), density of nodule (pure ground glass, OR = 4.899; 95%CI: 2.572–9.716;  $P < .001$ ; part-solid nodules, OR = 6.096; 95%CI: 3.153–14.186;  $P < .001$ ), nodule size (OR = 1.193; 95%CI: 1.107–1.290;  $P < .001$ ), GAGE7 (OR = 1.954; 95%CI: 1.054–3.624;  $P = .033$ ), and GBU4–5 (OR = 2.576; 95%CI: 1.380–4.806;  $P = .003$ ). The concordance index was 0.86 (95%CI: 0.83–0.91) and 0.88 (95%CI: 0.84–0.94) in the training and validation cohorts, respectively. The calibration curves showed good agreement between the predicted risk by the nomogram and real outcomes. We have developed and validated a preoperative prediction model for MPNs. The model could aid physicians in clinical treatment decision making.

**Abbreviations:** AAB = autoantibody, BPN = benign pulmonary nodule, CEA = carcinoembryonic antigen, CI = confidence interval, C-index = concordance index, CYFRA21-1 = cytokeratin 19 fragment, LDCT = low-dose computed tomography, MPN = malignant pulmonary nodule, NSE = neuron-specific enolase, OR = odds ratio, SPN = solitary pulmonary nodule, TAA = tumor-associated antigen.

**Keywords:** lung neoplasms, nomogram, risk factors, solitary pulmonary nodules

## 1. Introduction

Lung cancer is a major public health problem worldwide and it has the highest incidence and mortality among all malignant

tumors in China.<sup>[1]</sup> It is well known that early diagnosis and treatment of lung cancer can prolong life span and improve prognosis of patients.

With the widespread use of low-dose computed tomography (LDCT) examination which is an important method to screen early lung cancer, a frequently reported incidence of solitary pulmonary nodules (SPNs) has shown a significantly increasing trend in recent years.<sup>[2]</sup> SPN is a term used to describe single, round, well-circumscribed radiological opacity <3 cm in diameter.<sup>[3]</sup> However, in all cases, malignant pulmonary nodules (MPNs) account for <10% of these nodules, it shows that LDCT has a high false-positive rate.<sup>[4]</sup> This increases the psychological burden of patients with SPN and makes clinical decision-making still at empirical decisive period. Cancer remains undiagnosed as malignant or benign except for pathological examination.<sup>[5]</sup> Some invasive pathological examination such as surgical resection, fine needle biopsy, and bronchoscopy can identify the benign and malignant SPNs, but there are risks of respiratory impairment, pneumothorax, and bleeding, it should be avoided especially in the case of benign tumors. Currently, a number of studies on the risk factors related to MPN have been carried out: The risk factors include clinical features, serum tumor markers, imaging features, and gene tags.<sup>[6–9]</sup> Traditional biomarkers such as carcinoembryonic antigen (CEA), cytokeratin 19 fragment (CYFRA21-1), neuron-specific enolase (NSE) are widely used in the establishment of prediction model. However, the predictive

Editor: Maya Saranathan.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

<sup>a</sup> Department of Laboratory Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China, <sup>b</sup> National Key Clinical Department of Laboratory Medicine, Jiangsu Province Hospital, Nanjing Medical University, Nanjing, China.

\* Correspondence: Shiyang Pan, Department of Laboratory Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, China (e-mail: sypan@njmu.edu.cn).

Copyright © 2021 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Ren Z, Ding H, Cai Z, Mu Y, Wang L, Pan S. Development and validation of a prediction model for malignant pulmonary nodules: a cohort study. *Medicine* 2021;100:51(e28110).

Received: 3 June 2021 / Received in final form: 28 October 2021 / Accepted: 16 November 2021

<http://dx.doi.org/10.1097/MD.00000000000028110>

efficacy of these markers is not ideal, false-positive results often occur as a result of infection, benign tumors, pregnancy, and other factors.<sup>[10]</sup> It is important to find new potential markers to adapt to the fast pace of current clinical settlements. Tumor-associated autoantibodies (AABs) develop in response to tumor antigens and may be found in the plasma of asymptomatic lung cancer patients. Previous study found that an AABs panel has been validated in different screening cohorts to assess lung cancer risk, and may help differentiate between benign and malignant lesions.<sup>[11–15]</sup> Moreover, AABs samples are easily accessible and have been used for early lung cancer screening in clinic. Here we chose 7 AABs (SOX2, GAGE7, CAGE, MAGEA1, P53, GBU4–5, and PGP9.5) to assess their predictive value.

The purpose of this study was to identify clinical variables significantly associated with the risk of MPN and develop and validate a new clinical prediction model for MPN before SPN resection based on imaging, clinical characteristics, and laboratory parameters.

## 2. Materials and methods

### 2.1. Study design and participants

We retrospectively searched the radiological imaging system of the First Affiliated Hospital of Nanjing Medical University from June 2018 to December 2020 to identify patients who were diagnosed with SPN and underwent surgical resection and pathological examination. All the patients had no obvious pulmonary symptoms and SPNs were found unintentionally in the process of physical examination or other disease examination. The inclusion criteria were as follows: the final diagnoses were confirmed with histopathologic diagnosis based on tissue obtained from surgical resection; no extrapulmonary malignancy; and complete clinical, CT image, and laboratory data. The exclusion criteria were as follows: history of lung cancer treatment; history of other cancers; and incomplete laboratory data. Data on SPN patients collected from June 2018 to May 2020 were used as the training dataset, and data on SPN patients collected from June 2020 to December 2020 were used as the validation dataset. The current study was approved by the Institutional Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (No: 2020-SR-153).

### 2.2. Clinicopathological variables

Clinical features were collected including age of the patient, gender, smoking history. Imaging parameters mainly included SPN site (upper, middle or lower), diameter and density. All patients received a routine preoperative laboratory examination within 2 wk before SPN surgery that included whole blood count, coagulation function, serum tumor markers tests [CEA, CYFRA21-1, NSE, thymidine kinase1], and serum AABs tests (SOX2, GAGE 7, CAGE, MAGE, P53, GBU4-5, PGP9.5). For the derivative indicator involved, the neutrophil–lymphocyte ratio was calculated as follow: neutrophil–lymphocyte ratio = neutrophil count/lymphocyte count. Serum CEA, CYFRA21-1, and NSE were measured by electrochemiluminescence immunoassays using a Cobas e602 automated analyzer (Roche, Germany). Thymidine kinase1 was detected by chemiluminescence digital imaging system (Biovica, Sweden). A Sysmex XN series automated hematology analyzer (Sysmex, Japan) and a Sysmex CS5100 automated blood coagulation analyzer (Sysmex,

Japan) was used to determine the complete blood count and coagulation function, respectively. The concentrations of AABs were quantitated by enzyme linked immunosorbent assay (KAIBAOLUO, China).

### 2.3. Statistical analysis

Categorical variables are displayed as the number and percentage, and continuous variables are presented as the median (interquartile range). Categorical variables were compared using the chi-square test or Fisher exact test. As the continuous variables did not conform to the normal distribution, comparisons of these variables between 2 different groups were conducted using the Mann–Whitney test. A univariate logistic regression analysis was used to assess the significance of each variable in the training cohort for the prediction of MPN. All variables with  $P < .05$  in the univariate logistic regression analysis and other predictors with clinical relevance were incorporated into a multivariate logistic regression analysis. The nomogram for the prediction of MPN was established based on the results of the multivariate logistic regression analysis by using the rms package of R, version 4.0.4 (<http://www.r-project.org/>). To evaluate the prediction performance of the nomogram, we calculated the concordance index (C-index) with 1000 bootstrap samples to measure discrimination [the model's ability to distinguish between benign pulmonary nodule (BPN) and MPN patients and generated calibration plots to measure calibration consistency between the predicted probability and observed frequency of patients with MPN]. The optimal cutoff value of the nomogram was determined by maximizing the Youden index. All statistical analyses were performed using SPSS software, version 22 (SPSS, Inc., Chicago, IL) and R, version 4.0.4 (<http://www.r-project.org/>). This report followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines.<sup>[16]</sup>

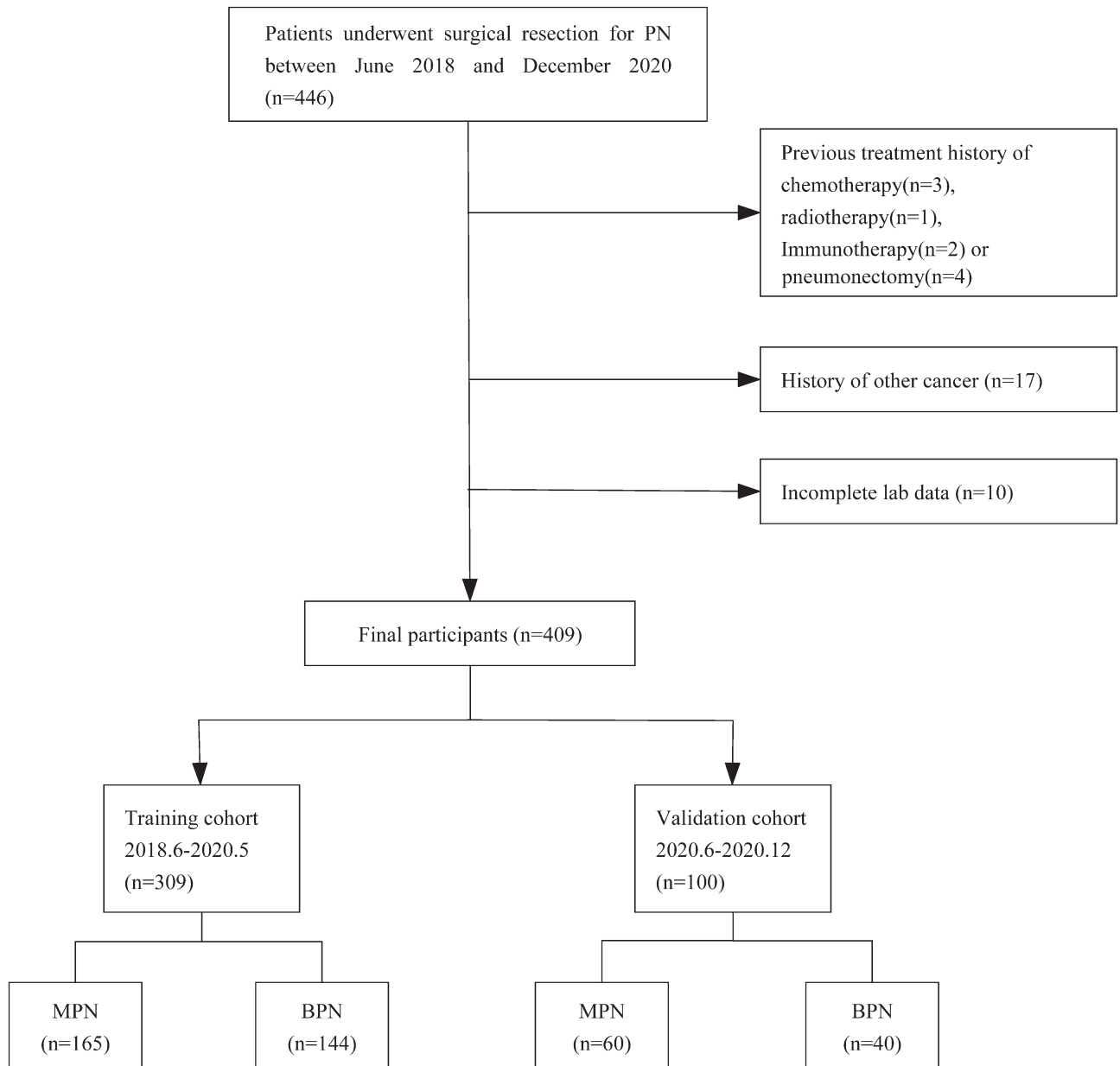
## 3. Results

### 3.1. Characteristics of the training and validation cohorts

In total, 446 SPN patients were included in this retrospective Study. Ultimately, 409 patients met the inclusion criteria. Among them, 309 patients whose data were collected between June 2018 and May 2020 formed the training cohort, and 100 patients whose data were collected between June and December 2020 formed the validation cohort (Fig. 1). The clinicopathologic characteristics of the patients are summarized in Table 1. The median ages of patients in the training and validation cohorts were 56 and 55 years, respectively. One hundred sixty-five (40.3%) patients were men and 225 (55.0%) patients were diagnosed as MPNs. The histological examination confirmed MPN in 165 (53.4%) patients in the training cohort and 60 (60.0%) patients in the validation cohort. There was no significant difference in the distribution of variables between the training and validation cohorts except for P53, hemoglobin, platelet distribution width.

### 3.2. Predictors selection

The results of the univariate logistic regression analysis of the clinical features in the training cohort are shown in Table 2. Gender (odds ratio [OR] = 1.465; 95% confidence interval [CI]:



**Figure 1.** Flow chart of the study population. BPN = benign pulmonary nodule, MPN = malignant pulmonary nodule.

0.915–2.344;  $P=.039$ ), SOX2 (OR=1.055; 95%CI=1.006–1.107;  $P=.027$ ), GAGE7 (OR=1.070; 95%CI=1.010–1.134;  $P=.021$ ), GBU4–5 (OR=1.107; 95%CI=1.050–1.167;  $P<.001$ ), NSE (OR=0.954; 95%CI=0.912–0.997;  $P=.037$ ), nodule size (OR=1.044; 95%CI=1.013–1.075;  $P=.005$ ), location of nodule (OR=2.717; 95%CI=1.355–3.476;  $P=.001$ ), nature of nodule (for pure ground glass vs solid, OR=7.037; 95%CI:3.845–12.879;  $P<.001$ ; for part-solid nodules vs solid, OR=11.913; 95%CI:5.957–23.822;  $P<.001$ ) were significant preoperative risk factors associated with MPN in the univariate analysis. All these predictors with a  $P$  value  $<.05$  as well as P53, hemoglobin, platelet distribution width which show difference between the training and validation cohorts and CEA, CYFRA21-1 which were mentioned in previous established model were selected for the multivariate analysis. In the multivariate

analysis, female (OR=2.487; 95%CI: 1.313–4.711;  $P=.005$ ), upper lobe of lung (OR=1.126; 95%CI: 1.054–1.204;  $P<0.001$ ), density of nodule (for pure ground glass vs solid, OR=4.899; 95%CI: 2.572–9.716;  $P<.001$ ; for part-solid nodules vs solid, OR=6.096; 95%CI: 3.153–14.186;  $P<.001$ ), nodule size (OR=1.193; 95%CI: 1.107–1.290;  $P<.001$ ), GAGE7 (OR=1.954; 95%CI: 1.054–3.624;  $P=.033$ ) and GBU4–5 (OR=2.576; 95%CI: 1.380–4.806;  $P=0.003$ ) were independently associated with the presence of MPN (Table 3).

### 3.3. Development and validation of a nomogram for preoperative MPN prediction

Based on the results of the multivariate analysis, we chose gender, tumor size, location of nodule, density of nodule, GAGE7, and

**Table 1****Comparison of participant characteristics in the training and validation cohorts.**

Characteristic	Training cohort (n = 309)	Validation cohort (n = 100)	P value
Median age (IQR), yr	56 (48,63)	55 (46,64)	.741
Gender (%)			
Male	123 (39.8)	42 (42.0)	.726
Female	186 (60.2)	58 (58.0)	
Smoking history (%)			
Yes	54 (17.5)	11 (11.0)	.156
No	255 (82.5)	89 (89.0)	
Family history (%)			
Yes	12 (3.9)	5 (5.0)	.575
No	297 (96.1)	95 (95.0)	
P53, U/mL	1.00 (0.30, 2.65)	0.73 (0.20, 2.58)	.043
PGP9.5, U/mL	0.30 (0.00, 1.90)	0.20 (0.00, 1.50)	.161
SOX2, U/mL	0.50 (0.10, 1.90)	0.70 (0.10, 1.70)	.531
GAGE7, U/mL	0.80 (1.40, 2.80)	0.60 (1.15, 2.98)	.209
GBU4-5, U/mL	1.90 (0.30, 15.75)	2.20 (0.30, 11.03)	.382
MAGE, U/mL	0.10 (0.00, 0.20)	0.10 (0.00, 0.10)	.664
CAGE, U/mL	0.00 (0.00, 0.10)	0.00 (0.00, 0.10)	.912
TK1, pmol/L	0.19 (0.07, 0.45)	0.36 (0.06, 0.45)	.174
CEA, ng/mL	1.74 (1.05, 2.69)	1.58 (1.00, 2.71)	.638
CYFRA21-1, ng/mL	1.81 (1.42, 2.39)	1.83 (1.46, 2.49)	.519
NSE, ng/mL	16.17 (14.00, 19.34)	15.78 (13.80, 18.64)	.373
WBC, (10 <sup>9</sup> /L)	5.54 (4.76, 6.67)	5.42 (4.79, 6.71)	.818
NLR	1.68 (1.29, 2.16)	1.75 (1.28, 2.19)	.887
HB, g/L	133 (125, 145)	128 (121, 146)	.031
PLT, 10 <sup>9</sup> /L	201 (166, 235)	210 (169, 237)	.204
PDW, %	13.60 (12.00, 16.31)	11.05 (10.80, 11.90)	.039
MPV, fL	10.80 (10.00, 11.70)	10.55 (10.13, 14.74)	.495
DD, mg/L	0.19 (0.13, 0.36)	0.21 (0.13, 0.38)	.431
Nodule size, mm	12.70 (9.00, 18.00)	13.05 (9.00, 18.53)	.095
Location of nodule (%)			.635
Upper	162 (54.4)	57 (57.0)	
middle	23 (7.4)	5 (5.0)	
Lower	124 (40.2)	38 (38.0)	
Nature of nodule (%)			.487
Solid	112 (36.2)	34 (34.0)	
Pure ground glass	104 (33.7)	37 (37.0)	
Part-solid nodules	93 (30.1)	29 (29.0)	

CEA = carcinoembryonic antigen, CYFRA21-1 = cytokeratin 19 fragment, DD = D-dimer, HB = hemoglobin, IQR = interquartile range, MPV = mean platelet volume, NLR = neutrophil-lymphocyte ratio, NSE = neuron-specific enolase, PDW = platelet distribution width, PLT = platelets, RDW = red blood cell distribution width, TK1 = thymidine kinase 1, WBC = white blood cells.

GBU4-5 for model development. The nomogram for predicting MPN in patients preoperatively is presented in Figure 2. The probability of MPN can be estimated by using this nomogram to calculate the total points for each patient. Further analysis indicated that the nomogram has excellent performance in distinguishing MPN from BPN. We generated calibration curves to evaluate the calibration of the prediction model. Calibration curves demonstrated acceptable model calibration, with good agreement between the observed frequency and predicted probability of patients with MPN in both datasets (Fig. 3).

In the training cohort, the C-index was 0.86 (95%CI: 0.83–0.91), and in the validation cohort, the C-index was 0.88 (95%CI: 0.84–0.94). According to the maximum Youden index, the optimal cutoff value for the prediction probability of the nomogram was 0.54. The sensitivity, specificity, positive predictive value, and negative predictive value when the model was used to differentiate between MPN and BPN were 80.1%,

**Table 2****Univariate logistic regression analysis of preoperative data of patients with malignant pulmonary nodules in the training cohort.**

Variable	OR (95%CI)	P value
Gender, female vs male	1.465 (0.915–2.344)	.039
Smoking history, yes vs no	1.482 (0.785–2.801)	.225
Family history, yes vs no	1.005 (0.947–1.086)	.892
P53, U/mL	1.047 (1.002–1.094)	.059
PGP9.5, U/mL	1.058 (0.941–1.190)	.343
SOX2, U/mL	1.055 (1.006–1.107)	.027
GAGE7, U/mL	1.070 (1.010–1.134)	.021
GBU4-5, U/mL	1.107 (1.050–1.167)	<.001
MAGE, U/mL	1.072 (0.964–1.193)	.200
CAGE, U/mL	1.001 (0.950–1.054)	.985
TK1, pmol/L	0.836 (0.670–1.044)	.115
CEA, ng/mL	1.054 (0.945–1.175)	.348
CYFRA21-1, ng/mL	1.063 (0.864–1.307)	.565
NSE, ng/mL	0.954 (0.912–0.997)	.037
WBC, (10 <sup>9</sup> /L)	0.816 (0.702–0.948)	.068
NLR	0.812 (0.645–1.021)	.074
HB, g/L	0.984 (0.969–1.000)	.052
PLT, 10 <sup>9</sup> /L	0.999 (0.995–1.003)	.469
PDW, %	1.004 (0.925–1.091)	.917
MPV, fL	1.062 (0.905–1.245)	.461
DD, mg/L	0.732 (0.484–1.108)	.140
Nodule size, mm	1.044 (1.013–1.075)	.005
Location of nodule, upper vs middle and lower	2.717 (1.355–3.476)	.001
Nature of nodule		
Pure ground glass vs solid	7.037 (3.845–12.879)	<.001
Part-solid nodules vs solid	11.913 (5.957–23.822)	<.001

CEA = Carcinoembryonic antigen, CI = Confidence interval, CYFRA21-1 = Cytokeratin 19 fragment, DD = D-Dimer, HB = Hemoglobin, MPV = Mean platelet volume, NLR = neutrophil-lymphocyte ratio, NSE = Neuron-specific enolase, OR = Odds ratio, PDW = platelet distribution width, PLT = Platelets, RDW = Red blood cell distribution width, TK1 = Thymidine kinase 1, WBC = White blood cells.

75.8%, 77.4%, and 72.2%, respectively, in the training cohort and 74.0%, 73.5%, 84.1%, and 75.5%, respectively, in the validation cohort (Table 4).

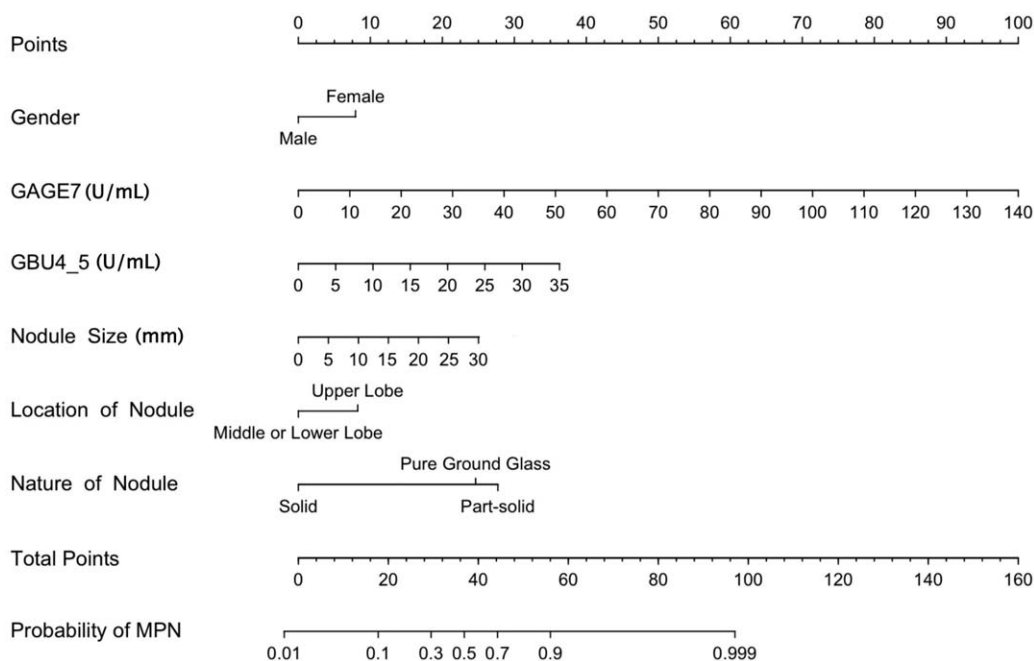
#### 4. Discussion

In this study, we conducted retrospective analysis of individual clinical features, image and laboratory data of 409 newly diagnosed SPN patients. Then a novel prediction model in predicting MPNs was developed by using multivariate logistic regression analysis. The identified novel prediction model could

**Table 3****Multivariate logistic regression analysis of preoperative data of patients with malignant pulmonary nodules in the training cohort.**

Variable	β	OR (95%CI)	P value
Gender, female vs male	0.911	2.487 (1.313–4.711)	.005
GAGE7, U/mL	0.670	1.954 (1.054–3.624)	.033
GBU4-5, U/mL	0.946	2.576 (1.380–4.806)	.003
Nodule size, mm	0.176	1.193 (1.107–1.290)	<.001
Location of nodule, upper vs middle or lower	0.119	1.126 (1.054–1.204)	<.001
Nature of nodule			
Pure ground glass vs solid	1.501	4.899 (2.572–9.716)	<.001
Part-solid vs solid	1.682	6.096 (3.153–14.186)	<.001

CI = confidence interval, OR = odds ratio.



**Figure 2.** Nomogram for predicting MPN preoperatively in patients with PN. When using the nomogram, find the position of each variable on the axis and the corresponding point vertically. Then, add the points of all variables, and determine the prediction probability of MPN on the bottom axis. MPN = malignant pulmonary nodule.

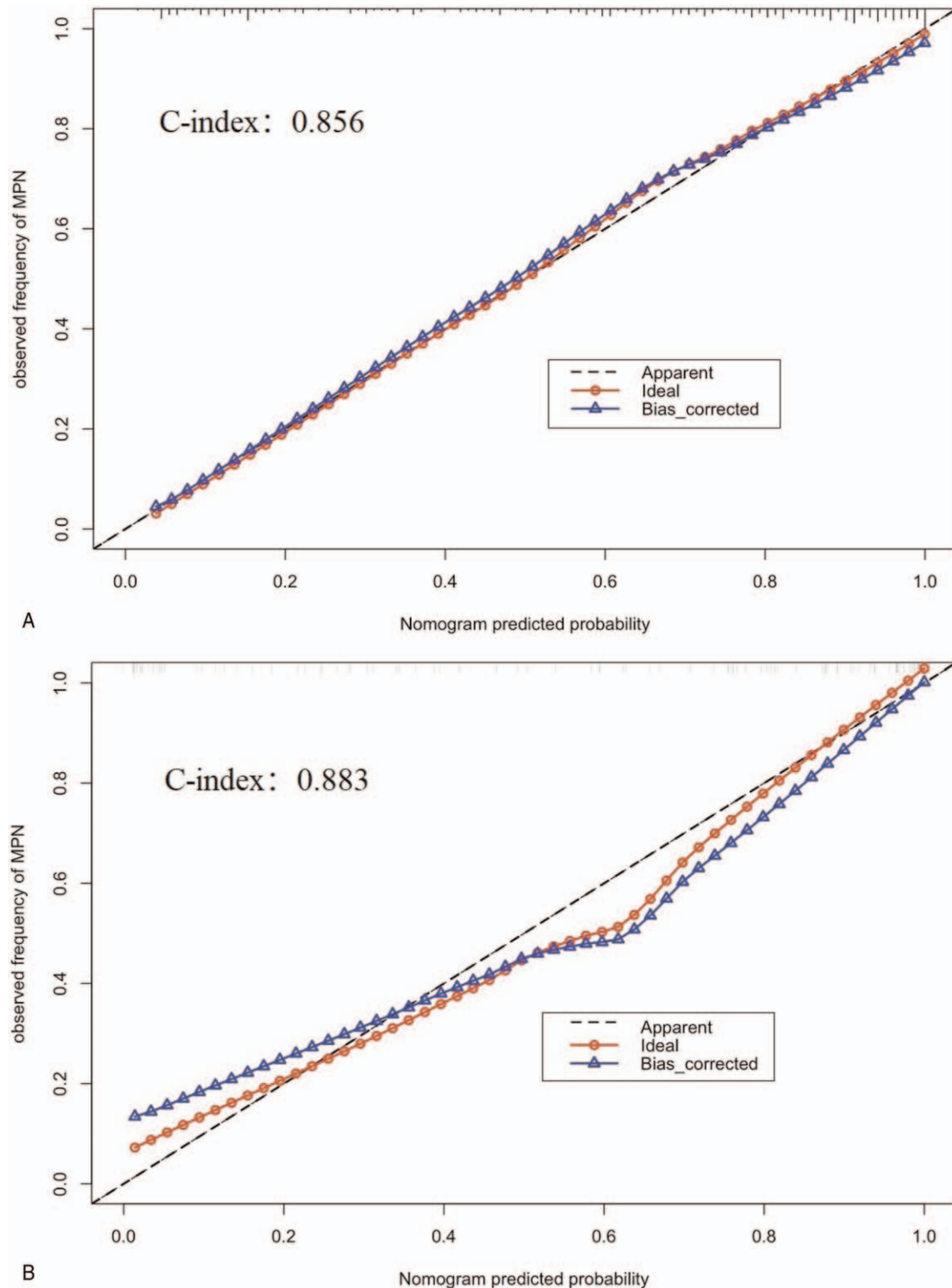
successfully classify the SPN patients into BPNs and MPNs and showed good agreement between the predicted probability and actual frequency of MPN.

Lung cancer is the number one cause of cancer death worldwide with 1.76 million associated deaths reported in 2018.<sup>[17]</sup> The key issue in the fight against this disease is the detection and diagnosis of all SPNs at an early stage. The benign nodules account for more than 90% of all SPNs.<sup>[18]</sup> The remaining MPN: primary lung cancer or metastatic malignant tumor is the focus of our early screening. From February 2015, LDCT screening entered the armamentarium of diagnostic tools broadly available to individuals at high-risk of developing lung cancer.<sup>[19]</sup> However, the high positive rate entails a burden of the diagnostic work-up, clinicians mainly make decisions based on image information and personal experience judgment. Excessive treatment or missed diagnosis of early lung cancer due to long-term follow-up are 2 major problems in clinical decision-making of diagnosis and treatment of SPN. Therefore, how to accurately predict MPN to formulate treatment plan is the main problem faced by clinicians.

Previous studies have showed that predictors, which serve to assess the lung cancer risk or constitute the components to build the lung cancer risk prediction models, can be categorized into the following groups: clinical/epidemiological (smoking history, age, family history, spirometry, emphysema); radiological (SPN features: diameter/volume, spiculation, lobulation, location in a lung lobe, relation to pulmonary fissures, calcification pattern; parameters derived by mathematically advanced image analysis); biochemical/genomic/epigenomic (protein and genomic validated clusters); and others: sputum cytology (cell-CT analysis), exhaled breath analysis.<sup>[20]</sup> Many clinical prediction models that combine clinical features, laboratory parameters, and imaging characteristics have been established to make accurate MPN predictions.

Although our model was also built from these 3 aspects, we applied AAB that is limited in current reports. In our report, female, nodule size, location, and density, GAGE7, GBU4-5 were identified as independent risk factors significantly associated with MPN.

Tumor-associated antigens (TAAs) produced by tumor cells during their development include tumor/testis-specific antigens, aberrantly-overexpressed or ectopically-expressed antigens, and stem cell transcriptional factors, etc.<sup>[21,22]</sup> Based on the immunoeediting theory, TAAs are captured by the immune system and lead to the formation of AABs via humoral immune responses. In our study, GAGE7 and GBU4-5 were independent risk factors of MPN. They are both cancer/testis antigens, which do not express in other somatic tissues except the testis and malignant tumors. Serological analysis of tumor antigens by recombinant cDNA expression cloning (SEREX) is a technology which aims at the identification of TAAs of potential diagnostic or therapeutic use.<sup>[23]</sup> GBU4-5 is a protein that recently identified by using SEREX technologies. It is a special helicase, which plays an important role in carcinogenesis because of its specificity and immunogenicity and may provide diagnostic and potentially immunotherapeutic cancer targets.<sup>[24]</sup> GBU4-5 encodes a DEAD-box domain which can elicit an autoantibody response.<sup>[25,26]</sup> These DEAD-box-containing proteins are involved in RNA processing, ribosome assembly, spermatogenesis, embryogenesis and cell growth and division which may well be important in the carcinogenic pathway.<sup>[27]</sup> GAGE gene belongs to a family of genes organized in clustered repeats. They have a high degree of predicted sequence identity, but differ by scattered single nucleotide substitution.<sup>[28]</sup> Cilensek et al<sup>[29]</sup> described an anti-apoptotic activity exerted by GAGE for the first time. They cloned GAGE7 from HeLa cells and showed that it renders transfected cells



**Figure 3.** Calibration curves of the clinical prediction model. A: Calibration plot for predicting MPN in the training cohort; B: calibration plot for predicting MPN in the validation cohort. C-index = concordance index, MPN = malignant pulmonary nodule.

resistant to apoptosis induced by interferon-gamma or by the death receptor Fas/CD95/APO-1. In the Fas pathway, the antiapoptotic activity of GAGE-7 maps downstream of caspase-8 activation and upstream of poly (ADP-ribose) polymerase (PARP) cleavage. Furthermore, GAGE7 renders the cells resistant to the therapeutic agents Taxol and gamma-

irradiation. Following the various apoptotic stimuli, the surviving GAGE7 transfectants actively proliferate and exhibit enhanced long term survival in colony formation assays.<sup>[29]</sup> In general, there is a functional link between GAGE7 and 2 aspects of human tumor progression, namely, resistance to Fas induced apoptosis and to chemo- and radio-therapy.

**Table 4**  
**Accuracy of the nomogram in predicting the risk of malignant pulmonary nodules at the optimal threshold value.**

Variable	Value (95%CI)	
	Training cohort	Validation cohort
Sensitivity, %	80.1 (75.9–87.2)	74.0 (67.7–85.2)
Specificity, %	75.8 (70.3–85.8)	73.5 (69.6–84.4)
Positive predictive value, %	77.4 (82.8–91.0)	84.1 (84.2–98.0)
Negative predictive value, %	72.2 (65.6–78.1)	75.5 (64.8–83.7)
Positive likelihood ratio	4.07 (2.80–5.90)	4.60 (3.60–6.90)
Negative likelihood ratio	0.29 (0.20–0.36)	0.32 (0.23–0.40)
Concordance index	0.86 (0.83, 0.91)	0.88 (0.84–0.94)
Predicted probability*	0.54	0.56

CI = confidence interval.

\* Predicted probability refers to the optimal cutoff value for malignant pulmonary nodules prediction based on the maximum Youden index.

The detection of AABs has been officially carried out in clinic rather than being limited to scientific research applications. Their concentrations were quantitated by enzyme linked immunosorbent assay which is simple and fast with good feasibility. The high specificity and stability of AABs indicates that they are suitable to improve the diagnostic performance of mixed panels of biomarkers and complement the results of high sensitivity imaging studies.<sup>[30]</sup> It will provide a reliable basis for clinicians to make appropriate treatment decisions with the gradual popularization of clinical application. We can also explore and design more diverse autoantibody profiles and apply them to more tumor diseases so that AABs could be a valuable tool for early detection not only of lung cancer but also of other malignant tumors in the future.

Imaging examination can directly reflect the morphological parameters of SPN such as size, location, shape, and density. According to the different solid components in the nodules SPNs can be divided into 3 types: pure ground glass nodules, part solid nodules, and solid nodules. The pathological mechanism of malignant pure ground glass nodule is that the cancer cells grow along the wall of bronchus and alveoli, and do not damage the normal lung scaffold structure of alveolar cavity and bronchus. At this stage, the nodules show pure ground glass density. With the growth of cancer cells, normal lung tissues such as alveoli and bronchus were gradually broken, while fibrous scar tissue proliferated, some solid tissue density appeared, and nodules gradually presented as part of solid nodules. As Fleischner society points out in the guidelines for the diagnosis and treatment of SPNs, part solid nodules are more likely to be malignant than pure ground glass nodules. Most solid nodules with smooth margin and diameter less than 6 mm are benign.<sup>[31]</sup> In addition, the malignant probability of SPNs will increase with the enlarge of size. Our research shows a consistent conclusion.

Traditional tumor markers such as CEA, CYFRA21-1 have been widely used in the establishment of prediction models, but univariate analysis showed that there was no significant difference between MPN and BPN in our study. These indicators have predictive value in high-risk groups of lung cancer.<sup>[6]</sup> however, they cannot be used for the screening of early lung cancer in asymptomatic patients.

Undeniably, our study still had some limitations. First, all the data analyzed in this study were obtained from a single institution, and data from other centers are needed to further verify the reliability of the model. Second, although AABs

detection has been carried out in clinic, its application time is still short, the sample size is therefore limited. The remaining 5 AABs did not show prediction efficiency, which may be due to the difference in sensitivity of each indicator or the limitation of sample size. Our research will continue to expand the sample size in the future.

## 5. Conclusion

In summary, we have developed and validated a preoperative prediction model for early lung cancer in patients with SPN. Because of the inclusion of cases that do not meet the surgical indications but put forward the operation requirements, the model has a wider application prospect than those based on surgical indications alone. With the inclusion of 1 demographic (female), 2 laboratory parameters (GAGE7 and GBU4-5), and 3 radiological characteristics (upper lobe of lung, nonsolid nodule, nodule size), our prediction model could effectively differentiate MPN from BPN and provide a reliable basis for clinicians to make appropriate therapeutic decisions.

## Author contributions

**Conceptualization:** Shiyang Pan.

**Data curation:** Zhen Ren, Hongmei Ding, Zhenzhen Cai.

**Formal analysis:** Hongmei Ding, Zhenzhen Cai.

**Funding acquisition:** Shiyang Pan.

**Investigation:** Zhen Ren.

**Methodology:** Zhen Ren, Hongmei Ding, Zhenzhen Cai, Yuan Mu, Shiyang Pan.

**Software:** Zhen Ren, Hongmei Ding, Yuan Mu.

**Writing – original draft:** Zhen Ren, Lin Wang.

**Writing – review & editing:** Shiyang Pan.

## References

- [1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–49.
- [2] Horeweg N, van Rosmalen J, Heuvelmans MA, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *Lancet Oncol* 2014;15:1332–41.
- [3] Ost D, Fein AM, Feinsilver SH. Clinical practice. The solitary pulmonary nodule. *N Engl J Med* 2003;348:2535–42.
- [4] Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA* 2012;307:2418–29.
- [5] Croswell JM, Baker SG, Marcus PM, et al. Cumulative incidence of false-positive test results in lung cancer screening: a randomized trial. *Ann Intern Med* 2010;152:505–12. W176–80.
- [6] Yang D, Zhang X, Powell CA, et al. Probability of cancer in high-risk patients predicted by the protein-based lung cancer biomarker panel in China: LCBP study. *Cancer* 2018;124:262–70.
- [7] He X, Xue N, Liu X, et al. A novel clinical model for predicting malignancy of solitary pulmonary nodules: a multicenter study in Chinese population. *Cancer Cell Int* 2021;21:115.
- [8] Li Y, Wang J. A mathematical model for predicting malignancy of solitary pulmonary nodules. *World J Surg* 2012;36:830–5.
- [9] 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;116:756–62.
- [10] Zimmerman R, Wahren B, Edsmyr F. Assessment of serial CEA determinations in urine of patients with bladder carcinoma. *Cancer* 1980;46:1802–9.
- [11] Chapman CJ, Healey GF, Murray A, et al. EarlyCDT®- lung test: improved clinical utility through additional autoantibody assays. *Tumour Biol* 2012;33:1319–26.

- [12] Healey GF, Lam S, Boyle P, et al. Signal stratification of autoantibody levels in serum samples and its application to the early detection of lung cancer. *J Thorac Dis* 2013;5:618–25.
- [13] Massion PP, Healey GF, Peek LJ, et al. Autoantibody signature enhances the positive predictive power of computed tomography and nodule-based risk models for detection of lung cancer. *J Thorac Oncol* 2017;12:578–84.
- [14] Boyle P, Chapman CJ, Holdenrieder S, et al. Clinical validation of an autoantibody test for lung cancer. *Ann Oncol* 2011;22:383–9.
- [15] Sullivan FM, Farmer E, Mair FS, et al. Detection in blood of autoantibodies to tumour antigens as a case-finding method in lung cancer using the EarlyCDT®-lung test (ECLS): study protocol for a randomized controlled trial. *BMC Cancer* 2017;17:187.
- [16] Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
- [17] Aberle DR, Adams AM, Berg CD, et al. Reduced lung cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
- [18] Mariusz A, Ewa W, Sylwia SS, et al. Risk factors assessment and risk prediction models in lung cancer screening candidates. *Ann Transl Med* 2016;4:151.
- [19] Brawley OW, Flenaugh EL. Low-dose spiral CT screening and evaluation of the solitary pulmonary nodule. *Oncology (Williston Park)* 2014;28:441–6.
- [20] National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Non-small cell lung cancer V.1.2020. Available at: [http://www.nccn.org/professionals/physician\\_gls/PDF/non-small\\_cell\\_lung\\_cancer.pdf](http://www.nccn.org/professionals/physician_gls/PDF/non-small_cell_lung_cancer.pdf). Accessed November 6, 2019.
- [21] Zhou S, Yi T, Zhang B, et al. Mapping the high throughput SEREX technology screening for novel tumor antigens. *Comb Chem High Throughput Screen* 2012;15:202–15.
- [22] Vigneron N, Stroobant V, Van den Eynde BJ, et al. Database of T cell-defined human tumor antigens: the 2013 update. *Cancer Immun* 2013;13:15.
- [23] Tureci O, Usener D, Schneider S, et al. Identification of tumor-associated autoantigens with SEREX. *Methods Mol Med* 2005;109:137–54.
- [24] Ziora D, Kornelia K, Jastrzebski D, et al. High resolution computed tomography in 2-year follow-up of stage I sarcoidosis. *Adv Exp Med Biol* 2013;788:369–74.
- [25] Türeci O, Mack U, Luxemberger U, et al. Humoral responses of lung cancer patients against tumor antigen NY-ESO-1. *Cancer Lett* 2005;236:64–71.
- [26] Xia Q, Kong XT, Zhang GA, et al. Proteomics-based identification of DEAD-box protein 48 as a novel autoantigen, a prospective serum marker for pancreatic cancer. *Biochem Biophys Res Commun* 2005;330:526–32.
- [27] Linder P. Dead-box proteins: a family affair – active and passive players in RNP remodeling. *Nucleic Acids Res* 2006;34:4168–80.
- [28] Kular RK, Yehiely F, Kotlo KU, et al. GAGE, an antiapoptotic protein binds and modulates the expression of nucleophosmin/B23 and interferon regulatory factor 1. *J Interferon Cytokine Res* 2009;29:645–55.
- [29] Cilensek ZM, Yehiely F, Kular RK, et al. A member of the GAGE family of tumor antigens is an anti-apoptotic gene that confers resistance to Fas/CD95/APO-1, interferon-gamma, taxol and gamma-irradiation. *Cancer Biol Ther* 2002;1:380–7.
- [30] Doseeva V, Colpitts T, Gao G, et al. Performance of a multiplexed dual analyte immunoassay for the early detection of non-small cell lung cancer. *J Transl Med* 2015;13:55.
- [31] 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;284:228–43.