

Establishment of a mathematic model for predicting malignancy in solitary pulmonary nodules

Man Zhang^{1,2}, Na Zhuo³, Zhanlin Guo², Xingguang Zhang⁴, Wenhua Liang^{5,6}, Sheng Zhao⁷, Jianxing He^{1,5,6}

¹The First Clinical College, Southern Medical University, Guangzhou 510515, China; ²Department of Thoracic Surgery, ³Department of Dermatology, Affiliated Hospital of Inner Mongolia Medical University, Hohhot 010050, China; ⁴School of Public Health, Inner Mongolia Medical University, Hohhot 010059, China; ⁵Department of Thoracic Surgery, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China; ⁶Guangzhou Institute of Respiratory Disease & China State Key Laboratory of Respiratory Disease & National Clinical Research Center for Respiratory Disease, Guangzhou 510120, China; ⁷Department of Radiology, Affiliated Hospital of Inner Mongolia Medical University, Hohhot 010050, China

Contributions: (I) Conception and design: M Zhang, J He; (II) Administrative support: Z Guo; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: M Zhang, N Zhuo, X Zhang, S Zhao; (V) Data analysis and interpretation: N Zhuo, X Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Jianxing He, MD, PhD, FACS. Department of Thoracic Surgery, The First Affiliated Hospital of Guangzhou Medical University; Guangzhou Institute of Respiratory Disease & China State Key Laboratory of Respiratory Disease, No. 151, Yanjiang Rd, Guangzhou 510120, China. Email: drjianxing.he@gmail.com.

Background: The aim of this study was to establish a model for predicting the probability of malignancy in solitary pulmonary nodules (SPNs) and provide guidance for the diagnosis and follow-up intervention of SPNs.

Methods: We retrospectively analyzed the clinical data and computed tomography (CT) images of 294 patients with a clear pathological diagnosis of SPN. Multivariate logistic regression analysis was used to screen independent predictors of the probability of malignancy in the SPN and to establish a model for predicting malignancy in SPNs. Then, another 120 SPN patients who did not participate in the model establishment were chosen as group B and used to verify the accuracy of the prediction model.

Results: Multivariate logistic regression analysis showed that there were significant differences in age, smoking history, maximum diameter of nodules, spiculation, clear borders, and Cyfra21-1 levels between subgroups with benign and malignant SPNs ($P < 0.05$). These factors were identified as independent predictors of malignancy in SPNs. The area under the curve (AUC) was 0.910 [95% confidence interval (CI), 0.857-0.963] in model with Cyfra21-1 significantly better than 0.812 (95% CI, 0.763-0.861) in model without Cyfra21-1 ($P = 0.008$). The area under receiver operating characteristic (ROC) curve of our model is significantly higher than the Mayo model, VA model and Peking University People's (PKUPH) model. Our model (AUC = 0.910) compared with Brock model (AUC = 0.878, $P = 0.350$), the difference was not statistically significant.

Conclusions: The model added Cyfra21-1 could improve prediction. The prediction model established in this study can be used to assess the probability of malignancy in SPNs, thereby providing help for the diagnosis of SPNs and the selection of follow-up interventions.

Keywords: Solitary pulmonary nodule (SPN); malignant tumor; logistic model; prediction; diagnosis

Submitted Jun 27, 2015. Accepted for publication Oct 09, 2015.

doi: 10.3978/j.issn.2072-1439.2015.10.56

View this article at: <http://dx.doi.org/10.3978/j.issn.2072-1439.2015.10.56>

Introduction

Owing to the extensive use of computed tomography (CT), the detection rate of solitary pulmonary nodules (SPNs) has shown a significantly increasing trend in recent years (1-3). Lung cancer screening trials with large sample sizes indicate that the detection rate of SPN ranges from 8% to 51%, with the vast majority at approximately 20% (4). In SPN cases, malignant nodules account for 5-69%, with an average rate of 40% (5,6). Early diagnosis and treatment of such malignant nodules greatly improves the overall survival rate and prognosis of patients with lung cancer (7,8). Therefore, correctly identifying malignancy in the detected SPN becomes a key point. The ideal goal is to diagnose and treat of malignant nodules early while avoiding unnecessary invasive examinations and surgery for benign nodules. The ultimate goal is to avoid unnecessary cost while allowing SPN patients to obtain the maximum cost benefit.

Differentiation of malignancy or benignancy in SPNs prior to invasive examination or positron emission tomography (PET)/CT scan depends primarily on empirical prediction, which is closely associated with the doctor's theoretical level, practical experience, and diagnostic ability. To reduce human factors and improve diagnostic accuracy, scholars have established models for predicting the probability of malignancy in SPNs based on a combination of clinical and imaging data. Such prediction models can be used to guide doctors in choosing interventions for the next step (9-11). The use of mathematical prediction models is currently recommended by both the American College of Clinical Pharmacy (ACCP) and the Specialty Committee of Lung Cancer, Chinese Anti-cancer Association. It is recommended to first calculate the probability of malignancy in the detected SPN and then perform targeted intervention in accordance with the predicted level of probability of malignancy (4).

Among the diagnostic prediction models for SPN, the Mayo model established the earliest by Swensen *et al.* (9). The Mayo model includes three clinical features (age, smoking history and past history of a malignant tumor) and three imaging features (nodule diameter, presence of spiculation, and location in the lobe). The items included in the Mayo model have an area under the receiver operating characteristic (ROC) curve (AUC) of 0.83. In addition, different diagnostic prediction models for SPNs have been established, such as Mayo model (9), VA model (10), Peking University People's (PKUPH) model (11) and Brock University model (12). According to their respective studies,

most of these models achieve a diagnostic accuracy of more than 80%.

Most of the existing prediction models for SPNs have been established from general clinical data and imaging features of SPN patients, while fewer models have included lung tumor markers. However, the detection of lung tumor markers is an important method in the screening, early diagnosis, and differential diagnosis of lung cancer. Moreover, tumor markers are unaffected by race or the environment. Carcinoembryonic antigen (CEA), cytokeratin-19 fragment (Cyfra21-1), and neuron-specific enolase (NSE) are currently commonly used as lung tumor markers and are available for routine detection in most hospitals. Combined detection of multiple tumor markers has been found to greatly improve the detection rate of lung cancer (13-16). Lung tumor markers are also used in combination with CT images to differentiate malignancy from benignancy in SPNs, which has proven to improve the detection rate of malignant nodules (17,18). However, few prediction models for SPNs have included lung cancer markers to date. Therefore, this study aimed to establish a diagnostic prediction model for SPNs by including lung tumor markers.

Materials and methods

Clinical data

In total, 312 patients with a clear pathological diagnosis of SPN by surgical resection or lung biopsy were reviewed. Of these, 18 were excluded because data were incomplete. A total of 294 patients were collected as group A to create a mathematical model. Patients were collected from The Affiliated Hospital of Inner Mongolia Medical University and The First Affiliated Hospital of Guangzhou Medical University from January 2005 to December 2011. The inclusion criteria were the following: (I) ≤ 3 cm diameter solitary round lesion in the lung, without atelectasis, significant enlargement of hilar and mediastinal lymph nodes, or pleural effusion; (II) clear pathological diagnosis; and (III) complete clinical medical records and CT image data. The patients included 153 men and 141 women, aged 32-80 (55.1 ± 10.7) years. Clinical data were collected from the selected patients, including gender, age, smoking history and quantity, family and past history of malignant tumors, and serum levels of CEA, NSE, and Cyfra21-1.

Another 120 patients with a clear pathological diagnosis of SPN by surgical resection or lung biopsy were collected

from January 2012 to December 2014. These patients served as group B and were used to verify the accuracy of the prediction model.

Imaging data

Plain and/or contrast-enhanced CT data on the patients were collected and independently reviewed by two experienced high-qualification physicians. Detailed records were made for the following CT features of the SPNs: nodule position and size; maximum nodule diameter measured in the lung window; presence or absence of a clear boundary, spiculation and lobulation; cavitation and calcification; vascular convergence; and pleural retraction signs. In cases of discrepancy between the descriptions by the two physicians, re-evaluation was performed by a third physician.

Statistical analysis

Data were analyzed using SPSS 13.0. First, univariate analysis was performed in group A for age, gender, smoking history and quantity, family and past history of malignant tumors, lesion position, maximum nodule diameter, lobulation, spiculation, clear border, cavitation, calcification, vascular convergence sign, pleural retraction sign, and serum levels of CEA, NSE, and Cyfra21-1. Categorical data were analyzed using the χ^2 test, and continuous data were analyzed using the *t*-test. Next, multivariate logistic regression analysis was used to screen independent predictors of the probability of malignancy in SPNs and establish a regression equation for predicting the probability of malignancy. Group B was used to verify the model by the maximum likelihood ratio test, Omnibus test, and Hosmer-Lemeshow test. An appropriate probability cutoff of malignancy or benignancy was chosen, and the sensitivity, specificity, accuracy, positive predictive value, and negative predictive value of the model were calculated. A *P* value of less than 0.05 was considered statistical significance.

Results

Pathological diagnosis

In group A (294 cases), there were 176 cases (59.9%) diagnosed as malignant SPNs, including adenocarcinoma (112 cases, 38.1%), squamous cell carcinoma (45 cases, 15.3%), bronchioloalveolar carcinoma (8 cases, 2.7%), small cell lung cancer (4 cases, 1.3%), large cell carcinoma

(3 cases, 1.0%), carcinoid tumor (2 cases, 0.6%), and adenosquamous carcinoma (2 cases, 0.6%). The other 118 cases (40.1%) were diagnosed as benign SPNs, including tuberculoma (61 cases, 20.7%), inflammatory pseudotumor (23 cases, 7.8%), hamartoma (21 cases, 7.1%), sclerosing hemangioma (5 cases, 1.7%), *Aspergillus* infection (3 cases, 1.0%), local cyst with concomitant infection (3 cases, 1.0%), organizing pneumonia (1 case, 0.3%), and fibrosis (1 case, 0.3%). In group B, there were 72 cases (60.0%) of malignant SPNs and 48 cases (40%) of benign SPNs. No significant differences were found between the two groups in gender, age or nodule diameter.

Univariate and multivariate analyses of independent predictors of SPNs

Univariate analysis showed that there were significant differences between the subgroups of benign and malignant SPNs in terms of age, smoking history, smoking quantity, family history of malignant tumor, tumor diameter, spiculation, lobulation, clear border, calcification, pleural retraction sign, and CEA, NSE, and Cyfra21-1 levels ($P < 0.05$) (Table 1). Furthermore, multivariate logistic regression analysis revealed that age, smoking history, maximum nodule diameter, spiculation, clear border, and Cyfra21-1 level were significantly different between the subgroups of benign and malignant SPNs ($P < 0.05$). These factors were identified as independent predictors for malignancy in SPNs (Table 2).

Establishment of logistic regression equation

Prediction model for the probability of malignancy in SPNs: $P = e^x / (1 + e^x)$, $x = -14.417 + (0.111 \times \text{age}) + (1.009 \times \text{smoking history}) + (2.597 \times \text{nodule diameter}) + (1.056 \times \text{spiculation}) + (-1.258 \times \text{clear border}) + (1.184 \times \text{Cyfra21-1})$, where the following are used: *e* is the natural logarithm; age is recorded by year; history of smoking, either previously or currently, is scored as 1 or otherwise as 0; nodule diameter refers to the maximum nodule diameter measured by chest CT prior to surgery (unit: cm); spiculation and clear border are both derived from the imaging report (1: yes, 0: no); and Cyfra21-1 represents the serum Cyfra21-1 level (units: ng/mL).

Verification of mathematical prediction model and selection of an appropriate probability cut off

Another 120 cases of SPNs (group B) were substituted into

Table 1 Relationship between clinical/CT signs of patients and malignancy/benignancy in SPNs

Characteristics	Benign	Malignant	χ^2	t value	P value
Case number	118	176			
Age (years)	50.11±10.15	61.01±11.36		8.411	0.000
Gender (%)			0.381		0.537
Male	54.2 (64/118)	50.6 (89/176)			
Female	45.8 (54/118)	49.4 (87/176)			
Smoking history (%)					
No	68.6 (81/118)	51.7 (91/176)	8.349		0.004
Smoking quantity, packs/year	162.0±47.8	258.9±71.1		8.853	0.000
Family history of cancer (%)					
No	94.9 (112/118)	85.2 (150/176)	6.835		0.009
Previous cancer history (%)					
No	98.3 (116/118)	98.3 (173/176)	0.000		1.000
Diameter, cm	1.82±0.50	2.32±0.42		9.090	0.000
Clear border (%)					
No	51.7 (61/118)	86.9 (153/176)	44.283		0.000
Spiculation (%)					
No	87.3 (103/118)	64.8 (114/176)	18.525		0.000
Lobulation (%)					
No	87.3 (103/118)	73.3 (129/176)	8.311		0.004
Calcification (%)					
No	87.3 (103/118)	96.0 (169/176)	7.785		0.005
Cavitation (%)					
No	89.8 (106/118)	88.6 (156/176)	0.104		0.747
Pleural retraction sign (%)					
No	87.3 (103/118)	75.0 (132/176)	6.650		0.010
Vascular convergence sign (%)					
No	95.8 (113/118)	92.0 (162/176)	1.615		0.204
Position (%)					
LUL	20.3 (24/118)	28.4 (50/176)	7.690		0.104
LLL	12.7 (15/118)	11.9 (21/176)			
RUL	30.5 (36/118)	29.0 (51/176)			
RML	20.3 (24/118)	10.2 (18/176)			
RLL	16.1 (19/118)	20.5 (36/176)			
CEA (ng/mL)	4.30±3.17	5.30±2.92		2.726	0.003
NSE (ng/mL)	11.53±6.78	14.18±6.64		3.322	0.001
Cyfra21-1 (ng/mL)	2.18±0.69	2.93±0.90		8.000	0.000

CT, computed tomography; SPNs, solitary pulmonary nodules; CEA, carcinoembryonic antigen.

Table 2 Multivariate regression analysis of factors affecting SPNs

Factor	Regression coefficient	Standard error	P	Odds ratio (95% confidence interval)
Age	0.111	0.020	0.000	1.117 (1.075-1.161)
Smoking history	1.009	0.391	0.010	2.744 (1.276-5.901)
Nodule diameter	2.597	0.427	0.000	13.420 (5.807-31.013)
Spiculation	1.056	0.467	0.024	2.874 (1.151-7.175)
Clear border	-1.258	0.408	0.002	0.284 (0.128-0.632)
Cyfra21-1 (ng/mL)	1.184	0.250	0.000	3.267 (2.001-5.333)
Constant	-14.417	1.858	0.000	1.117 (1.075-1.161)

SPNs, solitary pulmonary nodules.

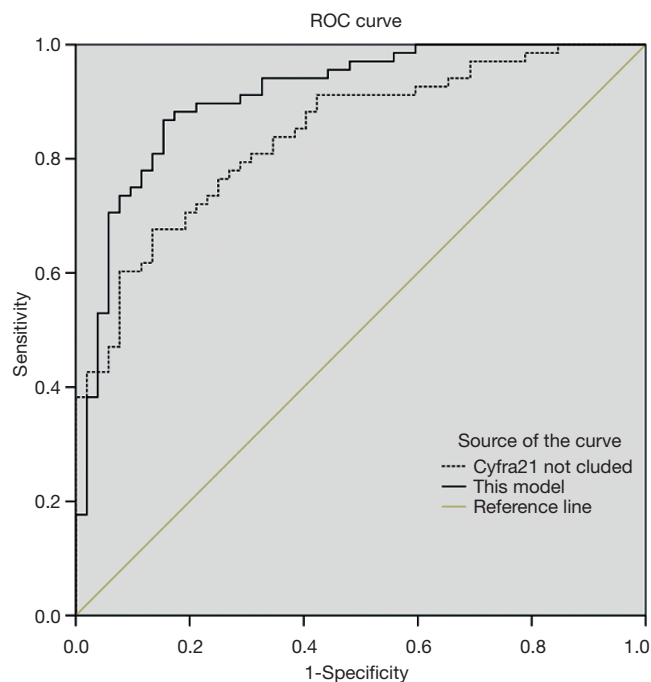


Figure 1 ROC curve generated using our proposed model with and without Cyfra21-1. The AUC was 0.910 (95% CI, 0.857-0.963) in model with Cyfra21-1 as compared with 0.812 (95% CI, 0.763-0.861) in model without Cyfra21-1 ($P=0.008$ for the difference in AUC). ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval.

the regression equation in accordance with the risk factors and assignments to generate the ROC curve (Figure 1). The AUC of group B was 0.910 [95% confidence interval (CI), 0.857-0.963]. An appropriate cut off point was selected at $P=0.5552$, and the model achieved a sensitivity of 86.8%, a specificity of 84.6%, a positive predictive value of 88.1%, and a negative predictive value of 83.0%.

Added value of the Cyfra21-1

The data for group B were substituted into our model with and without Cyfra21-1 to generate the respective ROC curves (Figure 1). The model with Cyfra21-1 was significantly better than the model without Cyfra21-1. The AUC was 0.910 (95% CI, 0.857-0.963) in model with Cyfra21-1 when compared with 0.812 (95% CI, 0.763-0.861) in model without Cyfra21-1 ($P=0.008$ for the difference in AUC), suggesting that adding Cyfra21-1 can improve prediction.

Validation and comparison of different predictive models

The data for group B were substituted into the proposed model, Mayo model, VA model, PKUPH mode and Brock University model to generate the respective ROC curves (Figure 2). The area under the ROC curve of the five models was 0.910, 0.752, 0.730, 0.833 and 0.878 (Table 3).

The area under ROC curve of our model is significantly higher than the Mayo model, VA model and PKUPH model (Table 3). A comparison of our model and Brock model, the AUC in our model was 0.910 (95% CI, 0.857-0.963) higher than the AUC in Brock model of 0.878 (95% CI, 0.837-0.929), the difference was not statistically significant ($P=0.350$).

Discussion

In this study, multivariate logistic regression analysis showed that age, smoking history, maximum nodule diameter, spiculation, and Cyfra21-1 level were identified as independent predictors for estimating malignancy in SPNs, whereas a clear nodule border was found to be a protective factor indicating the possibility of a benign SPN. Based on

the above results, a clinical prediction model for SPNs was established by including two general clinical indices (age and smoking history), three imaging indices (maximum nodule diameter, spiculation, and clear border of nodule), and a laboratory index (Cyfra21-1 level). Various independent risk factors in the model have been reported previously, such as age, smoking history (9,10), maximum nodule diameter (19),

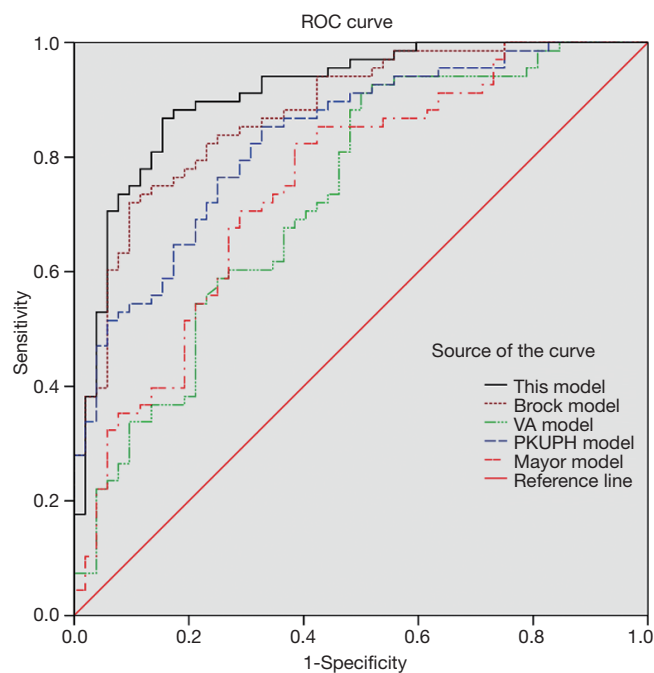


Figure 2 ROC curve of our proposed model and other models. The area under ROC curve of our model is significantly higher than the Mayo model VA model and PKUPH model. Our model (AUC =0.910) compared with Brock model (AUC =0.878, $P=0.350$). ROC, receiver operating characteristic; AUC, area under the curve; PKUPH, Peking University People's.

spiculation, and tumor border (11). One exception is that the Cyfra21-1 level was included in the diagnostic prediction model for SPNs for the first time. Although there is literature proposed that: Adding Cyfra21-1 into the prediction models might improve the accuracy of prediction for SPN, model containing Cyfra21-1 still hasn't been established yet (20).

Research has shown that tumor position and past history of a malignant tumor (9,10) are both independent predictors of malignant SPNs. However, no statistically significant difference was found in tumor position between the subgroups of benign and malignant SPNs in this study. One possible reason is that China has a high incidence of tuberculosis, which accounts for the relatively high proportion of tuberculosis in patients with benign SPNs included in the study. Moreover, tuberculosis occurs preferentially in the upper lobe, similarly to malignant tumors and thus resulting in no significant difference in nodule position between benign and malignant SPNs. In the present study, a past history of cancer had no reference value to distinguish malignancy from benignancy in SPNs, possibly due to the smaller number of SPN patients with a history of malignant tumors and the relatively small total sample size included in the study.

As for the application value of serum tumor markers in pulmonary nodules, some researchers find that: the tumor markers alone or in combination in the diagnosis of pulmonary nodules are limited (14). There was also a study of the combination of tumor markers and imaging in the diagnosis of SPN, which showed that the sensitivity and accuracy of the diagnosis are not improved (21). However, the research has indicated that the tumor markers exhibited higher specificity, which may be a useful supplement to the imaging diagnosis. Recently, some scholars have found

Table 3 Comparison of the area under the ROC curve (AUC) of our model with their models

Model	AUC	Standard deviation	P value	95% confidence interval	
				Lower limit	Upper limit
Proposed model	0.910	0.027	–	0.857	0.963
Our model without Cyfra21-1	0.812	0.025	0.008	0.763	0.861
Brock model	0.878	0.021	0.350	0.837	0.929
PKUPH model	0.833	0.026	0.040	0.782	0.884
Mayo model	0.752	0.045	0.003	0.664	0.841
VA model	0.730	0.047	0.001	0.638	0.822

ROC, receiver operating characteristic; AUC, area under the curve.

that CEA and CYFRA21-1 have higher positive rates in the malignant SPN patients, suggesting a certain value in the early diagnosis of malignant SPN (22). Xiao *et al.* (20) observed serum Cyfra21-1 as a new risk factor adding into the prediction models might improve the accuracy of prediction for SPN. Our results showed that serum Cyfra21-1 level was found to be significantly higher in the malignant SPN subgroup compared with the benign SPN subgroup. Multivariate analysis revealed that the serum Cyfra21-1 level was an independent predictor of the probability of malignancy in SPNs. Therefore, the present study was included the lung tumor marker Cyfra21-1 in a mathematical model for predicting malignancy or benignancy in SPNs. Our results show that the model added CYFRA21-1 increase the area under the ROC curve, that suggesting added Cyfra21-1 could improve prediction.

Serum Cyfra21-1 is the cytokeratin 19 fragment released during tumorigenesis in normal cells. Both adenocarcinoma and squamous cell carcinoma are associated with Cyfra21-1 expression. Thus, Cyfra21-1 is among the serum markers with the greatest diagnostic value in non-small cell lung cancer (23). Meta-analysis has shown that serum Cyfra21-1 is of great value in the diagnosis of non-small cell lung cancer; it is the lung tumor marker with the highest diagnostic efficacy compared with CEA and NSE (24,25). Previous studies have combined the tumor marker CEA with imaging features to establish a prediction model (18). In the present study, despite its significant difference between the two SPN subgroups, the serum CEA level was not an independent predictor of malignancy in SPNs.

Another 120 patients not participating in the model establishment were chosen and substituted into the proposed model, Mayo model, VA model, PKUPH mode and Brock University model. The accuracy of the proposed model was verified, and its diagnostic efficacy was compared with other four models. For the five models tested, the area under ROC curve of our model is significantly higher than the Mayo model, VA model and PKUPH model. Our model Compared with Brock model, the difference was not statistically significant. Brock model is the most accurate prediction tool described to date on the basis of CT and clinical information. It's based on a very large dataset from Canadian CT screening programmes, and now represents the most accurate prediction tool published to date.

The results that the area under the ROC curve of our model is larger and the AUC reaches 0.910, indicate our

model has the same prediction ability as the Brock model. Moreover, our results suggest that by adding Cyfa21-1 to our model, prediction of our model can be improved. Based on this result, we suppose that when building a new large-sample-based model, like Brock model, we may consider collecting the data of adding tumor markers when doing data collection and analysis, Adding tumor markers into the prediction models might improve the accuracy of prediction. The inclusion of tumor markers may further promote the diagnosis of the original model, but it also needs to be verified in future studies.

Based on our calculations, the optimal probability cut off for malignancy was determined to be $P=0.5552$. When the predicted probability of malignancy is greater than or equal to 0.5552, the possibility of a malignant SPN is considered, and further invasive examination or PET/CT is recommended for auxiliary diagnosis. When the predicted probability is less than 0.5552, the possibility of a benign SPN should be considered and follow-up observation recommended. The proposed model achieved a sensitivity of 86.8%, a specificity of 84.6%, a positive predictive value of 88.1%, and a negative predictive value of 83.0%.

Despite its relatively high accuracy, we need to emphasize that this prediction model cannot take the place of a pathological diagnosis. It only serves as a tool for use before targeted intervention following the detection of a SPN. The role of the prediction model is to guide intervention in the next step. Application of the prediction model, on the one hand, can enable timely diagnosis and treatment of the detected malignant SPNs, and on the other hand, it will avoid unnecessary invasive examination and surgery for benign SPNs, ultimately protecting patients from unnecessary medical costs, pain, and risk. This prediction model also has a few defects, such as the relatively small sample size. Moreover, most of the subjects were patients undergoing surgical treatment, indicating a potential selection bias.

In conclusion, a combination of risk factors for malignancy in SPN (age, smoking history, and diameter and shape of nodule) with tumor markers can enable accurate differentiation of malignancy from benignancy in SPNs. Adding Cyfra21-1 into the prediction model can improve the accuracy of prediction for SPN. Such noninvasive SPN evaluation methods, once confirmed to have excellent diagnostic efficacy by multi-center studies in a larger range of patients, will provide meaningful guidance for the diagnosis and treatment of SPNs.

Acknowledgements

Funding: This work was supported by Inner Mongolia Natural Science Foundation (2015MS0899).

Footnote

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

References

1. Khan AN, Al-Jahdali HH, Irion KL, et al. Solitary pulmonary nodule: A diagnostic algorithm in the light of current imaging technique. *Avicenna J Med* 2011;1:39-51.
2. Choromańska A, Macura KJ. Evaluation of solitary pulmonary nodule detected during computed tomography examination. *Pol J Radiol* 2012;77:22-34.
3. 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.
4. Gould MK, Donington J, Lynch WR, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143:e93S-120S.
5. Truong MT, Ko JP, Rossi SE, et al. Update in the evaluation of the solitary pulmonary nodule. *Radiographics* 2014;34:1658-79.
6. Krochmal R, Arias S, Yarmus L, et al. Diagnosis and management of pulmonary nodules. *Expert Rev Respir Med* 2014;8:677-91.
7. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395-409.
8. Aberle DR, DeMello S, Berg CD, et al. Results of the two incidence screenings in the National Lung Screening Trial. *N Engl J Med* 2013;369:920-31.
9. Swensen SJ, Silverstein MD, Ilstrup DM, et al. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med* 1997;157:849-55.
10. Gould MK, Ananth L, Barnett PG, et al. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest* 2007;131:383-8.
11. Li Y, Chen KZ, Wang J. Development and validation of a clinical prediction model to estimate the probability of malignancy in solitary pulmonary nodules in Chinese people. *Clin Lung Cancer* 2011;12:313-9.
12. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013;369:910-9.
13. Bekci TT, Senol T, Maden E. The efficacy of serum carcinoembryonic antigen (CEA), cancer antigen 125 (CA125), carbohydrate antigen 19-9 (CA19-9), carbohydrate antigen 15-3 (CA15-3), alpha-fetoprotein (AFP) and human chorionic gonadotropin (hCG) levels in determining the malignancy of solitary pulmonary nodules. *J Int Med Res* 2009;37:438-45.
14. Seemann MD, Beinert T, Fürst H, et al. An evaluation of the tumour markers, carcinoembryonic antigen (CEA), cytokeratin marker (CYFRA 21-1) and neuron-specific enolase (NSE) in the differentiation of malignant from benign solitary pulmonary lesions. *Lung Cancer* 1999;26:149-55.
15. Chu XY, Hou XB, Song WA, et al. Diagnostic values of SCC, CEA, Cyfra21-1 and NSE for lung cancer in patients with suspicious pulmonary masses: a single center analysis. *Cancer Biol Ther* 2011;11:995-1000.
16. Wang WJ, Tao Z, Gu W, et al. Clinical observations on the association between diagnosis of lung cancer and serum tumor markers in combination. *Asian Pac J Cancer Prev* 2013;14:4369-71.
17. Pecot CV, Li M, Zhang XJ, et al. Added value of a serum proteomic signature in the diagnostic evaluation of lung nodules. *Cancer Epidemiol Biomarkers Prev* 2012;21:786-92.
18. Yonemori K, Tateishi U, Uno H, et al. Development and validation of diagnostic prediction model for solitary pulmonary nodules. *Respirology* 2007;12:856-62.
19. Shi CZ, Zhao Q, Luo LP, et al. Size of solitary pulmonary nodule was the risk factor of malignancy. *J Thorac Dis* 2014;6:668-76.
20. Xiao F, Liu D, Guo Y, et al. Novel and convenient method to evaluate the character of solitary pulmonary nodule-comparison of three mathematical prediction models and further stratification of risk factors. *PLoS One* 2013;8:e78271.
21. Seemann MD, Seemann O, Dienemann H, et al. Diagnostic value of chest radiography, computed tomography and tumour markers in the differentiation of malignant from benign solitary pulmonary lesions. *Eur J*

- Med Res 1999;4:313-27.
22. Ni LF, Liu XM. Diagnostic value of serum tumor markers in differentiating malignant from benign solitary pulmonary nodules. *Beijing Da Xue Xue Bao* 2014;46:707-10.
 23. Kitamura S. CYFRA21-1. *Nihon Rinsho* 2005;63 Suppl 8:654-8.
 24. Cui C, Sun X, Zhang J, et al. The value of serum Cyfra21-1 as a biomarker in the diagnosis of patients with non-small cell lung cancer: A meta-analysis. *J Cancer Res Ther* 2014;10 Suppl:C131-4.
 25. Wang XJ, Liu JL, Xu CA. Diagnostic value of serum CYFRA21-1, CEA and NSE in non-small cell lung cancer: A meta analysis. *Chinese Clinical Oncology* 2011;12:1076-9.

Cite this article as: Zhang M, Zhuo N, Guo Z, Zhang X, Liang W, Zhao S, He J. Establishment of a mathematic model for predicting malignancy in solitary pulmonary nodules. *J Thorac Dis* 2015;7(10):1833-1841. doi: 10.3978/j.issn.2072-1439.2015.10.56