Received: 27 June 2020 | Revised: 21 November 2020 | Accepted: 29 November 2020

DOI: 10.1002/jcla.23682

RESEARCH ARTICLE

WILEY

Development of risk prediction models for lung cancer based on tumor markers and radiological signs

Yuqin Tu[1] | Yan Wu[2] | Yunfeng Lu[3] | Xiaoyun Bi[1] | Te Chen[1]

[1]Department of Medical Laboratory, The First Affiliated Hospital, Chongqing Medical University, Chongqing, China

[2]Department of Blood Transfusion, The First Affiliated Hospital, Chongqing Medical University, Chongqing, China

[3]Department of Radiology, The First Affiliated Hospital, Chongqing Medical University, Chongqing, China

Correspondence
Te Chen, Department of Medical Laboratory, the First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China.
Email: chente@cqmu.edu.cn

Abstract

Background: Accurate prediction of malignancy risk for pulmonary lesions with pleural effusion improves early diagnosis of lung cancer. This study aimed to develop and validate a model to predict lung cancer.

Methods: Clinical data of 536 patients with pulmonary diseases were collected. The risk factors were identified by regression analysis. Three prediction models were developed. The predictive performances of the models were measured by the area under the curves (AUCs) and calibrated with 1000 bootstrap samples to minimize the over-fitting bias. The net benefits of the models were evaluated by decision curve analysis. Finally, a separate cohort of 134 patients was used to validate the models externally.

Results: Seven independent risk factors were identified from 18 clinical variables, which included the pleural fluid carcinoembryonic antigen (CEA), serum cytokeratin-19 fragment (CYFRA 21-1), the ratio of CEA in the pleural fluid to serum, extrathoracic cancer history (>5 years), tumor size, vessel convergence, and lobulation. The AUCs of the three models were 0.976, 0.927, and 0.944 in the training set and 0.930, 0.845, and 0.944 in the external set, respectively. The accuracies of the three models were 89.6%, 81.4%, and 88.8%. Model 1 showed the best iteration fit (R^2 = 0.84, 0.68, and 0.73) and a higher net benefit on decision curve analysis when compared to the other two models.

Conclusion: The advantageous model could assess the risk of lung cancer in patients with pleural effusion and act as a useful tool for early identification of lung cancer.

KEYWORDS

lung cancer, pleural effusion, radiology, risk assessment, tumor markers

1 | INTRODUCTION

Lung cancer is one of the leading causes of cancer-related deaths worldwide, accounting for about 787,000 deaths each year in China.[1,2] Many patients are identified in the advanced stages of lung cancer during initial diagnosis. The age-standardized 5-year relative survival of lung cancer was only 19.7% in China, but if diagnosed at an early stage, then surgical resection offers a favorable prognosis.[3-5]

There are several methods for diagnosing lung cancer. The most commonly recommended method for screening lung cancer is computed tomography (CT), especially the low-dose CT (LDCT). Early screening of lung cancer through LDCT decreases the mortality rate by 20%.[6] Fluorodeoxyglucose positron emission tomography

(FDG-PET) is another screening method recommended. However, the metabolically active infection or inflammatory lesions unfortunately led to false-positive results and controversies.[7] Besides, other approaches, such as needle biopsy, bronchoalveolar lavage, or surgery are also available.[8] These methods are considered expensive, time-consuming, and invasive when compared to serological testing or CT. More importantly, many patients might have no surgical indications for these methods.[9-11]

The diagnostic performance of pleural effusion mostly depends on hydrothorax cytology. Nevertheless, the sensitivity of this method is only around 60%.[12] Several recent studies have reported the use of tumor markers in pleural effusion, achieving satisfactory results in lung cancer diagnosis, especially the carcinoembryonic antigen (CEA) and cytokeratin-19 fragment (CYFRA 21-1).[13-15] Notably, more significant performance improvement can be yielded by combining the biomarkers and radiological signs.[16] Meanwhile, a prediction model is currently regarded as the most useful method.[17] The prediction models for lung cancer diagnosis include the Mayo model, Peking University People's Hospital model (PKUPH), VA model, and Brock University model.[18-21] These models vary in performance, and none of the models involved clinical biomarkers and required further optimization.[17,22,23]

Therefore, our study intended to build new prediction models for lung cancer diagnosis based on real-world clinical data. Also, the external validation and net clinical benefits of these models were evaluated.

2 | MATERIALS AND METHODS

2.1 | Study population

Patients with lung cancer were enrolled from the First Affiliated Hospital of Chongqing Medical University from January 2017 to May 2020. Lung cancer in these patients was confirmed by pathology (surgical resection or biopsy). The staging of non–small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC) was determined according to the criteria of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual, 8th Edition. Only the first admission results were used for analysis in this study. The Clinical Research Ethics Committee of the First Affiliated Hospital of Chongqing Medical University has evaluated and approved the study protocol.

Lung cancer patients were selected based on the inclusion and exclusion criteria. The inclusion criteria were as follows: patients with (I) pleural effusion; (II) complete clinical data including age, gender, smoking status, history of extrathoracic malignant neoplasm (>5 years), and family cancer history; (III) complete imaging data including the maximum diameter (cm) of pulmonary mass, tumor site, speculation, lobulation, vascular convergence, and air bronchogram; and (IV)complete tumor markers results of CEA, CYFRA 21-1, carbohydrate antigen 19-9 (CA 19-9), alpha-fetoprotein (AFP), and ferritin (FER) in pleural effusion (fCEA, fCYFRA 21-1, fCA 19-9, fAFP, fFER), and CEA and CYFRA 21-1 in serum (sCEA and sCYFRA 21-1). The exclusion criteria were as follows: patients (I) without any lumpy shadows on CT; (II) with chronic renal insufficiency, diabetes, liver cirrhosis, and other diseases that can cause false-positive results of tumor markers; (III) with history of malignancy within the past 5 years (persistent or recurrent malignant neoplasm); (IV) who are near to death; (V) who are pregnant or breast-feeding; (VI) who underwent organ transplantation; (VII) with chronic viral infections; and (VIII) taking immunosuppressive medications.

According to these criteria, 402 patients were selected as training set from January 2017 to July 2019 for the prediction models (Figure 1). Another 134 patients were selected from August 2019 to May 2020 as external validation set using the same procedure.

2.2 | Tumor markers assay

The serum and pleural effusion samples were collected and detected on the same day. The tumor markers, including CEA, CYFRA 21-1, CA 19-9, AFP, and FER, were determined by using the electrochemiluminescence method (Cobas e602 immunoassay system; Roche Diagnostics, Switzerland). The quality control measures were strictly controlled when conducting the tests.

2.3 | Radiological imaging acquisition and analysis

Plain or enhanced CT data were collected by scanning the patient's chest using Philips Brilliance CT 64-channel scanner (Philips Healthcare, Netherlands). At least two experienced radiologists interpreted all the CT images together, and radiological signs were recorded by reaching a consensus.

2.4 | Prediction model development

Firstly, the potential risk factors were identified through univariate analysis based on the clinical data, radiological signs, and biomarkers. The variables with $p < 0.10$ in univariate analysis and clinically significant variables were used as candidate parameters for model development. The best-fit model 1 (M1) with chosen variables was then found by multivariate analyses through stepwise regression selection procedure (backward, $p < 0.05$). Considering that some of the data might not be available in the real world, two other models (M2 and M3) have also been developed based on two different datasets. Compared with M1, the two datasets lacked tumor markers and radiological signs, respectively.

2.5 | Statistical analysis

R x64 4.0.1 (The University of Auckland, New Zealand) and SPSS 25.0 (IBM, USA) softwares were used for statistical analysis. Normally distributed continuous variables are presented as means ± standard deviation, non-normally distributed continuous variables are presented as median (interquartile range), and categorical variables are

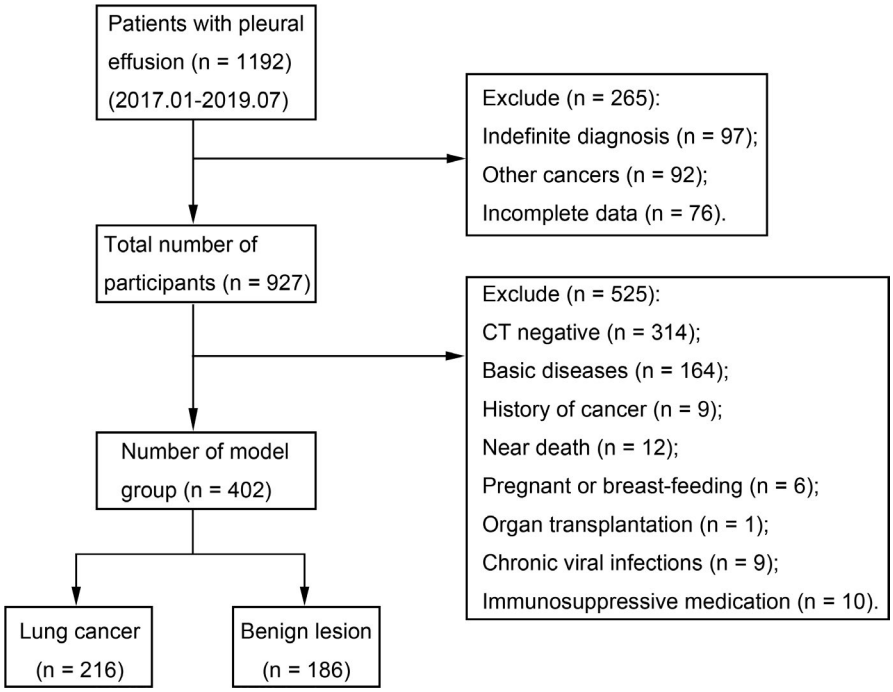

FIGURE 1 Flow diagram. Basic diseases, chronic renal insufficiency, diabetes, liver cirrhosis, and other diseases which can cause false-positive of tumor markers; CT negative, no lump shadow on CT; History of cancer, history of the extrathoracic tumor in the past 5 years; Incomplete data, cases with incomplete clinical data; Other cancers, cancers except for lung cancer

presented as frequencies (percentages). Categorical variables were expressed as frequency and percentage. The levels of tumor markers were $\log_{10}$-transformed. The difference between the groups was evaluated by the Mann-Whitney *U* test or unpaired Student's *t* test, while categorical variables were compared by using the chi-square test. The nomograms were elaborated based on model coefficients by using RMS package, and predictive performances of the models were measured by the AUCs of pROC package and validated using the external sample. The models were calibrated using bootstrap resampling with 1000 samples and confirmed by the Hosmer-Lemeshow goodness-of-fit test. The diagnostic accuracies of the models were compared with external validation data. Finally, decision curve analysis (DCA) was used to assess whether the improved models could predict the net benefit. All tests were two-sided, and $p < 0.05$ was considered to be statistically significant.

3 | RESULTS

3.1 | Patient characteristics

There were 216 lung cancer cases, including adenocarcinoma (159, 73.6%), squamous cell carcinoma (30, 13.9%), and SCLC (27, 12.5%). The remaining 186 cases included were with benign diseases, such as tuberculosis (123, 66.1%), pneumonia (30, 16.1%), and other diseases (33, 17.7%). Table 1 shows that the patients in the cancer group were considerably older, had higher levels of fCEA, fCYFRA 21-1, fCA 19-9, fAFP, sCYFRA 21-1, sCEA, and fCEA/sCEA, and the diameter of the tumor was larger than those with benign diseases ($p < 0.05$). Males were found to be more susceptible to lung diseases when compared to females ($p < 0.05$). Table 2 shows the diagnostic performance of different tumor markers in serum and pleural effusion. Although family tumor history showed no significant difference between the two groups ($p > 0.10$), it was still considered as a candidate factor based on clinical and previous research experience. What's more, the external validation set included 67 lung cancer cases and 67 patients with benign diseases. The clinical characteristics, including the history of malignancy, tumor size, imaging signs, and tumor markers, were comparable to those of the training set (Table 1).

3.2 | Development of the prediction models

Eighteen clinical variables were analyzed for model establishment, and 10 candidate variables, including age, gender, smoking history, family history of cancer, spiculation, air bronchogram, tumor site, fCYFRA 21-1, fAFP, and fCA 19-9 were abandoned in multivariate analysis. Therefore, the parameters of M1 consisted of extrathoracic cancer history (>5 years), sCYFRA 21-1, fCEA/sCEA, diameter, and vessel convergence. M2 contained the history of malignancy (>5 years), diameter, vessel convergence, and lobulation, and M3 included the history of malignancy (>5 years), fCEA, sCEA, and sCYFRA21-1 (Table 3). The diagnostic nomograms show the weights of each parameter in our models (Figure 2).

3.3 | Evaluation and validation of the models

In the training set, compared with M2 (AUC = 0.927, 95% CI: 0.901–0.953) and M3 (AUC = 0.944, 95% CI: 0.922–0.966), M1 (AUC = 0.976, 95% CI: 0.963–0.989) demonstrated the highest predictive performance (Figure 3). In the external set, the

TABLE 1 Clinical characteristics of study subjects

	Cancer	Control	*P*	χ^2/t
Training set				
No. of patients (%)	216 (54)	186 (46)	–	–
Age, years	65.1 ± 10.4	59.4 ± 19.3	0.017	3.742
Male, no. (%)	134 (62.0)	136 (73.1)	0.018	5.565
Smoking, no. (%)	127 (58.8)	93 (50.0)	0.077	3.121
History (>5 years), no. (%)	12 (5.6)	1 (0.54)	0.005	8.042
Family, no. (%)	12 (5.6)	7 (3.8)	0.399	0.713
Diameter, cm	4.3 ± 2.4	1.1 ± 1.1	<0.001	16.563
Lobulation, no. (%)	167 (77.3)	37 (19.9)	<0.001	131.84
Spiculation, no. (%)	88 (40.7)	20 (10.8)	<0.001	45.742
Vessel convergence, no. (%)	21 (9.7)	2 (1.1)	<0.001	13.853
Air bronchogram, no. (%)	60 (27.8)	14 (7.5)	<0.001	27.288
Location-upper lobe, no. (%)	108 (50.0)	112 (60.2)	0.040	4.209
fCYFRA 21-1 (ng/ml)	51.5 (19.6–189.2)	13.7 (5.1–31.5)	<0.001	–
fCEA (ng/ml)	48.7 (4.8–420.7)	1.4 (0.7–1.9)	<0.001	–
fFER (ng/ml)	1404 (737–2972)	1300 (675–2931)	0.656	–
fCA 19-9 (U/ml)	11.9 (2.8–223.7)	4.3 (2.3–7.4)	<0.001	–
fAFP (IU/ml)	1.5 (1.0–2.1)	1.3 (1.0–1.7)	0.001	–
sCYFRA 21-1 (ng/ml)	5.9 (3.2–12.2)	2.1 (1.3–3.3)	0.002	–
sCEA (ng/ml)	6.5 (2.8–30.4)	2.1 (1.3–3.2)	<0.001	–
fCEA/sCEA	3.5 (1.1–13.5)	0.6 (0.5–0.8)	<0.001	–
Validation set				
No. of patients (%)	67 (50)	67 (50)	–	–
Age, years	66.7 ± 12.8	64.6 ± 17.5	0.750	0.807
Male, no. (%)	47 (70.1)	44 (65.7)	0.580	0.308
History (>5 years), no. (%)	9 (13.4)	2 (3.0)	0.028	4.853
Diameter, cm	3.4 ± 2.2	1.5 ± 2.1	<0.001	5.306
Lobulation, no. (%)	57 (85.1)	16 (23.9)	<0.001	50.585
Spiculation, no. (%)	46 (68.7)	4 (6.0)	<0.001	56.280
Vessel convergence, no. (%)	20 (29.9)	2 (3.0)	<0.001	17.620
fCEA (ng/ml)	24.4 (3.5–245.6)	1.2 (0.8–1.7)	<0.001	–
sCYFRA 21-1 (ng/ml)	5.3 (3.1–11.6)	2.3 (1.6–2.7)	<0.001	–
sCEA (ng/ml)	5.6 (2.7–25.5)	2.2 (1.2–3.1)	<0.001	–
fCEA/sCEA	2.7 (0.9–9.6)	0.7 (0.4–0.9)	<0.001	–

Note: Quantitative data are expressed as means ± SD or median (interquartile range).

Abbreviations: "–" not available; fAFP, alpha-fetoprotein in pleural fluid; Family, family cancer history; fCA 19-9, carbohydrate antigen 19-9 in pleural fluid; fCEA, carcinoembryonic antigen in pleural fluid; fCEA/sCEA, the ratio of CEA in the pleural fluid to serum; fCYFRA 21-1, cytokeratin-19 fragment in pleural fluid; fFER, ferritin in pleural fluid; History (>5 years), the history of extrathoracic malignancy (>5 years); sCEA, carcinoembryonic antigen in serum; sCYFRA 21-1, cytokeratin-19 fragment in serum.

AUC of M1 was 0.930 (95% CI: 0.884–0.975), while that of M2 and M3 was 0.845 (95% CI: 0.774–0.916) and 0.944 (95% CI: 0.905–0.983), respectively. M1 showed the best consistency between the predicted probability and the observed frequency (Figure 4), and the *p* values of all the three models in the Hosmer-Lemeshow goodness-of-fit test were >0.05 ($p = 0.727$, 0.230, and 0.750, respectively), indicating the proper fit model. Meanwhile, M1 gave the best results of iteration fitting of the three models ($R^2 = 0.84$, 0.68, and 0.73, respectively). In the external validation set, the accuracy of M1 was 89.6%, and M2 and M3 were 81.4% and 88.8%, respectively (Table 4). Moreover, the three models were superior over the baseline model, and M1 was considered as the optimal model with net benefits as shown via DCA (Figure 5).

TABLE 2 The diagnostic performance of CEA, CYFRA 21-1, FER, CA 19-9, AFP, and fCEA/sCEA in lung cancer

	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Accuracy
fCEA (ng/ml)	78.2 (72.1–83.6)	93.5 (89.0–96.6)	0.90 (0.87–0.93)	0.85
fCYFRA 21-1 (ng/ml)	68.1 (61.4–74.2)	71.5 (64.4–77.9)	0.76 (0.72–0.81)	0.70
fFER (ng/ml)	51.4 (44.5–58.2)	54.8 (47.4–62.1)	0.52 (0.46–0.57)	0.53
fCA 19-9 (U/ml)	48.6 (41.8–55.5)	91.4 (86.4–95.0)	0.69 (0.63–0.74)	0.68
fAFP (IU/ml)	44.2 (37.4–51.1)	73.5 (66.5–79.7)	0.59 (0.63–0.74)	0.57
sCEA (ng/ml)	56.9 (50.1–63.6)	91.9 (87.0–95.4)	0.79 (0.75–0.84)	0.73
sCYFRA 21-1 (ng/ml)	67.6 (60.9–73.8)	84.4 (78.4–89.3)	0.83 (0.79–0.87)	0.75
fCEA/sCEA	75.5 (69.2–81.0)	93.5 (89.0–96.6)	0.87 (0.84–0.91)	0.84

Abbreviations: CI, confidence interval; fAFP, alpha-fetoprotein in pleural fluid; fCA 19-9, carbohydrate antigen 19-9 in pleural fluid; fCEA, carcinoembryonic antigen in pleural fluid; fCEA/sCEA, the ratio of CEA in the pleural fluid to serum; fCYFRA 21-1, cytokeratin-19 fragment in pleural fluid; fFER, ferritin in pleural fluid; sCEA, carcinoembryonic antigen in serum; sCYFRA 21-1, cytokeratin-19 fragment in serum.

TABLE 3 Results of multivariate logistic regression analysis in three models

	Model 1			Model 2			Model 3		
	Coefficient	*p*	OR	Coefficient	*p*	OR	Coefficient	*p*	OR
History (>5 years)	3.7	0.004	42.2	2.8	0.014	16.8	2.7	0.022	15.1
Diameter, cm	1.1	<0.001	3.0	1.0	<0.001	2.7	–	–	–
Vessel convergence	2.1	0.032	8.4	1.8	0.024	6.1	–	–	–
Lobulation sign	–	–	–	1.3	<0.001	3.6	–	–	–
Log fCEA (ng/ml)	–	–	–	–	–	–	3.4	<0.001	28.9
Logs CEA (ng/ml)	–	–	–	–	–	–	−1.5	0.011	0.24
Logs CYFRA 21-1 (ng/ml)	2.5	<0.001	12.5	–	–	–	2.7	<0.001	15.3
Log (fCEA/sCEA)	3.1	<0.001	21.2	–	–	–	–	–	–
Constant	−4.5	<0.001	0.01	−3.0	<0.001	0.05	−2.7	<0.001	0.07

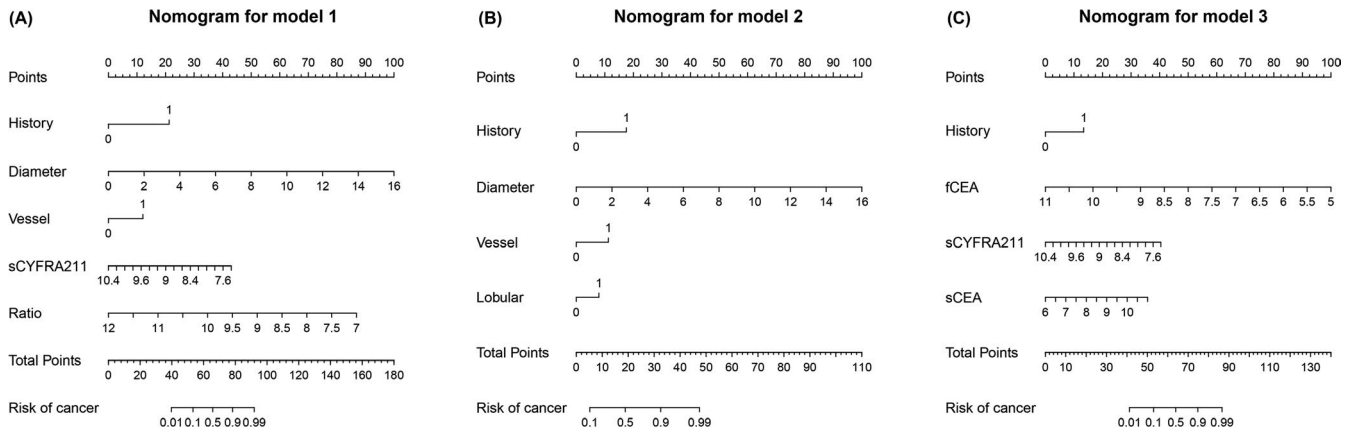

FIGURE 2 Nomograms of the three models. fCEA, the CEA in pleural fluid; History, the history of extrathoracic cancer (>5 years); Lobular, lobulation sign; Ratio, fCEA/sCEA, the ratio of CEA in the pleural fluid to serum; sCYFRA 21-1, the CYFRA 21-1 in serum; sCEA, the CEA in serum; Vessel, vessel convergence

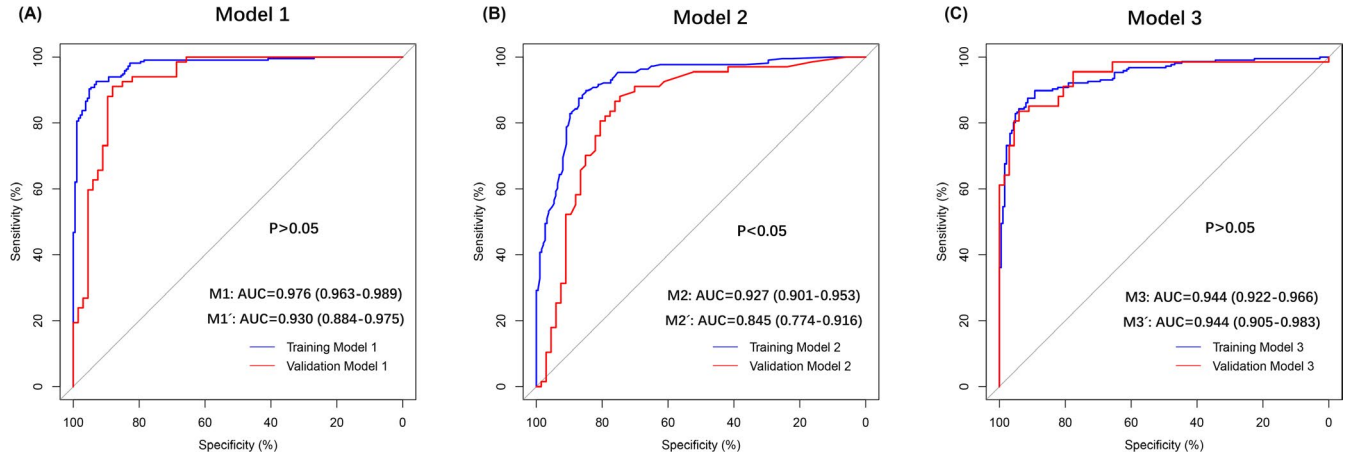

FIGURE 3 ROC curves of the three models in the training set and validation set. The M1′, M2′, and M3′ are the ROCs of the three models in the validation set

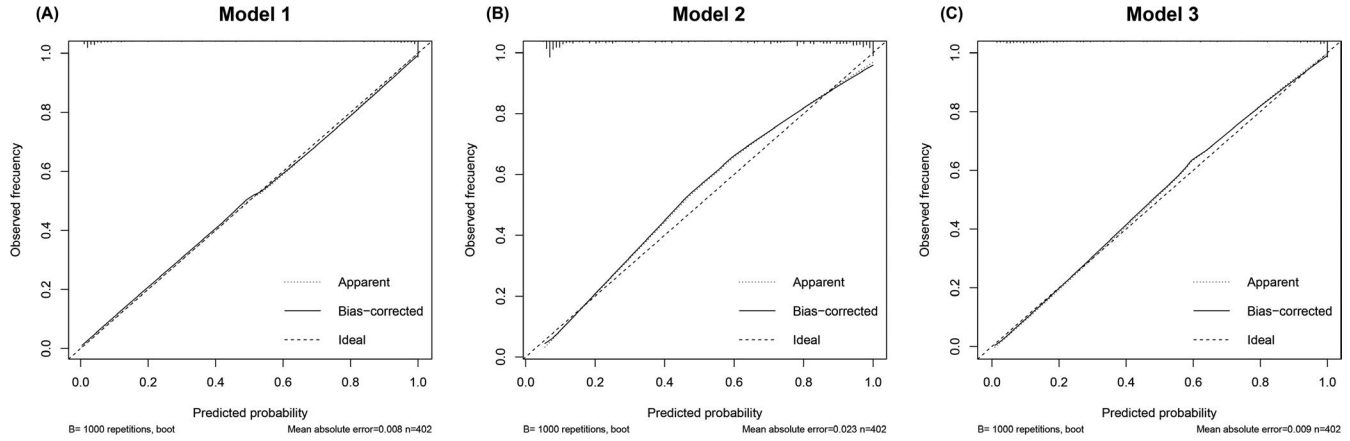

FIGURE 4 Calibration curves of the models

TABLE 4 The diagnostic classification table for external validation set

Model	Observed cancer	Predicted cancer		Percentage correct (%)
		No	Yes	
1	No	59	8	88.1
	Yes	6	61	91.0
	Overall (%)		89.6	
2	No	50	17	74.6
	Yes	8	59	88.1
	Overall (%)		81.4	
3	No	63	4	94.0
	Yes	11	56	83.6
	Overall (%)		88.8	

4 | DISCUSSION

Although there are varied test results with regard to diagnosis currently, the doctors hardly make full use of these results to estimate whether patients are at risk of cancer or not, especially in the early stage of lung cancer.[4] The National Comprehensive Cancer Network (NCCN) guidelines on lung cancer screening have endorsed the utilization of risk prediction model for the identification of high-risk individuals to supplement the National Lung Screening Trial's (NLST) eligibility criteria. The United States Preventive Services Task Force (USPSTF) is currently used for risk-based screening as it revises its recommendations on lung cancer screening.[24,25] Therefore, in the last few years, numerous risk prediction models have been published for lung cancer screening. However, more research is needed to optimize the risk-based lung cancer screening.[24] In this study, three prediction models were developed to predict the probability of malignancy of pulmonary lesions and pleural effusion. Internal and external validations for each model were also conducted, and M1 was found to be the best-fit model of the three prediction models. More importantly, all parameters in our models were objective, readily available, and required no additional detection.

During the data collection phase, 80 patients were not included due to incomplete tumor markers or imaging data. Therefore, the other two models (M2 and M3) with specific types of data were

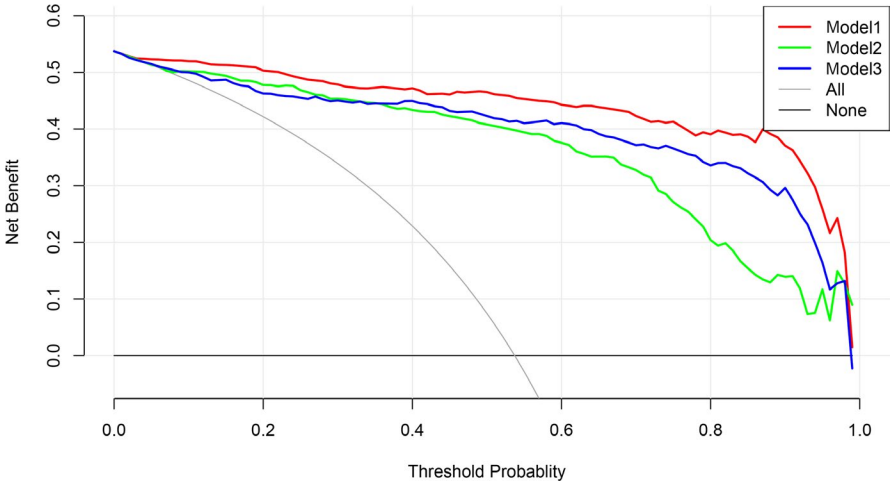

FIGURE 5 Decision curve analysis of the risk of malignancy in lung lesions

built by masking to meet this question, in which M2 was based on radiological signs, and M3 was based on tumor markers. These two models can predict lung cancer and can be used by clinicians according to the condition of the patients. Interestingly, the performance comparisons for M1 and M3 were slightly different between internal and external validations. The main reason for this could be due to the diversity of data in both training and validation sets. However, our validation results, except M2, showed that the differences in AUCs between the validation and training sets of the other two models remained small. Therefore, this does not affect the diagnostic ability of M1 in a single patient. It is worth noting that different lung cancer subtypes for analyses and the three models demonstrated high diagnostic accuracy for different types of cancers when grouped. In the validation set, the diagnostic accuracies of adenocarcinoma, squamous cell carcinoma, and SCLC were 93.0%, 87.4%, and 90.6% for M1, and 94.3%, 90.2%, and 87.4% for M2, and 92.6%, 88.0%, and 90.9% for M3. Nonetheless, we also believed that our model probably requires validation in more extensive data.

The parameters of the best-fit model generally have a closer and safer relationship with lung cancer when compared to other clinical results. For example, the extrathoracic cancer history (>5 years) as a parameter could assist in stably predicting the weight of our nomograms, which is similar to other reports, including the Mayo model and other models.[19,26] Another example, the fCEA/sCEA is shown to have a more reliable performance when compared with fCEA and sCEA individually, and this finding is consistent with that of the previous research studies.[23,27-29] Additionally, radiological signs were another reason for the excellent performance of our models. Many studies have reported that the tumor diameter size acts as an independent factor in judging benign and malignant pulmonary lesions.[16] Yang found that lobulation showed an association with high risk of malignant tumors.[26] Other studies have shown that lumps are more likely to be destructive if vessel convergence signs are present.[30] Furthermore, our models differed from other models in many ways, such as in target population and parameters.[18-26,31,32] On one hand, our models focused on patients with lung diseases and pleural effusion, in which 73.6% of cases were adenocarcinomatous in the lung cancer group, and 66.1% of subjects had tuberculosis in the benign group. This composition ratio is similar to that of epidemiological distribution of lung diseases with pleural effusion.[5,33] On the other hand, all parameters in our models were easily obtained and have reliable repeatability, eliminating most of the errors caused by subjective factors.

However, the development of a prediction model based on potential variables is challenging. Some biomarkers, such as CA125, CA153, HE4, and CA242, metalloproteinase-9 (MMP-9) have been used in other studies.[34-36] But these factors have low specificities for diagnosing lung cancer and were usually used to screen female malignancies. Besides, with the development of genomics, proteomics, radiomics, and liquid biopsy techniques, increasing indicators help in predicting the risk of lung cancer in a better way.[37] Unfortunately, most of these markers still require more clinical confirmations and cannot be widely used as a routine test in the hospitals currently. More importantly, the indicators obtained from new methods encourage us to develop better models rather than replacing them. Therefore, more characteristics with better specificity and sensitivity should be incorporated for predicting the model.

However, this study has several limitations. Firstly, M1 does not apply to patients without pleural effusion or with surgical contraindications, and so, M2 should be chosen as an alternative model. Secondly, the diagnostic performance of the same indicator from serum is lower than that from pleural effusion. For example, in our work, the AUC of the CEA in serum was only 0.79, but the AUC of the CEA in pleural effusion was 0.90. Therefore, indicators in pleural effusion might have potential advantages in lung cancer risk prediction. Thirdly, the baseline ranges of nomograms were limited, especially the diameter, and might have low diagnostic performance in some patients.

In conclusion, our prediction models were developed and confirmed by combining serum and pleural effusion tumor markers with radiological signs. Our models can reasonably assess lung cancer risk in patients with pulmonary diseases and provide a new valuable tool for early diagnosis of lung cancer.

ACKNOWLEDGMENTS

The authors would like to thank the clinical laboratory technician, pulmonologist, and radiologist of the First Affiliated Hospital of Chongqing Medical University, Chongqing, China, who have participated in our study.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

Te Chen contributed to study concept, design, and article writing. Yuqin Tu contributed to data collection, analysis, interpretation, and article writing. Yan Wu and Yunfeng Lu contributed to provision of study materials or patients.. Xiaoyun Bi provided administrative support:. All authors read and approved the final article.

DATA AVAILABILITY STATEMENT

The data used to support the findings of this study are restricted by the Clinical Research Ethics Committee of the First Affiliated Hospital of Chongqing Medical University to protect patient privacy. The data may be available upon request but not for all due to the data protection laws.

ORCID

Te Chen https://orcid.org/0000-0001-7153-4016

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.
2. Zheng R, Sun K, Zhang S, et al. Report of cancer epidemiology in China, 2015. *Chinese Journal of Oncology*. 2019;41(1):19-28.
3. Zeng H, Chen W, Zheng R, et al. Changing cancer survival in China during 2003–15: a pooled analysis of 17 population-based cancer registries. *Lancet Glob Health*. 2018;6(5):e555-e567.
4. Blandin Knight S, Crosbie PA, Balata H, Chudziak J, Hussell T, Dive C. Progress and prospects of early detection in lung cancer. *Open Biol*. 2017;7(9):1–12.
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7-30.
6. Mazzone PJ, Silvestri GA, Patel S, et al. Screening for lung cancer: chest guideline and expert panel report. *Chest*. 2018;153(4):954-985.
7. Porcel JM, Hernandez P, Martinez-Alonso M, Bielsa S, Salud A. Accuracy of fluorodeoxyglucose-PET imaging for differentiating benign from malignant pleural effusions: a meta-analysis. *Chest*. 2015;147(2):502-512.
8. Desai NR, Lee HJ. Diagnosis and management of malignant pleural effusions: state of the art in 2017. *J Thorac Dis*. 2017;9(Suppl 10):S1111-S1122.
9. Veltri A, Bargellini I, Giorgi L, Almeida P, Akhan O. CIRSE Guidelines on Percutaneous Needle Biopsy (PNB). *Cardiovasc Intervent Radiol*. 2017;40(10):1501-1513.
10. Guo Z, Shi H, Li W, et al. Chinese multidisciplinary expert consensus: Guidelines on percutaneous transthoracic needle biopsy. *Thorac Cancer*. 2018;9(11):1530-1543.
11. Miller RJ, Casal RF, Lazarus DR, Ost DE, Eapen GA. Flexible Bronchoscopy. *Clin Chest Med*. 2018;39(1):1-16.
12. Porcel JM, Esquerda A, Vives M, Bielsa S. Etiology of pleural effusions: analysis of more than 3,000 consecutive thoracenteses. *Arch Bronconeumol*. 2014;50(5):161-165.
13. Volaric D, Flego V, Zauhar G, Bulat-Kardum L. Diagnostic value of tumour markers in pleural effusions. *Biochem Med (Zagreb)*. 2018;28(1):10706.
14. Enz N, Fragoso F, Gamrekeli A, Lippek F, Jungraithmayr W. Carcinoembryonic antigen-positive pleural effusion in early stage non-small cell lung cancer without pleural infiltration. *J Thorac Dis*. 2018;10(5):E340-E343.
15. Feng M, Zhu J, Liang L, et al. Diagnostic value of tumor markers for lung adenocarcinoma-associated malignant pleural effusion: a validation study and meta-analysis. *Int J Clin Oncol*. 2017;22(2):283-290.
16. Du Q, Yan C, Wu S-G, et al. Development and validation of a novel diagnostic nomogram model based on tumor markers for assessing cancer risk of pulmonary lesions: a multicenter study in Chinese population. *Cancer Lett*. 2018;420:236-241.
17. Zhang M, Zhuo N, Guo Z, et al. Establishment of a mathematic model for predicting malignancy in solitary pulmonary nodules. *J Thorac Dis*. 2015;7(10):1833-1841.
18. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *New Engl J Med*. 2013;369(10):910-919.
19. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med*. 1997;157(8):849-855.
20. 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(5):313-319.
21. Gould MK, Ananth L, Barnett PG, Veterans Affairs SCSG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest*. 2007;131(2):383-388.
22. Ferreiro L, Gude F, Toubes ME, et al. Predictive models of malignant transudative pleural effusions. *J Thorac Dis*. 2017;9(1):106-116.
23. Pan Y, Bai W, Chen J, et al. Diagnosing malignant pleural effusion using clinical and analytical parameters. *J Clin Lab Anal*. 2019;33(2):e22689.
24. Toumazis I, Bastani M, Han SS, Plevritis SK. Risk-Based lung cancer screening: a systematic review. *Lung Cancer*. 2020;147:154-186.
25. Tang W, Peng Q, Lyu Y, et al. Risk prediction models for lung cancer: perspectives and dissemination. *Chin J Cancer Res*. 2019;31(2):316-328.
26. Yang L, Zhang Q, Bai L, et al. Assessment of the cancer risk factors of solitary pulmonary nodules. *Oncotarget*. 2017;8(17):29318-29327.
27. Trape J, Sant F, Franquesa J, et al. Evaluation of two strategies for the interpretation of tumour markers in pleural effusions. *Respir Res*. 2017;18(1):103.
28. Zhai K, Wang W, Wang Y, Liu JY, Zhou Q, Shi HZ. Diagnostic accuracy of tumor markers for malignant pleural effusion: a derivation and validation study. *J Thorac Dis*. 2017;9(12):5220-5229.
29. Bai Y, Shen W, Zhu M, et al. Combined detection of estrogen and tumor markers is an important reference factor in the diagnosis and prognosis of lung cancer. *J Cell Biochem*. 2019;120(1):105-114.
30. Li X, Zhang Q, Jin X, Cao L. Combining serum miRNAs, CEA, and CYFRA21-1 with imaging and clinical features to distinguish benign and malignant pulmonary nodules: a pilot study: Xianfeng Li et al.: Combining biomarker, imaging, and clinical features to distinguish pulmonary nodules. *World J Surg Oncol*. 2017;15(1):107.
31. Tammemagi MC, Ten Haaf K, Toumazis I, et al. Development and validation of a multivariable lung cancer risk prediction model that includes low-dose computed tomography screening results: a secondary analysis of data from the national lung screening trial. *JAMA Netw Open*. 2019;2(3):e190204.

32. Gray EP, Teare MD, Stevens J, Archer R. Risk prediction models for lung cancer: a systematic review. *Clin Lung Cancer*. 2016;17(2):95-106.
33. Bai C, Choi CM, Chu CM, et al. Evaluation of pulmonary nodules: clinical practice consensus guidelines for Asia. *Chest*. 2016;150(4):877-893.
34. Chien LH, Chen CH, Chen TY, et al. Predicting lung cancer occurrence in never-smoking females in Asia: TNSF-SQ, a prediction model. *Cancer Epidemiol Biomarkers Prev*. 2020;29(2):452-459.
35. Integrative Analysis of Lung Cancer Etiology and Risk Consortium for Early Detection of Lung Cancer, Guida F, Nan S, et al. Assessment of lung cancer risk on the basis of a biomarker panel of circulating proteins. *JAMA Oncol*. 2018;4(10):e182078.
36. Porcel JM, Esquerda A, Martinez-Alonso M, Bielsa S, Salud A. Identifying thoracic malignancies through pleural fluid biomarkers: a predictive multivariate model. *Medicine (Baltimore)*. 2016;95(10):e3044.
37. Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. *Ann Glob Health*. 2019;85(1):8.

How to cite this article: Tu Y, Wu Y, Lu Y, Bi X, Chen T. Development of risk prediction models for lung cancer based on tumor markers and radiological signs. *J Clin Lab Anal*. 2021;35:e23682. https://doi.org/10.1002/jcla.23682