ORIGINAL RESEARCH

Development and validation of a novel diagnostic model for assessing lung cancer metastasis in a Chinese population based on multicenter real-world data

Yiyong Yao, [ID][1],* Cunling Yan,[2],*
Wei Zhang,[3],* San-Gang Wu,[4],*
Jie Guan,[2],* Gang Zeng,[1],* Qiang Du,[1]
Chun Huang,[1] Hui Zhang,[5]
Huiling Wang,[6] Yanfeng Hou,[2]
Zhiyan Li,[2] Lixin Wang, [ID][7]
Yijie Zheng,[8] Xun Li[9]

[1]Department of Respiratory Medicine, Suzhou Municipal Hospital, Nanjing Medical University, Suzhou, People's Republic of China; [2]Department of Clinical Laboratory, Peking University First Hospital, Beijing, People's Republic of China; [3]Department of Biostatistics, School of Public Health, Fudan University, Shanghai, People's Republic of China; [4]Department of Radiation Oncology, Xiamen Cancer Hospital, The First Affiliated Hospital of Xiamen University, Xiamen, People's Republic of China; [5]Department of Laboratory, Suzhou Municipal Hospital, Nanjing Medical University, Suzhou, People's Republic of China; [6]Department of Respiratory Medicine, The Second Affiliated Hospital, Dalian Medical University, Dalian, People's Republic of China; [7]Department of TCM and Western Medicine, Shanghai Pulmonary Hospital Affiliated to Tongji University, Shanghai, People's Republic of China; [8]Medical Scientific Affairs, Abbott Diagnostics Division, Abbott Laboratories, Asian Pacific Group, Shanghai, People's Republic of China; [9]Department of Laboratory Medicine, The First Affiliated Hospital, School of Medicine, Xiamen University, Xiamen, People's Republic of China

*These authors contributed equally to this work

Correspondence: Yijie Zheng
Medical Scientific Affairs, Abbott Diagnostics Division, Abbott Laboratories, 388 Nanjing Xi Rd, Shanghai 200000, People's Republic of China
Tel +86 212 315 4961
Fax +86 216 334 6331
Email yijiezheng2015@163.com

Xun Li
Department of Laboratory Medicine, The First Affiliated Hospital, School of Medicine, Xiamen University, 55 Zhenhai Rd, Xiamen, Fujian 361003, People's Republic of China
Tel + 865922139507
Email xli2001@xmu.edu.cn

Background: Accurate disease staging plays an important role in lung cancer's clinical management. However, due to the limitation of the CT scan, it is still an unmet medical need in practice. In the present study, we attempted to develop diagnostic models based on biomarkers and clinical parameters for assessing lung cancer metastasis.

Methods: This study consisted of 799 patients with pulmonary lesions from three regional centers in China. It included 274 benign lesions patients, 326 primary lung cancer patients without metastasis, and 199 advanced lung cancer patients with lymph node or organ metastasis. The patients were divided into nodules group and masses group according to tumor size.

Results: Four nomogram models based on patient characteristics and tumor biomarkers were developed and evaluated for patients with nodules and masses, respectively. In patients with pulmonary nodules, the AUC to identify metastatic lung cancer from unidentified nodules (including benign nodules and lung cancer, model 1) reached 0.859 (0.827–0.887, 95% CI). Model 2 was used to predict metastasis in patients with lung cancer with AUC of 0.838 (0.795–0.876, 95% CI). In patients with pulmonary masses, the AUC to identify metastatic lung cancer from unidentified masses (model 3) reached 0.773 (0.717–0.823, 95% CI). Model 4 was used to predict metastasis in patients with lung cancer and AUC reached 0.731 (0.771–0.793, 95% CI). Decision curve analysis corroborated good clinical applicability of the nomograms in predicting metastasis.

Conclusion: All new models demonstrated promising discrimination, allowing for estimating the risk of lymph node or organ metastasis of lung cancer. Such integration of blood biomarker testing with CT imaging results will be an efficient and effective approach to benefit the accurate staging and treatment of lung cancer.

Keywords: CT imaging pulmonary lesions, biomarker, nomogram models, lung cancer metastasis, multicenter real-world

Introduction

Accurate disease staging plays an important role in lung cancer management by informing treatment choices and prognosis.[1] Scagliotti[2] reported that neoadjuvant chemotherapy could benefit patients with clinical stage IIB/IIIA non-small cell lung cancer (NSCLC) (including $T_{1-2}N_{1-2}M_0$, $T_{3-4}N_{0-1}M_0$), instead of patients with IB/IIA (including $T_{2a-2b}N_0M_0$) NSCLC. As the stage of lung cancer advances, the

prognosis becomes significantly worse, as evidenced by the 5-year survival rate of 60–92% for localized cancer, but only 13–53%% and 0–10% for advanced and metastatic disease, respectively.[3,4] Obviously, accurate diagnosis of lymph node and organ metastasis is very important in the diagnosis and treatment of lung cancer.

CT scan is the most commonly used noninvasive method for lung cancer evaluation. However, its ability is limited in evaluating the status of lymph nodes due to the fact that the size of mediastinal lymph nodes (LDs) is hardly correlated with tumor involvement.[5] In several reports, the sensitivity and specificity of chest CT scan for mediastinal lymph nodes' staging were only 52–57% and 81–85%, respectively (1,6). Positron-emission tomography (PET)/CT is another noninvasive technique with improved performance in assessing the status of lymph nodes, with a pooled sensitivity of 74–84% and specificity of 85–89% for assessing lymph node metastasis (1,6). PET/CT is better than traditional CT scan because it can detect the biological activity of cells. However, granulomatous lesions, infections, and other inflammatory diseases may result in false positive results on PET/CT. Furthermore, some well-differentiated low-grade malignancies, particularly bronchioloalveolar cell carcinoma and typical carcinoid tumors, may produce false negative results on PET/CT.[6–9] Additionally, the size of lesion must be adequate for detection by PET/CT because the lower limit of spatial resolution of current generation of PET scanners is approximately 7–10 mm.[6] The accessibility of this technology may also be limited due to economic cost and radiation exposure. Therefore, a simple, convenient, reproducible, noninvasive method is desirable to facilitate the assessment of lymph node or organ metastasis.

Blood-based biomarkers have been widely used in clinical practice, such as in testicular, prostate, colorectal, breast, and ovarian cancers. The clinical quality requirements and clinical recommendations are published by the National Academy of Clinical Biochemistry (NACB).[10–12] Human epididymis protein 4 (HE4) has been shown to be a new ideal biomarker of lung cancer.[13–15] Numerous studies have confirmed that tumor markers increase significantly in advanced lung cancer, compared with early lung cancer.[16–18] Blood levels of biomarkers are indicative of tumor burden, which is correlated with tumor stage. The prevalence of tumor markers is higher in advanced stage.[19,20] Given the importance of metastasis detection, we investigated the combination of biomarkers and clinical parameters in assessing the state of LDs or organ metastasis on CT. Four nomogram-based models were built to assist assessment of the lesions detected by CT scan. To the best of our knowledge, this study was the first attempt to develop a diagnostic model based on combination of biomarkers and clinical parameters for assessing lung cancer metastasis.

Materials and methods
Overall study design
The current study was designed to evaluate a combination of 5 biomarkers, including CEA, SCC, CYRFRA21-1, ProGRP and HE4, and clinical information to predict the occurrence of LDs or organ metastasis in patients with unidentified
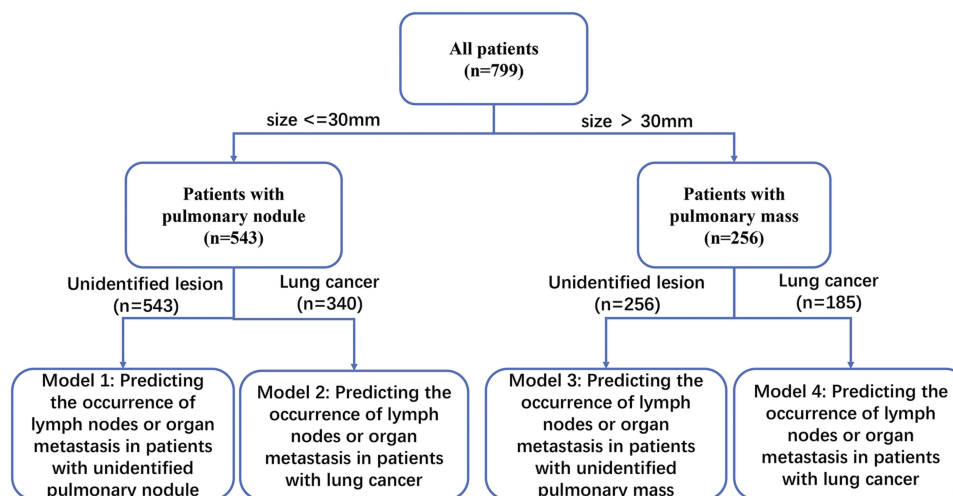

Figure 1 Overall study design. Firstly, all patients were divided into nodule group and mass group according to the size of the lesion; secondly, patients were divided into unidentified lesions group (before pathological confirmation) and lung cancer group (after pathological confirmation).

lesions (before pathological confirmation) and patients with lung cancer (after pathological confirmation) (Figure 1). To this end, all patients were divided into nodule group and masses group according to the size of the lesion, and four predictive models were developed: two for the pulmonary nodule and the other two for pulmonary masses.

Patients

A total of 799 participants with pulmonary lesions were enrolled from three centers (Suzhou Hospital Affiliated to Nanjing Medical University, the First Affiliated Hospital of Xiamen University, and Peking University First Hospital) in China from October 2015 to August 2018. The diagnosis of lung cancer was confirmed for all patients by pathology. Lymph node or organ metastasis was confirmed by pathology (surgical resection and/or biopsy). Other metastases were diagnosed by imaging, such as ultrasonic or CT. The specific stages were based on the pathological evaluation and imaging. The staging of NSCLC and small cell lung cancer (SCLC) was determined according to the criteria of AJCC Cancer Staging Manual, 8th Edition and the Veterans Administration Lung Cancer Study Group.[3,30] All patients were naive to antineoplastic therapy, radiotherapy or chemotherapy before surgery or cancer diagnosis. The study protocol was approved by the ethics committees of the hospitals. The institutional review board numbers are K2017002 for Suzhou Hospital Affiliated to Nanjing Medical University, 2017[1294] for Peking University First Hospital and KYX-2015-020 for the First Affiliated Hospital of Xiamen University respectively. All patients signed written informed consents and this study was conducted in accordance with the Declaration of Helsinki. The patients with benign lesion included hamartoma, tuberculoma, intrapulmonary lymph nodes, and fungus ball.

Analysis of tumor markers and image acquisition

A commercial chemiluminescent microparticle immunoassay (CMIA) method was used to test blood samples for CEA, SCC, CYFRA21-1, HE4, and proGRP on ARCHITECT i2000SR (Abbott Laboratories, Chicago, IL). All samples in the training set were run in duplicate. All patients were examined by regular multi-detector CT (MDCT) scan. The maximum dimension on axial CT images was measured and recorded by two senior consultant radiologists.

Statistical analysis

Continuous variables are described as mean (SD) or median (p25, p75) and count variables are described as frequency and percentage. All analyses were stratified by tumor size and tumor lesions. An independent-sample t-test and paired sample t-test or Mann-Whitney U test was used to compare continuous variables, while count variables were compared by χ2 test or Fisher exact test. Cut-off for each marker was determined by minimum p-value method[31] with 1000-time bootstrap replications on purpose of overcoming the variability of a single sample. The cut-off of a single bootstrap sample was estimated by minimum p-value approach and the mean of 1000 individual cut-off values as the final cut-off.

Variables including age, sex, CEA, SCC, CYFRA21-1, HE4, and proGRP were involved in building models. Predicting factors with $p<0.1$ in univariate analysis were used as candidate risk factors for stepwise multivariate logistic regression model (LRM) with a backward selection. A nomogram was elaborated upon LRM coefficients by using RMS package of R. The requirement of sample size was 150 for LRM modeling with 10 variables. A larger sample size will allow a more stable statistical model. The predictive performance of the nomogram was measured by concordance index (C index) and calibrated with 1000 bootstrap samples. Finally, through decision curve analyses (DCAs),[32,33] we evaluated whether the model improved the predictive net benefit. All tests were both sided and 0.05 was set as the P-value for significance. Statistical analyses were performed using R programming language v.3.5.2.

Results
Patient characteristics

A total of 799 patients with pulmonary lesions were enrolled, including 274 patients with benign lesion, 326 patients with primary lung cancer without metastasis, and 199 patients with advanced lung cancer with lymph node or organ metastasis. The patients included 543 patients with pulmonary nodule and 256 patients with pulmonary masses. Lymph node or organ metastasis was confirmed by pathology. The characteristics of the subjects were presented in Table 1.

Blood level of tumor markers

Compared with lung cancer without metastasis (including lung cancer and benign lesion) in pulmonary nodules, CEA, CYFRA21-1, Pro-GRP, and HE4 showed higher levels in lung cancer with metastasis. Compared with others in

Table 1 Patient demographic and clinical data

Variable	Patients with pulmonary nodules (lesion size≤30mm)			Patients with pulmonary mass (lesion size>30mm)		
	Lung cancer with metastasis (n=58)	Lung cancer without metastasis (n=282)	Benign lesion (n=203)	Lung cancer with metastasis (n=141)	Lung cancer without metastasis (n=44)	Benign lesion (n=71)
Sex						
Male (%)	37 (63.8)	114 (40.4)	119 (58.6)	112 (79.4)	34 (77.3)	52 (73.2)
Female (%)	21 (36.2)	168 (59.6)	84 (41.4)	29 (20.6)	10 (22.7)	19 (26.8)
Age (years)						
Median (Q1, Q3)	63 (53,70)	61 (50,68)	51 (34,62)	64 (57,70)	68 (60,74)	50 (40,59)
Subtype						
Adenocarcinoma (%)	43 (74.0)	264 (93.6)	NA	63 (44.7)	24 (54.5)	NA
SqCa (%)	9 (15.5)	13 (4.6)	NA	46 (32.6)	17 (38.6)	NA
Adenosquamous (%)	0	1 (0.4)	NA	0	1 (2.3)	NA
SCLC (%)	2 (3.5)	2 (0.7)	NA	23 (16.3)	0	NA
Others [a] (%)	4 (7.0)	2 (0.7)	NA	9 (6.4)	2 (4.6)	NA
Size (mm)						
Median (Q1, Q3)	23.0 (20.0,28.0)	15.0 (10.0,20.0)	14.0 (8.0,20.0)	48.0 (40.0,60.0)	48.0 (36.0,55.0)	37.0 (34.0,50.0)
T Stage						
x (%)	2 (3.5)	0	NA	0	0	NA
Tis (%)	0	17 (6.0)	NA	0	0	NA
1 (%)	31 (53.4)	233 (82.6)	NA	0	1 (2.3)	NA
2 (%)	14 (24.1)	26 (9.3)	NA	49 (34.8)	23 (52.3)	NA
3 (%)	2 (3.5)	4 (1.4)	NA	48 (34.0)	16 (36.4)	NA
4 (%)	9 (15.5)	2 (0.7)	NA	44 (31.2)	4 (9.1)	NA
N Stage						
0 (%)	5 (8.6)	282 (100)	NA	1 (0.7)	44 (100)	NA
1 (%)	16 (27.6)	0	NA	38 (26.9)	0	NA
2 (%)	23 (39.7)	0	NA	65 (46.1)	0	NA
3 (%)	14 (24.1)	0	NA	37 (26.2)	0	NA
M Stage						
0 (%)	43 (74.1)	282 (100)	NA	94 (66.7)	44 (100)	NA
1 (%)	13 (22.4)	0	NA	42 (29.8)	0	NA
x (%)	2 (3.5)	0	NA	5 (3.5)	0	NA

Note: [a]Others, including neuroendocrine carcinoma and large cell lung cancer.
Abbreviations: NA, not applicable; SqCa, squamous cell carcinoma; SCLC, small cell lung cancer.

pulmonary masses, only CEA, CYFRA21-1, and HE4 showed higher levels in lung cancer with metastasis (Table 2). When it was compared to patients with lesion without metastasis in pulmonary nodules, CEA, CYFRA21-1, and HE4 showed higher levels in lung cancer with metastasis. While other comparisons in pulmonary masses, CEA, SCC, and CYFRA21-1 exhibited higher levels in lung cancer with metastasis (Table 2).

Cut-off and the distribution of tumor markers

The present study divided patients into four cohorts according to the size of lesion and pathology (Figure 1). Cut-off values of markers for each model were determined by minimum p-value method with 1000-time bootstrap replications to overcome the variability of a single sample. The details of cut-off and distribution of tumor markers based on different

Table 2 Serum level of tumor markers in different lesions

Biomarker	Lung cancer with metastasis	Lung cancer without metastasis	Benign lesion	P_1	P_2
	Median (Q1, Q3)	Median (Q1, Q3)	Median (Q1, Q3)		
Lesion size≤30mm					
CEA (ng/mL)	5.43 (2.67, 9.40)	2.02 (1.31, 3.32)	2.03 (1.25, 2.84)	<0.001	<0.001
SCC (ng/mL)	0.90 (0.50, 1.10)	0.70 (0.50, 1.10)	0.70 (0.50, 1.00)	0.194	0.274
CYFRA21-1 (ng/mL)	2.19 (1.47, 3.92)	1.57 (1.03, 2.19)	1.41 (0.99, 2.02)	<0.001	<0.001
Pro-GRP (pg/mL)	27.94 (13.39, 46.81)	21.56 (12.64, 34.55)	19.62 (12.15, 34.65)	0.049	0.073
HE4 (pmol/L)	73.40 (54.00, 106.80)	50.75 (39.30, 68.70)	48.70 (33.70, 70.80)	<0.001	<0.001
Lesion size>30mm					
CEA (ng/mL)	4.73 (2.65, 10.98)	3.47 (2.20, 5.78)	2.29 (1.40, 3.11)	<0.001	0.048
SCC (ng/mL)	0.80 (0.50, 1.80)	1.20 (0.70, 2.25)	0.70 (0.50, 1.10)	0.716	0.048
CYFRA21-1 (ng/mL)	3.74 (2.19, 6.26)	2.66 (1.66, 4.72)	1.24 (0.90, 1.65)	<0.001	0.046
Pro-GRP (pg/mL)	26.72 (12.11, 44.14)	34.58 (22.34, 43.91)	19.00 (12.31, 32.39)	0.709	0.063
HE4 (pmol/L)	93.30 (65.50, 131.70)	81.15 (60.80, 112.25)	61.00 (40.90, 91.80)	<0.001	0.173

Notes: P1, lung cancer with metastasis vs lesion without metastasis (including lung cancer and benign lesion); P2, lung cancer with metastasis vs lung cancer without metastasis.

Table 3 The distribution of tumor markers based on the cut-off values

	CEA		SCC		CYFRA21-1		pro-GRP		HE4	
Lesion size≤30mm/lung cancer with metastasis vs lesion without metastasis (including lung cancer and benign lesion)										
Cut-off	≤5.58	>5.58	≤0.78	>0.78	≤2.55	>2.55	≤39.04	>39.04	≤78.22	>78.22
Group 1, n (%)	29(50.0)	29(50.0)	23(39.7)	35(60.3)	37(63.8)	21(36.2)	38 (65.5)	20 (34.5)	32 (55.2)	26 (44.8)
Group 2, n (%)	457(94.2)	28(5.8)	254(52.4)	231(47.6)	413(85.2)	72(14.8)	391(80.6)	94(19.4)	393(81.0)	92(19.0)
P	<0.001		0.067		<0.001		0.008		<0.001	
Lesion size≤30mm/lung cancer with metastasis vs lung cancer without metastasis										
Cut-off	≤6.12	>6.12	≤0.81	>0.81	≤2.89	>2.89	≤40.66	>40.66	≤79.81	>79.81
Group 1, n (%)	31 (53.5)	27 (46.5)	28 (48.3)	30 (51.7)	40 (69.0)	18 (31.0)	38(65.5)	20(34.5)	32(55.2)	26(44.8)
Group 3, n (%)	261 (92.5)	21 (7.5)	173 (61.4)	109 (38.6)	248 (87.9)	34 (12.1)	235 (83.3)	47 (16.7)	235 (83.3)	47 (16.7)
P	<0.001		0.065		<0.001		0.002		<0.001	
Lesion size>30mm/lung cancer with metastasis vs lesion without metastasis (including lung cancer and benign lesion)										
Cut-off	≤3.78	>3.78	≤1.85	>1.85	≤1.95	>1.95	≤43.94	>43.94	≤78.86	>78.86
Group 4, n (%)	54 (38.3)	87 (61.7)	106 (75.2)	35 (24.8)	27 (19.2)	114 (80.8)	104 (73.8)	37 (26.2)	51 (36.2)	90 (63.8)
Group 5, n (%)	86 (74.8)	29 (25.2)	100 (87.0)	15 (13.0)	75 (65.2)	40 (34.8)	95 (82.6)	20 (17.4)	72 (62.6)	43 (37.4)
P	<0.001		0.018		<0.001		0.090		<0.001	
Lesion size>30mm/lung cancer with metastasis vs lung cancer without metastasis										
Cut-off	≤3.44	>3.44	≤1.06	>1.06	≤2.91	>2.91	≤22.12	>22.12	≤96.35	>96.35
Group 4, n (%)	48 (34.0)	93 (66.0)	88 (62.4)	53 (37.6)	54 (38.3)	87 (61.7)	62 (44.0)	79 (56.0)	75 (53.2)	66 (46.8)
Group 6, n (%)	22 (50.0)	22 (50.0)	21 (47.7)	23 (52.3)	24 (54.5)	20 (45.5)	11 (25.0)	33 (75.0)	30 (68.2)	14 (31.8)
P	0.057		0.084		0.057		0.025		0.080	

Notes: Group 1: lesion size≤30mm and lung cancer with metastasis; Group 2: lesion size≤30mm and lesion without metastasis (including lung cancer and benign lesion); Group 3: lesion size≤30mm and lung cancer without metastasis. Group 4: lesion size>30mm and lung cancer with metastasis; Group 5: lesion size>30mm and lesion without metastasis (including lung cancer and benign lesion); Group 6: lesion size>30mm and lung cancer without metastasis.

cut-off values were shown in Table 3. Univariate analysis was used to screen risk factors, and candidate risk factors were analyzed by stepwise multivariate logistic regression model (LRM) with a backward selection.

Model 1: prediction of lymph node or organ metastasis in patients with unidentified pulmonary nodule (size≤30 mm)

A total of 543 patients with pulmonary lesions were analyzed, including 58 lung cancer patients with metastasis, 282 lung cancer patients without metastasis, and 203 patients with benign nodule. Variables, such as sex, age, size of lesion, CYFRA21-1, HE4, SCC, ProGRP, and CEA were examined in stepwise multivariate analysis. Only size of lesion and CEA were significant independent risk factors for occurrence of lymph node or organ metastasis in patients with unidentified lesion ($p<0.05$). The result of logistic regression was presented in Table 4. The combination of all independent risk factors can be expressed as:

$$\text{Logit}(p) = 0.12 \times \text{Size} + 2.51 \times \text{CEA} - 4.94$$

CEA=1, if the level of CEA is over 5.58, otherwise =0. The AUC for model 1 in identifying metastatic lung cancer from unidentified nodule (including benign disease and lung cancer) reached 0.859 (95% CI, 0.827 to 0.887), (Table 4). The prognostic nomogram that integrated all significant independent factors for metastasis was shown in Figure 2A. Calibration curve indicated good agreement between prediction and observation in the probability of metastasis (Figure 2B). DCA was applied to render clinical validity to the nomograms (Figure 2C). It corroborated good clinical applicability of the nomograms in predicting metastasis because the ranges of threshold probabilities were wide and practical.

Model 2: prediction of lymph node or organ metastasis in patients with lung cancer (size≤30mm)

Different from model 1, patients with benign nodule were excluded. This model was used to predict the occurrence of metastasis in lung cancer patients (size≤30mm). The study population consisted of 340 patients, including 58 lung cancer patients with metastasis and 282 lung cancer patients without metastasis. The result of multivariate analysis was similar to model 1. The size of lesion and CEA were independent risk factors for occurrence of lymph node or organ

metastasis in patients with lung cancer ($p< 0.05$). The result of logistic regression was presented in Table 4. The combination of all independent risk factors can be expressed as:

$$\text{Logit}(p) = 0.11 \times \text{Size} + 2.08 \times \text{CEA} - 4.15$$

CEA=1, if the level of CEA is over 6.12, otherwise = 0. The AUC for model 2 in identifying metastatic lung cancer from lung cancer (size≤30mm) reached 0.838 (95% CI, 0.795 to 0.876), (Table 4). The prognostic nomogram that integrated all significant independent factors for metastasis was shown in Figure 2D. Calibration curve showed good agreement between prediction and observation in the probability of metastasis (Figure 2E). DCA was applied to render clinical validity to the nomograms (Figure 2F). It corroborated good clinical applicability of the nomograms in predicting metastasis because the ranges of threshold probabilities were wide and practical.

Model 3: prediction of lymph node or organ metastasis in patients with unidentified pulmonary masses (size >30 mm)

In this model, we analyzed 256 patients with pulmonary masses, including 141 lung cancer patients with metastasis, 44 lung cancer patients without metastasis, and 71 patients with benign masses. Sex, age, size of lesion, CYFRA21-1, HE4, SCC, ProGRP, and CEA were the variables included in stepwise multivariate analysis. Only CEA and CYFRA21-1 were independent risk factors for occurrence of lymph node or organ metastasis in patients with unidentified masses ($p<0.05$). The result of logistic regression was presented in Table 4. The combination of all independent risk factors can be expressed as:

$$\text{Logit}(p) = 1.01 \times \text{CEA} + 1.73 \times \text{CYFRA21} - 1 - 1.25$$

CEA=1, if the level of CEA is over 3.38, otherwise = 0; CYFRA21-1=1, if the level of CYFRA21-1 is over 1.95, otherwise = 0. The AUC for model 3 in identifying metastatic lung cancer from unidentified masses (including benign disease and lung cancer) reached 0.773 (95% CI, 0.717 to 0.823), (Table 4). The prognostic nomogram that integrated all significant independent factors for metastasis was shown in Figure 3A. Calibration curve showed good agreement between prediction and observation in the probability of metastasis (Figure 3B). DCA was applied to render clinical validity to the nomograms (Figure 3C). It

Table 4 The results of logistic regression based on different lesions

Parameter	β	OR (95% CI)	P-value	AUC (95% CI)
Model 1				0.859 (0.827, 0.887)
Intercept	−4.94	NA	<0.001	
Lesion size (mm)	0.12	1.13 (1.08, 1.18)	<0.001	
CEA (>5.58 vs ≤5.58)	2.51	12.29 (6.24, 24.20)	<0.001	
Model 2				0.838 (0.795, 0.876)
Intercept	−4.15	NA	<0.001	
Lesion size (mm)	0.11	1.12 (1.07, 1.17)	<0.001	
CEA (>6.12 vs ≤6.12)	2.08	8.00 (3.88 16.46)	<0.001	
Model 3				0.773 (0.717, 0.823)
Intercept	−1.25	NA	<0.001	
CEA (>3.78 vs ≤3.78)	1.01	2.73 (1.50, 4.98)	0.003	
CYFRA21-1 (>1.95 vs ≤1.95)	1.73	5.63 (3.09, 10.27)	<0.001	
Model 4				0.731 (0.661, 0.793)
Intercept	4.46	NA	0.002	
Age (years)	−0.05	0.95 (0.91, 0.99)	0.025	
SCC (>1.06 vs ≤1.06)	−0.75	0.47 (0.22, 0.99)	0.048	
CYFRA21-1 (>2.91 vs ≤2.91)	0.97	2.65 (1.23, 5.72)	0.013	
pro-GRP (>22.12 vs ≤22.12)	−1.12	0.33 (0.14, 0.75)	0.009	
HE4 (>96.35 vs ≤96.35)	0.94	2.56 (1.16, 5.66)	0.020	

Notes: Model 1: lung cancer with metastasis vs lesion without metastasis (including lung cancer and benign lesion) when lesion size ≤30mm; Model 2: lung cancer with metastasis vs lung cancer without metastasis when lesion size ≤30mm; Model 3: lung cancer with metastasis vs lesion without metastasis (including lung cancer and benign lesion) when lesion size > 30mm; Model 4: lung cancer with metastasis vs lung cancer without metastasis when lesion size > 30mm.
Abbreviation: NA, not applicable.

corroborated good clinical applicability of the nomograms in predicting metastasis because the ranges of threshold probabilities were wide and practical.

Model 4: prediction of lymph node or organ metastasis in patients with lung cancer (size >30mm)

This model was used to predict the occurrence of metastasis in lung cancer patients (size>30mm), which included 141 lung cancer patients with metastasis and 44 lung cancer patients without metastasis. Patients with benign nodule were excluded. Age, SCC, CYFRA21-1, pro-GRP, and HE4 were independent risk factors for occurrence of lymph node or organ metastasis in patients with lung cancer (size>30mm) (p< 0.05). The result of logistic regression is presented in Table 4. The combination of all independent risk factors can be expressed as:

$$\text{Logit(p)} = -0.05 \times \text{age} - 0.75 \times \text{SCC} + 0.97 \times \text{CYFRA21} - 1 - 1.12 \times \text{pro} - \text{GRP} + 0.94 \times \text{HE4} + 4.46$$

SCC=1, if the level of SCC is over 1.06, otherwise = 0; CYFRA21-1=1, if the level of CYFRA21-1 is over 2.91, otherwise = 0; pro-GRP =1, if the level of pro-GRP is over

22.12, otherwise = 0; HE4=1, if the level of HE4 is over 96.35, otherwise = 0. The AUC for model 4 in identifying metastatic lung cancer from lung cancer (size>30mm) reached 0.731 (95% CI, 0.661 to 0.793), (Table 4). The prognostic nomogram that integrated all significant independent factors for metastasis was shown in Figure 3D. Calibration curve showed good agreement between prediction and observation in the probability of metastasis (Figure 3E). DCA was applied to render clinical validity to the nomograms (Figure 3F). It corroborated good clinical applicability of the nomograms in predicting metastasis because the ranges of threshold probabilities were wide and practical.

Discussion

Outcome of lung cancer patients are subjected to clinical and pretreatment evaluation, especially for lymph nodes. Such evaluations are important for decisions regarding invasive procedures or treatments, such as neoadjuvant chemotherapy. From 2007 to 2013, among all the reported lung cancer cases, only 16% were diagnosed at early stage when the cancer was limited to primary site. Around 22% were diagnosed after the cancer had spread to regional lymph nodes or directly beyond the primary site.
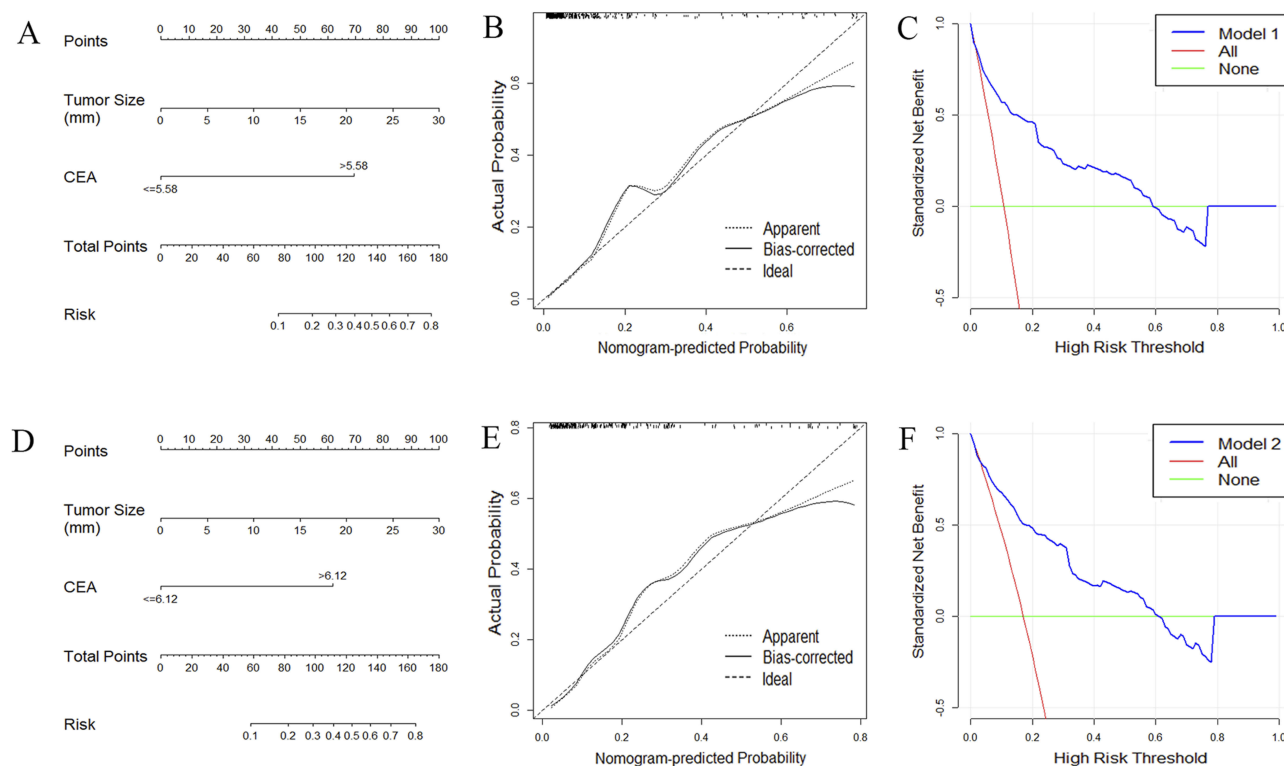
Figure 2 Development of nomogram to predict the occurrence of lymph node or organ metastasis in patients with nodules. (**A**) Model 1: for predicting the occurrence of lymph node or organ metastasis in patients with unidentified pulmonary nodule (size≤30 mm). To use the nomogram, an individual patient's value is located on each variable axis, and a line is drawn upward to determine the number of points received for each variable value. The sum of these numbers is located on the Total Points axis, and a line is drawn downward to Risk axes to determine the risk of metastasis in patients with unidentified lesions. (**B**) The calibration curve for predicting the occurrence of lymph node or organ metastasis in patients with unidentified pulmonary nodules. (**C**) Decision curve analysis of nomograms based on model 1. (**D**) Model 2: for predicting the occurrence of lymph node or organ metastasis in patients with lung cancer (size≤30mm). (**E**) The calibration curve for predicting the occurrence of lymph node or organ metastasis in patients with lung cancer (size≤30mm). (**F**) Decision curve analysis of nomograms based on model 2.

Unfortunately, at least half of cases (57%) were diagnosed at late stage after the cancer had already metastasized.[2,21] In order to improve the clinical efficiency, we developed a diagnosis model to differentiate cancer and benign nodules.[22] However, this is far from fulfilling the current clinical need. Accurate clinical evaluation of lymph node or organ metastasis is still a challenge but important for newly diagnosed lung cancers. In the present study, we built a novel diagnostic model to assess lung cancer metastasis based on combination of biomarkers, lesion size, and characteristics of patients, which will provide new decision support in clinical practice. In general, four different models were built in our article. When we found lesions in CT screening test, the nature of the lesion was unclear and Models 1 and 3 can assist physician to assess risk of malignancy and metastasis of the lesions. When the nature of the lesion was identified by biopsy, bronchoscope or other methods, Models 2 and 4 can be used. The current golden standard for metastasis is still pathological

examination. PET/CT also provides valuable information. However, the cost is too high. Thus, one important clinical application is to assist the metastasis risk assessment before further clinical decision.

Due to its simple, convenient, inexpensive, and reproducible application, tumor markers have been used in lung cancer assessment for decades. However, most of the previous studies have focused on their diagnostic and prognostic utilities. The implication of these tumor markers in the evaluation of lymph node or organ metastasis has not been systematically evaluated. Thus, this large-scale study further explored the potential clinical application of tumor biomarkers in metastasis assessment of lung cancer. The performances of the five common biomarkers (CEA, CYFRA21-1, Pro-GRP, SCC, and HE4) exhibited different levels among subgroup analysis.

Data from the Dutch Belgian Randomized Lung Cancer Screening (NELSON) trial of CT scan screening showed that malignancy was associated with larger nodule size for solid
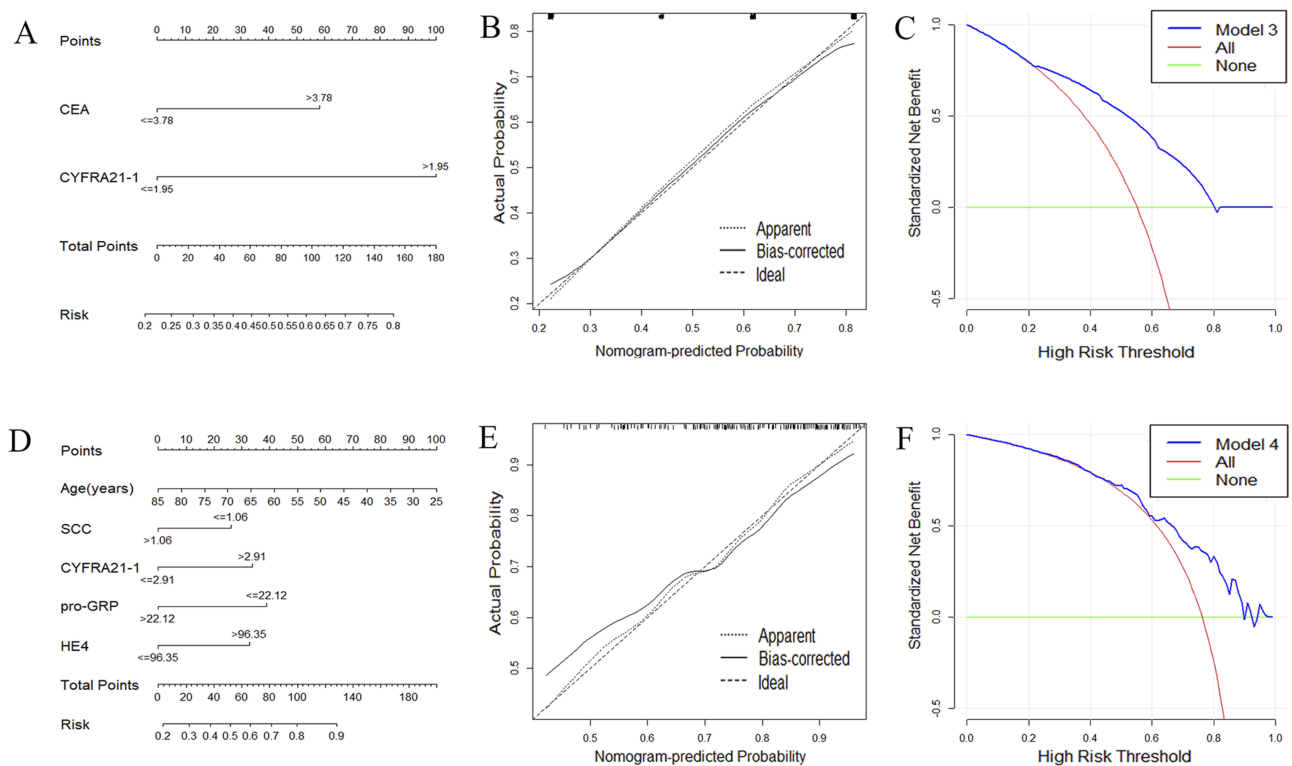
Figure 3 Development of nomogram to predict the occurrence of lymph node or organ metastasis in patients with mass. (**A**) Model 3: for predicting the occurrence of lymph node or organ metastasis in patients with unidentified pulmonary mass (size >30 mm). To use the nomogram, an individual patient's value is located on each variable axis, and a line is drawn upward to determine the number of points received for each variable value. The sum of these numbers is located on the Total Points axis, and a line is drawn downward to Risk axes to determine the risk of metastasis in patients with unidentified lesions. (**B**) The calibration curve for predicting the occurrence of lymph node or organ metastasis in patients with unidentified pulmonary mass. (**C**) Decision curve analysis of nomograms based on model 3. (**D**) Model 4: for predicting the occurrence of lymph node or organ metastasis in patients with lung cancer (size >30mm). (**E**) The calibration curve for predicting the occurrence of lymph node or organ metastasis in patients with lung cancer (size >30 mm). (**F**) Decision curve analysis of nomograms based on model 4.

nodules.[23] Besides, 30mm is the boundary between pulmonary nodules and pulmonary masses.[6] So we chose 30mm as cutoff. We classified pulmonary lesions into nodules and masses according to the size of lesions. According to different clinical applications, four models to assess lung cancer metastasis have been developed. In pulmonary nodules group and pulmonary masses group, one model was designed to assess the risk of lymph node or organ metastasis in case of unidentified lesions, and the other one was developed to evaluate the risk of lymph node or organ metastasis in lung cancer only. For patients with pulmonary nodules, independent risk factors for metastasis are size and CEA. This may be related to the high percentage of adenocarcinoma patients in the patient pools because this population is known to have high CEA levels. Additionally, the AUC for diagnosing lymph node or organ metastasis is better when the diameter of lesion is smaller than 30 mm. This may be due to the accumulation of tumor biomarkers in late stage tumors regardless of metastasis or not, as the level of tumor marker is correlated to lymph node metastasis and lesion size.[20] Although the level of tumor

markers is less efficient in predicting lymph node metastasis at later stage, fortunately, metastasis of larger lesions (≥30 mm) is much easier to detect by CT/MRI. Thus, the proposed models in this study can be used as a tool and decision support supplementary to CT/MRI in assessing lung cancer metastasis, especially in smaller lesions (< 30 mm).

For patients with pulmonary masses, independent risk factors for metastasis include CEA, CYFRA21-1, and HE4. Results show that SCC and pro-GRP are in negative correlation to the risk of metastasis. Nevertheless, previous research[24] reports that SCC has nothing to do with metastasis. The results may be attributed to selection bias. Besides, pro-GRP is mainly used in the diagnosis and evaluation of neuroendocrine tumors. The number of neuroendocrine tumors included in our trial is small.

Modeling and statistical construction is a common practice to assist diagnosis or estimate prognosis. A growing body of evidence suggests that the combination of biomarkers and tumor morphology, such as size, can improve the accuracy of the model. For example, in the field of lung cancer

diagnosis, Yang et al built a diagnostic model based on bio-markers and the most predictive imaging features of lung cancer.[25] The model demonstrates better accuracy on risk prediction than the ACCP lung nodule model. In addition, the information provided by the biomarker is no longer limited to its normal range. More useful information can be generated when biomarkers are combined with clinical data. Wang et al developed a prognostic nomogram for resectable intrahepatic cholangiocarcinoma based on tumor markers and clinico-pathologic data.[26] The model can provide more accurate pre-diction of patient survival compared with the staging systems currently available.[26] The risk prediction models proposed in this study were based on nomogram, which can provide indi-vidualized, evidence-based, and highly accurate estimation of cancer metastasis risk by combining multiple independent variables and assigning an appropriate weight to each variable based on its prognostic value.[27] Additionally, nomograms are easy to use and can facilitate management-related decision making.[28,29] To the best of our knowledge, this was the first attempt to develop a validated nomogram for assessing the risk of lymph node or organ metastasis in lung cancer based on tumor markers and the size of lung lesions.

There are some limitations to our study. Firstly, the number of patients included in our study was still lim-ited. A larger sample size will provide a more reliable model. Secondly, internal validation was used in the study. The validity of these models needs to be verified further with larger external data. Thirdly, the perfor-mance of CT scan alone in diagnosing mediastinal lymph node metastasis was not evaluated in our analy-sis. We will carry out further randomized prospective study to investigate whether the diagnosis of mediastinal lymph node metastasis can be improved compared with CT alone. Fourthly, only several commonly used tumor markers were analyzed in this study, which may not be enough to reflect the full profile of tumor markers or clinical biomarkers.

In general, a novel pretreatment evaluation nomogram model based on blood level of tumor markers and size of pulmonary lesion was developed and validated. This pro-posed model is valuable in estimating the risk of lymph node or organ metastasis for unidentified lung lesions or identified lung cancer.

Acknowledgments

Author contributions

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

Disclosure

Yijie Zheng is an employee of Abbott Diagnostics Division, Abbott Laboratories. The authors report no other conflicts of interest in this work.

References

1. Ettinger DS, Wood DE, Aisner DL, et al. Non-small cell lung cancer, version 5.2017, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2017;15:504–535.
2. Scagliotti GV, Pastorino U, Vansteenkiste JF, et al. Randomized phase III study of surgery alone or surgery plus preoperative cisplatin and gemcitabine in stages IB to IIIA non-small-cell lung cancer. *J Clin Oncol*. 2012;30:172–178. doi:10.1200/JCO.2010.33.7089
3. Goldstraw P, Mitchell A, Bolejack V, et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage grouping in forthcoming (Eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2016;11(1):39–51. doi:10.1016/j.jtho.2015.09.009
4. Howlander N, Noone AM, Krapcho M, et al. editors. *Vintage 2009 Populations. National Cancer Institute*. Bethesda, MD: SEER Cancer Statistics Review; 1975–2009. Available from: http://seer.cancer.gov/csr/1975_2009_pops09(2012). Accessed September 09, 2019.
5. Vogel P, Daschner H, Lenz J, Schäfer R. Correlation of lymph node size and metastatic involvement of lymph nodes in bronchial cancer. *J Langenbecks Arch Chir*. 1990;.375:141–144. doi:10.1007/BF00206806
6. Silvestri GA, Gonzalez AV, Harris LJ, et al. Methods for staging non-small cell lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *J Chest*. 2013;143(5 Suppl):e211S–e250S. doi:10.1378/chest.12-2355
7. Detterbeck FC, Falen S, Rivera MP, Halle JS, Socinski MA. Seeking a home for a PET, part 1: defining the appropriate place for positron emission tomography imaging in the diagnosis of pulmonary nodules or masses. *J Chest*. 2004;125:2294–2299. doi:10.1378/chest.125.6.2294
8. Detterbeck FC, Falen S, Rivera MP, Halle JS, Socinski MA. Seeking a home for a PET, part 2: defining the appropriate place for positron emission tomography imaging in the staging of patients with suspected lung cancer. *Chest*. 2004;125:2300–2308. doi:10.1378/chest.125.6.2300
9. Rice SL, Friedman KP. Clinical PET-MR imaging in breast cancer and lung cancer. *J PET Clin*. 2016;11(4):387–402. doi:10.1016/j.cpet.2016.05.008

10. Diamandis EP, Hoffman BR, Sturgeon CM. National academy of clinical biochemistry laboratory medicine practice guidelines for the use of tumor markers. *J Clin Chem*. 2008;54(11):1935–1939. doi:10.1373/clinchem.2008.105494

11. Sturgeon CM, Hoffman BR, Diamandis EP; National Academy of Clinical Biochemistry. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines for use of tumor markers in clinical practice: quality requirements. *Clin Chem*. 2008;54(8):e1–e10. doi:10.1373/clinchem.2007.094144

12. Sturgeon CM, Hoffman BR, Diamandis EP; National Academy of Clinical Biochemistry. National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers. *Clin Chem*. 2008;54(12):e11–79. doi:10.1373/clinchem.2008.105601

13. Zeng Q, Liu M, Zhou N, Liu L, Song X. Serum human epididymis protein 4 (HE4) may be a better tumor marker in early lung cancer. *J Clin Chim Acta*. 2016;455:102–106. doi:10.1016/j.cca.2016.02.002

14. Huang W, Wu S, Lin Z, Chen P, Wu G. Evaluation of HE4 in the diagnosis and follow up of non-small cell lung cancers. *J Clin Lab*. 2017;63:461–467.

15. Wojcik E, Tarapacz J, Rychlik U, et al. Human epididymis protein 4 (HE4) in patients with small-cell lung cancer. *J Clin Lab*. 1625-32;62:2016.

16. Tozzoli R, Basso SM, D'Aurizio F, Metus P, Lumachi F. Evaluation of predictive value of pleural CEA in patients with pleural effusions and histological findings: a prospective study and literature review. *J Clin Biochem*. 2016;49:1227–1231. doi:10.1016/j.clinbiochem.2016.08.006

17. Chu XY, Hou XB, Song WA, Xue ZQ, Wang B, Zhang LB. Diagnostic values of SCC, CEA, Cyfra21-1 and NSE for lung cancer in patients with suspicious pulmonary masses: a single center analysis. *J Cancer Biol Ther*. 2011;11:995–1000. doi:10.4161/cbt.11.12.15526

18. Yang DW, Zhang Y, Hong QY, et al. Role of a serum-based biomarker panel in the early diagnosis of lung cancer for a cohort of high-risk patients. *J Cancer*. 2015;121:3113–3121. doi:10.1002/cncr.29551

19. Hui L, Rixv L, Xiuying Z. A system for tumor heterogeneity evaluation and diagnosis based on tumor markers measured routinely in the laboratory. *J. Clin Biochem*. 2015;.48:1241–1245. doi:10.1016/j.clinbiochem.2015.07.027

20. Molina R, Marrades RM, Augé JM, et al. Assessment of a combined panel of six serum tumor markers for lung cancer. *J Am J Respir Crit Care Med*. 2016;193:427–437. doi:10.1164/rccm.201404-0603OC

21. Howlader N, Noone AM, Krapcho M, et al. eds. *SEER Cancer Statistics Review*. Bethesda, MD: National Cancer Institute; 1975-2014. Available from: https://seer.cancer.gov/csr/1975_2014(2017). Accessed September 09, 2019.

22. Du Q, Yan C, Wu SG, 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. *J Cancer Lett*. 2018;420:236–241. doi:10.1016/j.canlet.2018.01.079

23. Horeweg N, de Koning HJ, Oudkerk M, 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;15:1332–1341. doi:10.1016/S1470-2045(14)70389-4

24. Chen ZQ, Huang LS, Zhu B. Assessment of seven clinical tumor markers in diagnosis of non-small-cell lung cancer. *J Dis Markers*. 2018;2018:9845123.

25. 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. *J Cancer*. 2018;124:262–270.

26. Wang Y, Li J, Xia Y, et al. Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. *J Clin Oncol*. 2013;31:1188–1195. doi:10.1200/JCO.2012.41.5984

27. Shariat SF, Capitanio U, Jeldres C, Karakiewicz PI. Can nomograms be superior to other prediction tools?J. *BJU Int*. 2009;103:492–495. doi:10.1111/j.1464-410X.2008.08073.x

28. Dong F, Shen Y, Gao F, et al. Nomograms to predict individual prognosis of patients with primary small cell carcinoma of the bladder. *J Cancer*. 2018;9(7):1152–1164. doi:10.7150/jca.23344

29. Lei Z, Li J, Wu D, et al. Nomogram for preoperative estimation of microvascular invasion risk in hepatitis B virus-related hepatocellular carcinoma within the Milan criteria. *J JAMA Surg*. 2016;151:356–363. doi:10.1001/jamasurg.2015.4257

30. Kalemkerian GP, Gadgeel SM. Modern staging of small cell lung cancer. *J Natl Compr Canc Netw*. 2013;11:99–104.

31. Mazumdar M, Glassman JR. Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. *J Stat Med*. 2000;19:113–132. doi:10.1002/(SICI)1097-0258(20000115)19:1<113::AID-SIM245>3.0.CO;2-O

32. Vickers AJEE. Decision curve analysis: a novel method for evaluating prediction models. *J Med Decis Macking*. 2006;26:565–574. doi:10.1177/0272989X06295361

33. Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and interpreting decision curve analysis: a guide for investigators. *J Eur Urol*. 2018;74(6):796–804. doi:10.1016/j.eururo.2018.08.038