



Research paper

Development and validation of a novel scoring system developed from a nomogram to identify malignant pleural effusion



Sufei Wang^{a,1}, Shan Tian^{b,1}, Yuan Li^c, Na Zhan^d, Yingyun Guo^b, Yu Liu^e, Juanjuan Xu^a, Yanling Ma^a, Shujing Zhang^a, Siwei Song^a, Wei Geng^a, Hui Xia^a, Pei Ma^a, Xuan Wang^a, Tingting Liao^a, Yanran Duan^f, Yang Jin^{a,*}, Weiguo Dong^{b,*}

^a Department of Respiratory and Critical Care Medicine, NHC Key Laboratory of Pulmonary Diseases, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, No.1277 Jiefang Avenue, Wuhan, Hubei 430022, China

^b Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan University, Wuhan, Hubei 430060, China

^c Department of Oncology, Renmin Hospital of Wuhan University, Wuhan University, Wuhan, Hubei 430060, China

^d Department of Pathology, Renmin Hospital of Wuhan University, Wuhan University, Wuhan, Hubei 430060, China

^e Health Checkup Department, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430022, China

^f School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430022, China

ARTICLE INFO

Article History:

Received 8 February 2020

Revised 29 June 2020

Accepted 13 July 2020

Available online xxx

Key words:

Malignant pleural effusion

Clinical features

Nomogram

Diagnostic value

Scoring system

ABSTRACT

Background: This study aimed to establish and validate a novel scoring system based on a nomogram for the differential diagnosis of malignant pleural effusion (MPE) and benign pleural effusion (BPE).

Methods: Patients with PE and confirmed aetiology who underwent diagnostic thoracentesis were included in this study. One retrospective set ($N = 1261$) was used to develop and internally validate the predictive model. The clinical, radiological and laboratory features were collected and subjected to logistic regression analyses. The primary predictive model was displayed as a nomogram and then modified into a novel scoring system, which was externally validated in an independent set ($N = 172$).

Findings: The novel scoring system was composed of fever (3 points), erythrocyte sedimentation rate (4 points), effusion adenosine deaminase (7 points), serum carcinoembryonic antigen (CEA) (4 points), effusion CEA (10 points) and effusion/serum CEA (8 points). With a cutoff value of 15 points, the area under the curve, specificity and sensitivity for identifying MPE were 0.913, 89.10%, and 82.63%, respectively, in the training set, 0.922, 93.48%, 81.51%, respectively, in the internal validation set and 0.912, 87.61%, 81.36%, respectively, in the external validation set. Moreover, this scoring system was exclusively applied to distinguish lung cancer with PE from tuberculous pleurisy and showed a favourable diagnostic performance in the training and validation sets.

Interpretation: This novel scoring system was developed from a retrospective study and externally validated in an independent set based on six easily accessible clinical variables, and it exhibited good diagnostic performance for identifying MPE.

Funding: NFSC grants (no. 81572942, no. 81800094).

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Pleural effusion (PE) is a common but difficult problem in clinical settings. Approximately 1.5 million patients with PE are diagnosed each year in the United States [1]. PE is a symptom caused by more than 50 diseases and is usually classified as malignant pleural effusion (MPE) or benign pleural effusion (BPE). The incidence of MPE is

higher than 200,000 cases/year in the United States [2], with lung cancer (36%) and breast cancer (26%) accounting for the vast majority of cases due to metastasis to the pleura [3,4]. With regard to BPE, tuberculous pleurisy effusion (TPE) is the most common cause [5] and a prominent problem in developing countries, including China [6]. The prognosis for patients with MPE is extremely poor in clinical practice [7,8], and delaying treatment due to misdiagnosis will directly lead to an increase in mortality. Therefore, developing a useful method that can identify MPE as early as possible with precision is highly important.

Cytology is most commonly used in clinical practice to identify tumour cells [9], which are the most powerful evidence in support of

* Corresponding authors.

E-mail addresses: whuhjy@126.com (Y. Jin), dongweiguo@whu.edu.cn (W. Dong).

¹ Sufei Wang and Shan Tian contributed equally to this work.

Research in context

Evidence before this study

Differentiating malignant pleural effusion (MPE) from benign pleural effusion (BPE) remains challenging in clinical practice; thus, developing a useful method that can precisely identify MPE as early as possible is highly important. Previous studies have attempted to differentiate MPE from BPE using various indexes. In addition to the limited sample sizes, most clinical studies have not established a quick and novel scoring system for physicians to use in clinical practice. More importantly, no investigations have used another independent set to externally validate their predictive model.

Added value of this study

We attempted to develop a predictive model using multivariate logistic regression based on a combination of the most widely available clinical features and radiological parameters and a variety of laboratory indexes. Then, the primary predictive model was displayed as a nomogram and modified into a novel scoring system for convenient clinical use. Finally, we not only internally validated the diagnostic performance of the modified scoring model in a retrospective study but also externally verified the scoring system in an independent set.

Implications of all the available evidence

In the present study, we selected the most significant indexes (fever, ADA, ESR, serum CEA, effusion CEA and effusion/serum CEA) to construct a predictive model. With the aim of establishing a novel scoring system, we converted the nomogram into a scoring system. This scoring system showed good diagnostic performance. Furthermore, we also validated the scoring system in an independent set to evaluate its predictability in an independent population, and the result was encouraging. This feasible and novel scoring system for differentiating patients with MPE and BPE should be further validated in multicentre prospective studies in the near future.

independent cohort to externally validate their predictive model [17,18]. Hence, the results from the previous studies were inconclusive, and the predictive models for diagnosing MPE were not good enough to be recommended in routine clinical practice.

In the present study, our primary goal was to develop a predictive model using multivariate logistic regression based on a combination of clinical features to distinguish MPE from BPE. Then, the primary predictive model was displayed as a nomogram and modified into a novel scoring system for convenient clinical use. In addition, we not only internally validated the diagnostic performance of the modified scoring model in a retrospective study but also externally verified the scoring system in another tertiary hospital. Our second aim was to exclusively apply the scoring system to differentiate lung cancer with PE from TPE and to evaluate the system's discriminative and calibration abilities for these two diseases.

2. Materials and methods

2.1. Patients and study design

This clinical study was performed at Renmin Hospital of Wuhan University and Union Hospital, Tongji Medical College, Huazhong University of Science and Technology (Wuhan Union Hospital). This clinical study consisted of two stages. The first stage (training set and internal validation set) retrospectively identified patients with PE in Renmin Hospital of Wuhan University from January 2014 to April 2018. The second stage (external validation set) included PE patients in Wuhan Union Hospital from August 2019 to December 2019. All patients who underwent diagnostic thoracentesis at the Departments of Respiratory, Oncology and Chest Surgery were included in this clinical study. All patients needed to have a confirmed cause of PE for final inclusion in this study.

Participants who met the following inclusion criteria were included in this study: (a) confirmed to have PE by chest CT, X-ray or ultrasonography; and (b) underwent diagnostic thoracentesis or, in some cases, pleural biopsy. The exclusion criteria were as follows: (a) indeterminable cause of PE; (b) younger than 18 years old; (c) no information available; and (d) reluctant to participate in this study.

The primary aim of the present study was to develop a scoring system with high predictive accuracy to accurately differentiate MPE from BPE, and the secondary aim was to differentiate lung cancer with PE from TPE. The patients were divided into three groups (training set, internal validation set and external validation set). The first group (the training set) included 70% of the patients with PE ($N = 883$) from the retrospective study carried out at Renmin Hospital of Wuhan University to develop the scoring system to discriminate MPE patients from patients with BPE. The second group (the internal validation set) included the remaining 30% participants with PE ($N = 378$) from Renmin Hospital of Wuhan University to validate the diagnostic performance of this scoring system. The third group (the external validation set) included prospectively recruited patients with PE ($N = 172$) in Wuhan Union Hospital, independent of the patients in Renmin Hospital of Wuhan University, to further validate the predictive model. This study was conducted in line with the Declaration of Helsinki. The institutional ethics committees of Renmin Hospital of Wuhan University (No. WDRY 2019-K014) and Wuhan Union Hospital (No. 2019-S075) reviewed and approved this study protocol. Moreover, the second-stage study performed at Wuhan Union Hospital was registered on the Clinical Trials website (No. NCT03997669). All patients were required to provide written informed consent.

2.2. Data collection

The clinical information, including demographic data, objective symptoms, and radiological and laboratory features, of the included patients were obtained from the electronic medical record (EMR).

MPE. In a recent clinical study that included 725 patients with MPE, the sensitivity of cytology for diagnosing MPE was only 63% [10]. Moreover, the assessment of cytology results is relatively subjective, and the diagnostic accuracy of this method largely depends on the expertise of the pathologist. Additionally, reactive mesothelial cells are easily confused with tumour cells due to similar morphological features [11]. Although cell block plays an important role in the identification of MPE, this method is usually time-consuming and has high costs, and its sensitivity is less than 50% [12]. Although pleural biopsy has a relatively higher diagnostic accuracy [13], it is invasive and associated with more complications, as well as higher costs. Hence, it is important to design a cost-effective method that is more accurate and less invasive to identify MPE [14].

A variety of biological markers have been investigated in several studies for the differential diagnosis of MPE and BPE, but most of the biomarkers have been used individually, yielding an unfavourable diagnostic performance [15,16]. The results from a few clinical studies showed that no single index from blood or effusion biochemistry could provide sufficient evidence to accurately distinguish MPE from BPE in clinical practice. Of course, some studies have also attempted to differentiate MPE from BPE using various indexes. In addition to the limited sample sizes, most clinical studies have not established a quick and novel scoring system for physicians to use in clinical practice [14]. More importantly, no investigations used another

The following clinical features were abstracted from the platform: demographic characteristics (age and gender), objective symptom (fever), radiological features (site of PE, volume of PE) and laboratory indexes. Additionally, some ratios commonly used in clinical practice were automatically calculated and incorporated into this study. All laboratory indexes were converted to categorical variables according to the reference range, and the combined ratios were converted to binary variables according to the optimal cutoff values via receiver operating characteristic (ROC) analyses. Additionally, fever was defined as a body temperature over 37.5 °C. The volume of PE was categorized as mild (<500 ml), moderate (500–1000 ml) or severe (>1000 ml).

2.3. Diagnostic criteria for MPE and BPE

If cancer cells were detected in the PE with cytological smears, cell blocks, or pleural biopsy, then the patient was diagnosed with MPE. In contrast, BPE was identified by a known aetiology, such as TPE or parapneumonic effusion, without any signs of cancer. TPE was established when Lowenstein–Jensen cultures or acid-fast stains of PE, sputum, bronchoalveolar lavage fluid, or pleural biopsy were positive. Parapneumonic effusion was diagnosed when PE coincided with pneumonia and was resolved after antibiotic treatment. Moreover, other aetiologies of PE followed strict clinical criteria (Table S1).

2.4. Statistical analysis

SPSS 22.0 (IBM Inc., Chicago, IL, USA) was used to conduct the statistical analyses. Categorical data (laboratory indexes) were expressed as frequencies with percentages. The differences between MPE and BPE groups were analysed with the χ^2 test or Fisher's exact test, whereas the continuous variable (age) was presented as the mean with standard deviation, and differences between the MPE and BPE groups were compared with Student's t-test. Logistic regression analyses were performed in the training set, and the statistically significant variables with an area under the curve (AUC) > 0.6 were further selected for multivariable logistic regression analyses. The regression coefficients were regarded as the weights for the variables in the predictive model. The nomogram applied to create the scoring system was developed with independent risk factors based on multivariate logistic analysis using the rms package in R. A model score for each patient was automatically calculated and evaluated by a ROC

curve. The optimal cutoff value was attained based on the Youden index, and specificity, sensitivity, positive likelihood ratio (PLR) and negative likelihood ratio (NLR) were applied to assess the diagnostic performance of the nomogram both in the training set and validation sets. Two-sided $P < 0.05$ was regarded as statistically significant.

3. Results

3.1. Study populations

A total of 1261 participants from Renmin Hospital of Wuhan University were included in this research and 70% ($N = 883$) were randomly assigned to the training set while the remaining participants ($N = 378$) were included in the internal validation set. Furthermore, 172 patients from Wuhan Union Hospital were included in the external validation set. A detailed flow diagram of patient selection is listed in Fig. 1. The distributions of age, gender and MPE/BPE ratio in the training and validation sets are presented in Table 1.

3.2. Construction of a nomogram and novel scoring system in the training set

In the training set, most of the indexes included in this study were significantly different between the patients with BPE and MPE (Table S2). Moreover, 23 parameters remained significant after the univariate regression analysis (Table S3). To construct a highly accurate predictive model, the significant parameters with an AUC > 0.6 were further subjected to a multivariate regression model. Six valuable factors (fever, ESR, effusion ADA, serum CEA, effusion CEA and effusion serum CEA) were selected to establish the predictive model (Fig. 2). Then, a nomogram (Fig. 3a) to discriminate MPE from BPE was built on the basis of a multivariate logistic regression model. This diagnostic nomogram possessed good discriminative ability, as reflected by an AUC of 0.914 (95% CI=0.888–0.935) (Figure S1a). The calibration curve showed that this diagnostic nomogram exhibited good calibration (Fig. 3b). Moreover, decision curve analysis (DCA) was applied to assess the clinical utility of the diagnostic nomogram. As shown in Fig. 3c, if the threshold probability is 0.45, patients with PE would benefit more from using this diagnostic nomogram than the treating all or treating none scenarios.

To make this predictive model more convenient for physicians to use in clinical practice, we modified the nomogram into a scoring

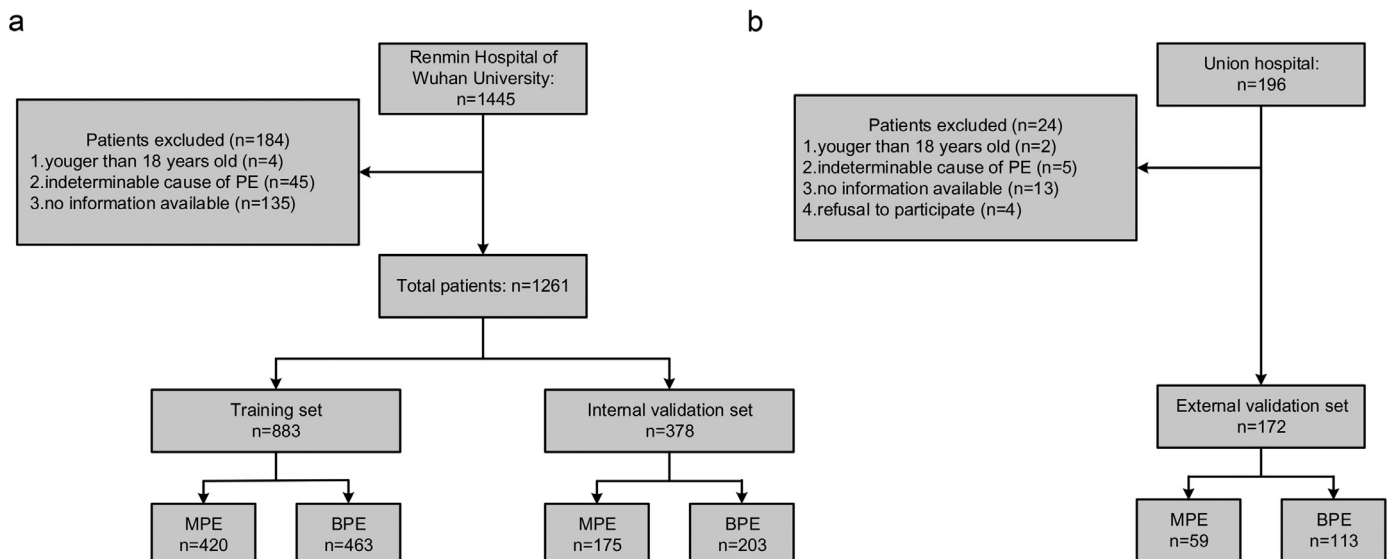


Fig. 1. Flowchart of participant selection: (a) Renmin Hospital of Wuhan University set; and (b) Wuhan Union Hospital set.

Table 1
Baseline characteristics of the training set and validation set.

Features	Training set (N = 883)	Internal validation set (N = 378)	External validation set (N = 172)	Score (points) median (IQR)
Age (years)	60.62±16.43	60.44±16.78	57.78±1.13	–
Gender:				
Female	312 (35.33%)	143 (37.83%)	62 (36.05%)	–
Male	571 (64.67%)	235 (62.17%)	110 (63.95%)	–
MPE				
Lung cancer	316 (35.80%)	131 (34.66%)	51 (29.65%)	32.0 (24.0–33.0)
Breast cancer	12 (1.36%)	3 (0.79%)	2 (1.16%)	34.0 (29.0–36.0)
Lymphoma	24 (2.72%)	6 (1.59%)	1 (0.58%)	21.0 (8.0–26.0)
Mesothelioma	6 (0.68%)	1 (0.26%)	1 (0.58%)	28.5 (22.0–32.0)
Ovarian cancer	5 (0.57%)	4 (1.06%)	0 (0%)	26.5 (14.0–32.0)
Other cancers	57 (6.45%)	30 (7.93%)	4 (2.33%)	28.5 (18.0–32.0)
BPE				
Tuberculous pleurisy	149 (16.87%)	77 (20.37%)	56 (32.56%)	3.0 (0–7.0)
Parapneumonic effusion	202 (22.87%)	83 (21.96%)	35 (20.35%)	10.0 (4.0–14.0)
Heart failure	37 (4.19%)	12 (3.17%)	7 (4.07%)	12.5 (10.0–14.0)
Pulmonary embolism	2 (0.22%)	1 (0.26%)	0 (0%)	14
Empyema	13 (1.47%)	7 (1.82%)	5 (2.91%)	3.0 (0–4.0)
Other benign diseases	60 (6.8%)	23 (6.08%)	10 (5.81%)	10.0 (3.0–14.0)

MPE=malignant pleural effusion; BPE=benign pleural effusion; IQR= interquartile range.

Other benign diseases: cirrhosis, nephrotic syndrome, pericardial disease, hypoproteinaemia, parasitic infection, systemic lupus erythematosus, rheumatoid arthritis, chylothorax.

Other cancers: oesophageal cancer, gastric cancer, renal cancer, liver cancer, cervical cancer, colorectal cancer, cholangiocarcinoma, endometrial cancer, nasopharyngeal cancer, pancreatic cancer, adrenal carcinoma, prostate cancer, thyroid cancer and cancer of unknown origin.

system with integer points: fever (3 points), ESR (4 points), effusion ADA (7 points), serum CEA (4 points), effusion CEA (10 points) and effusion/serum CEA (8 points) (Table 2).

3.3. Diagnostic performance of the scoring system in the training set

In the training set, based on the cutoff value of 15 points (Table 3), patients with PE were more likely to be diagnosed with MPE when the total number of points was more than 15 while the PE patients were less likely to be diagnosed with MPE when the total number of points was less than 15. Moreover, this scoring system showed good discriminative power in differentiating between MPE and BPE, as reflected by an AUC of 0.913 (95% CI=0.887–0.935) (Fig. 4a). When the optimal cutoff points were set at 15, the corresponding specificity, sensitivity, PLR and NLR values were 89.10%, 82.63%, 7.58 and 0.20, respectively (Table 4). Additionally, this scoring system also showed good calibration, implying that no significant difference existed between the predicted probability of the scoring system and the actual probability (Fig. 4c).

3.4. Diagnostic performance of the scoring system in the validation set

In the internal validation set, this scoring model exhibited favourable discriminative power, as reflected by an AUC of 0.922 (95% CI=0.883–0.952) (Fig. 4b). When the optimal cutoff points were also set at 15, the corresponding specificity, sensitivity, PLR and NLR values were 93.48%, 81.51%, 12.5 and 0.20, respectively (Table 4). Moreover, this scoring system also showed good calibration in the internal validation set, as displayed in Fig. 4e.

External validation was performed in an independent set to further validate the diagnostic performance of the scoring system in discriminating MPE from BPE. Good discrimination (AUC=0.912, 95% CI=0.859–0.950) was also observed in the external validation set (Fig. 4c). When the optimal cutoff score was also set at 15, the corresponding specificity, sensitivity, PLR and NLR values were 87.61%, 81.36%, 6.57 and 0.21, respectively (Table 4). Furthermore, as shown in Fig. 4f, this scoring system possessed good calibration in the external validation set.

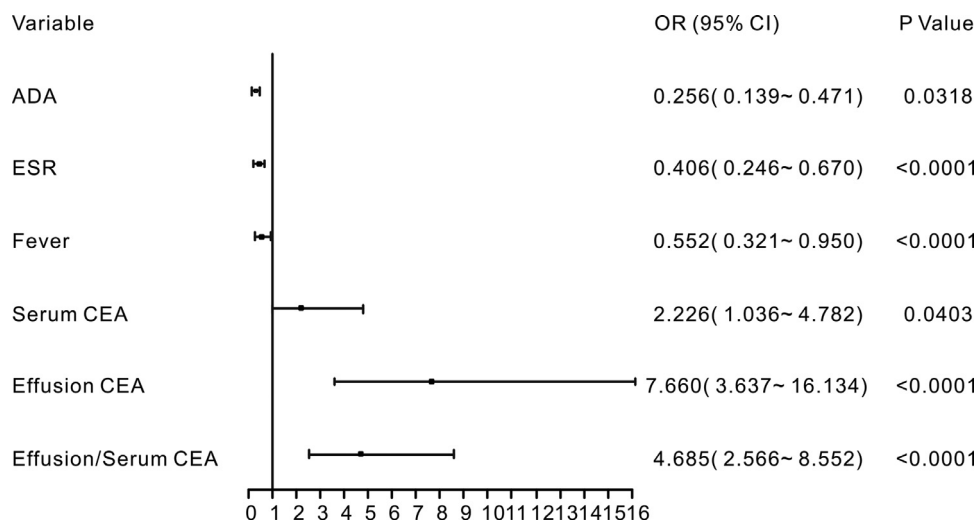


Fig. 2. Forest plot of the significant parameters in the multivariate regression analysis.

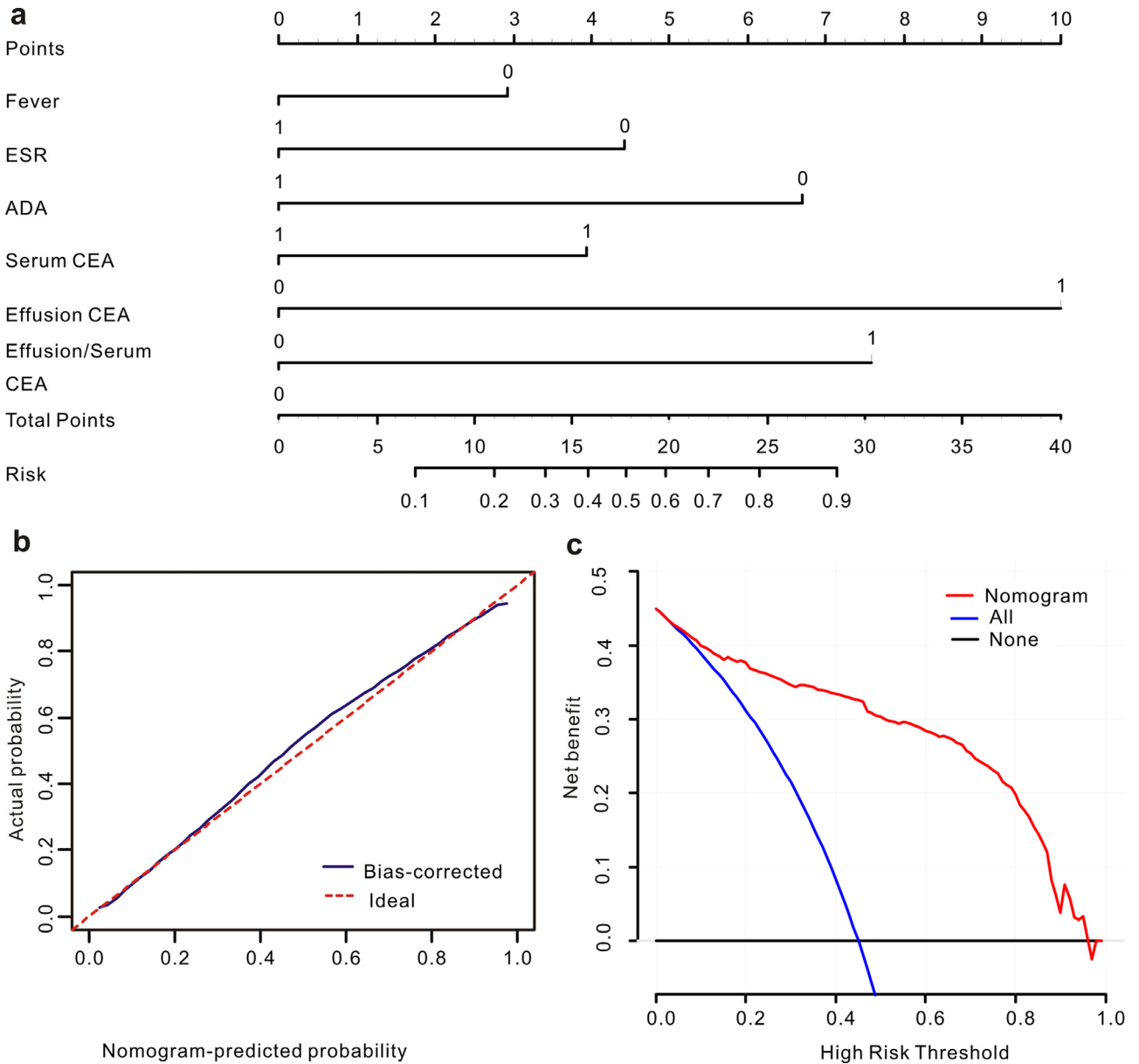


Fig. 3. Calibration and clinical use of a diagnostic nomogram for the discrimination of MPE and BPE. (a) Diagnostic nomogram for identifying MPE from BPE. (b) Calibration curve of the diagnostic nomogram. (c) DCA of the diagnostic nomogram.

3.5. Diagnostic significance of the scoring system for differentiating between lung cancer with PE and TPE

As TPE with atypical symptoms and laboratory indexes is usually confused with lung cancer with PE in a clinical setting, we exclusively applied the scoring system to differentiate between these two diseases. Encouragingly, the scoring system modified from the nomogram exhibited favourable diagnostic performances in the training and validation sets. As shown in Fig. 5a-c, the AUC values of this scoring system for distinguishing lung cancer with PE from TPE were 0.978 (95% CI=0.954–0.991) in the training set, 0.966 (95% CI=0.922–0.989) in the internal validation set, and 0.968 (95% CI=0.914–0.992) in the external validation set. The values of the other diagnostic parameters, such as specificity, sensitivity, PLR and NLR, when the total score was over 11 points are listed in Table 4. In addition, as shown in Fig. 5d-f, this scoring system used for the differential diagnosis of lung cancer with PE and TPE was also well calibrated in all three data sets.

4. Discussion

MPE is a common symptom of lung cancer, breast cancer and other cancer-related pleural lesions and indicates an advanced stage of cancer [19,20]. Moreover, the presence of MPE can seriously impair quality of life and shorten the survival of patients; the median survival time of patients with MPE ranges from 3 to 12 months [21,22]. However, patients with BPE, such as TPE and parapneumonic effusions, if treated in time, can usually be clinically cured. Therefore, promptly discriminating MPE from BPE with high accuracy is essential for therapeutic decisions and thus improving the prognosis of patients with MPE.

There have been a variety of attempts to create a more effective method to discriminate between MPE and BPE. Unfortunately, no specific method for differentiating MPE from BPE is perfect in clinical settings. Yang et al. [23] established a scoring system based on five PET-CT parameters for the differential diagnosis of MPE and BPE, and the sensitivity and specificity in the derivation cohort were 83.3% and

Table 2

A novel scoring system developed from a nomogram of the training set.

Parameters	Score generated from nomogram (points)	Score modified from nomogram (points)
Fever (No)	2.92	3
ESR (≤ 43 mm/h)	4.42	4
ADA (≤ 25.74 U/L)	6.7	7
Serum CEA (> 5 ng/mL)	3.93	4
Effusion CEA (> 5 ng/mL)	10	10
Effusion/serum CEA (> 1.66)	7.58	8

ESR=erythrocyte sedimentation rate; ADA=adenosine deaminase; CEA=carcinoembryonic antigen.

92.2%, respectively, with a cutoff value of 4. In addition, Porcel et al. [24] constructed a CT scan-based scoring system, and this scoring model achieved a sensitivity of 88% and a specificity of 94% with a

cutoff value of 7 points in the validation set. Moreover, Liu et al. [25] combined Raman spectra bands of PE with orthogonal partial least squares discriminant analysis to distinguish MPE from BPE, and its sensitivity and specificity reached 92.2% and 93.8%, respectively. Despite the relatively higher diagnostic accuracy, the above methods could not be applied in most hospitals, especially primary hospitals, due to the strict requirements for medical equipment. Furthermore, these methods are not good enough to be recommended in routine clinical practice due to inconvenience. Hence, designing a method that is not only feasible and simple but that could also accurately differentiate MPE from BPE with high calibration is important. In contrast, the indexes involved in our scoring system were easily obtained, and more importantly, the total cost of the indexes involved in the scoring model was acceptable (180 RMB) and lower than that of the commonly used cell block (220 RMB).

Although numerous clinical features are different between MPE patients and patients with BPE, the clinical significance of a single

Table 3

ROC analysis of the scoring system for identifying MPE in the training set.

Cutoff score	Youden index	Sensitivity% (95%CI)	Specificity% (95%CI)	PLR (95%CI)	NLR (95%CI)
> 12	0.6425	90.73 (86.5–94.0)	73.52 (68.3–78.3)	3.43 (2.8–4.1)	0.13 (0.09–0.20)
> 14	0.7101	83.78 (78.7–88.1)	87.23 (83.1–90.7)	6.56 (4.9–8.8)	0.19 (0.14–0.25)
> 15	0.7173	82.63 (77.5–87.0)	89.1 (85.2–92.3)	7.58 (5.5–10.4)	0.20 (0.15–0.25)
> 17	0.7165	82.24 (77.0–86.7)	89.41 (85.5–92.6)	7.76 (5.6–10.7)	0.20 (0.15–0.26)
> 18	0.6881	77.22 (71.6–82.2)	91.59 (88.0–94.4)	9.18 (6.4–13.3)	0.25 (0.20–0.31)

CI=confidence interval; PLR=positive likelihood ratio; NLR=negative likelihood ratio.

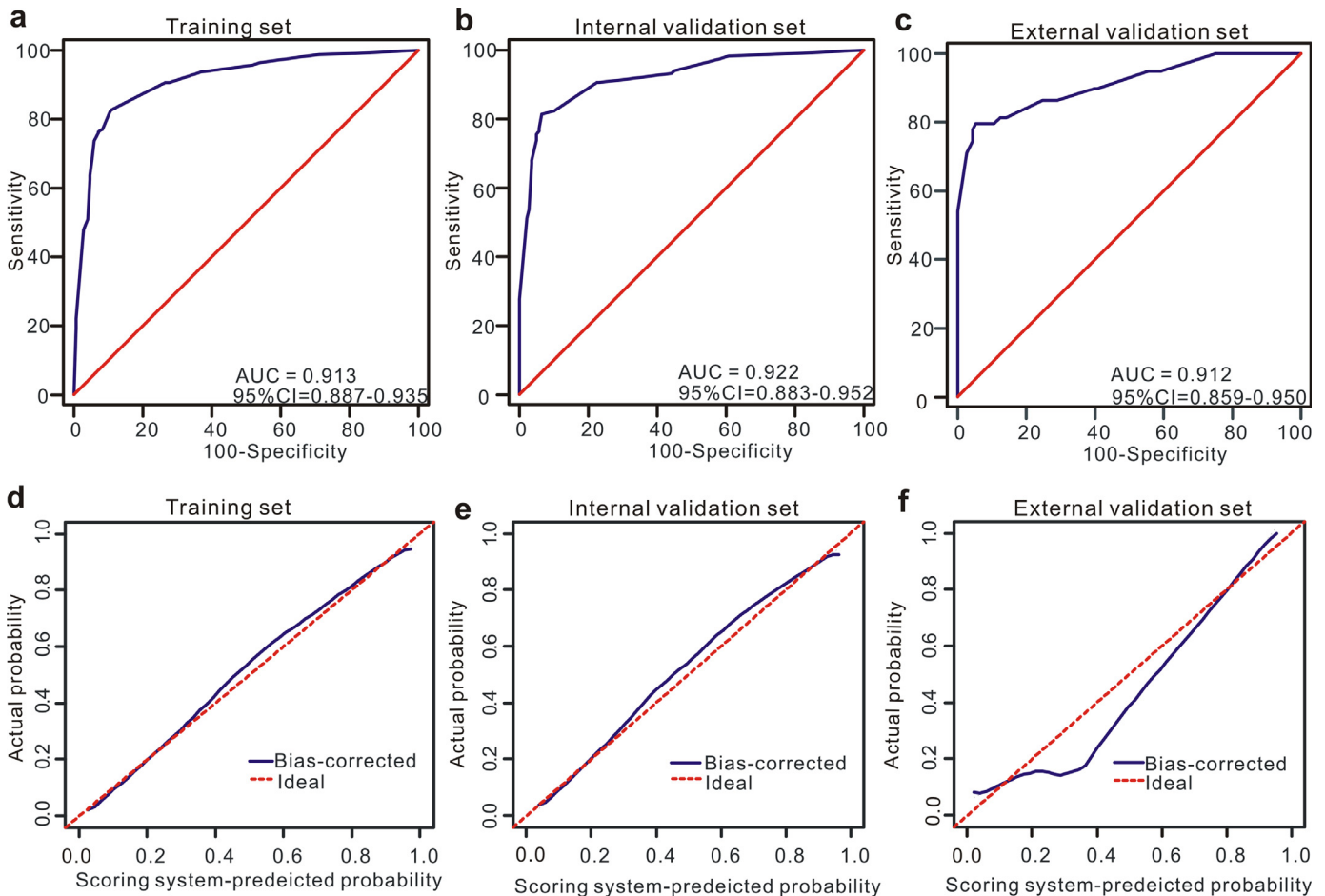


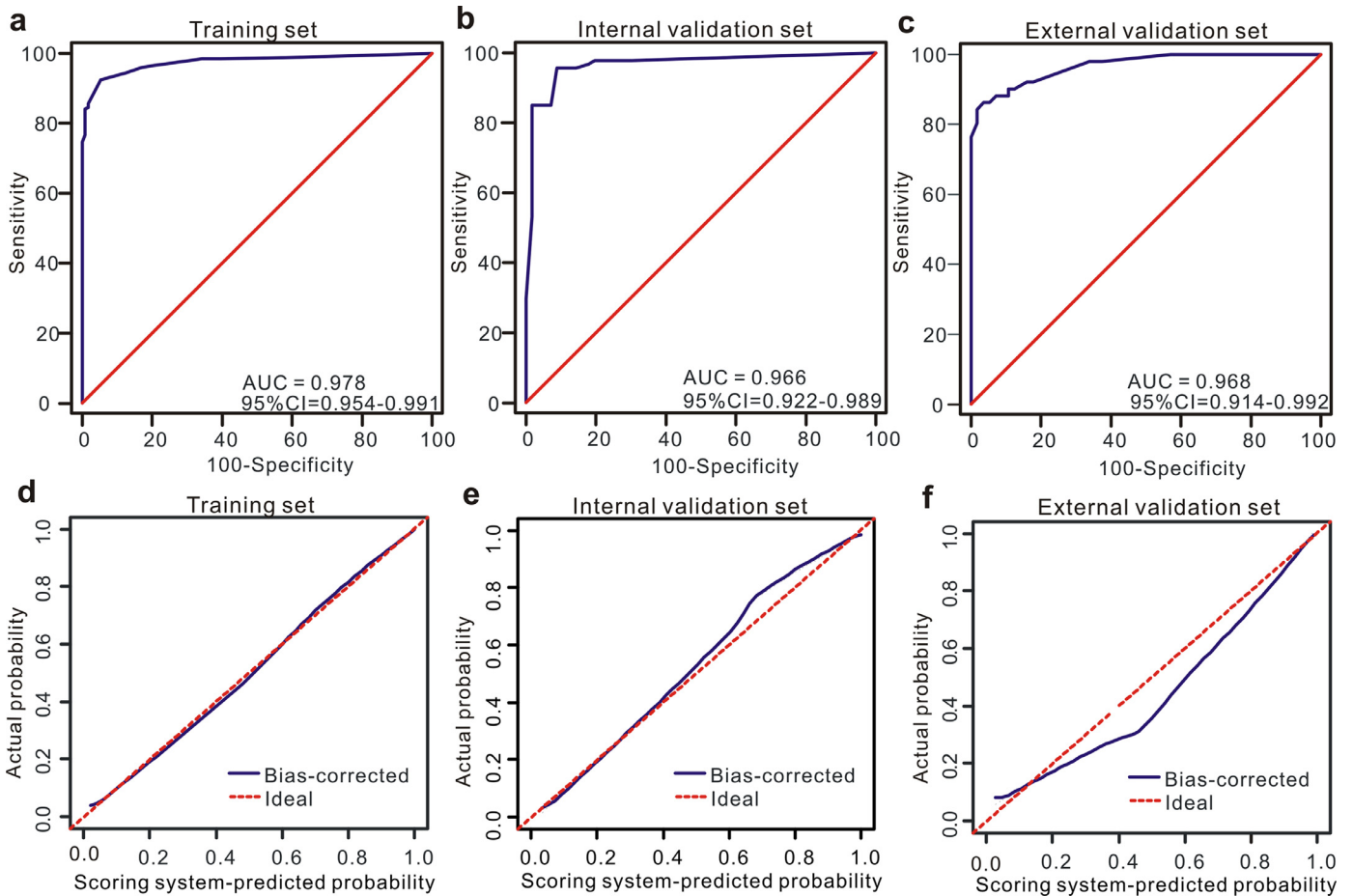
Fig. 4. Discrimination and calibration of the scoring system for discrimination of MPE and BPE. ROC curves of the scoring system in the training set (a), internal validation set (b) and external validation set (c). Calibration curves of the scoring system in the training set (d), internal validation set (e) and external validation set (f).

Table 4

Diagnostic performance of the scoring system in differentiating MPE from BPE and lung cancer with PE from TPE in the training and validation sets.

Variables	MPE/BPE			Lung cancer with PE/TPE		
	Training set	Internal validation set	External validation set	Training set	Internal validation set	External validation set
AUC (95%CI)	0.913 (0.887–0.935)	0.922 (0.883–0.952)	0.912 (0.859–0.950)	0.978 (0.954–0.991)	0.966 (0.922–0.989)	0.968 (0.914–0.992)
Sensitivity (95%CI)	82.63% (77.5%–87.0%)	81.51% (73.4%–88.0%)	81.36% (69.1%–90.3%)	92.42% (87.8%–95.7%)	95.56% (89.0%–98.8%)	90.2% (78.6%–96.7%)
Specificity (95%CI)	89.10% (85.2%–92.3%)	93.48% (88.0%–97.0%)	87.61% (80.1%–93.1%)	94.17% (91.9%–99.4%)	89.47% (78.5%–96.0%)	89.29% (78.1%–96.0%)
PLR (95%CI)	7.58 (5.5–10.4)	12.5 (6.6–23.6)	6.57 (4.0–10.9)	32.35 (10.59–98.74)	9.08 (4.30–19.42)	8.42 (3.89–18.02)
NLR (95%CI)	0.20 (0.15–0.25)	0.20 (0.14–0.29)	0.21 (0.12–0.36)	0.078 (0.05–0.13)	0.05 (0.02–0.14)	0.11 (0.05–0.26)

AUC=area under curve; CI=confidence interval; PLR=positive likelihood ratio; NLR=negative likelihood ratio; TPE=tuberculous pleural effusion.

**Fig. 5.** Diagnostic ability and calibration of the scoring system for discrimination of lung cancer with PE and TPE. ROC curves of the scoring system in the training set (a), internal validation set (b) and external validation set (c). Calibration curves of the scoring system in the training set (d), internal validation set (e) and external validation set (f).

index in the discrimination between the two conditions is quite limited due to either its low sensitivity or specificity. In recent years, with the development of analytical approaches, the construction of mathematical models based on multiple markers has been increasingly applied in the field of medicine [26–30]. This approach combines a series of significant parameters to generate a predictive model to achieve a better diagnostic performance. In the present study, we selected the most significant indexes (fever, ADA, ESR, serum CEA, effusion CEA and effusion/serum CEA) based on the β -coefficient generated by multivariate regression analysis to construct a predictive model. With the aim of establishing a novel scoring system, we converted the nomogram into a scoring system. This scoring system created in this study showed good diagnostic performance in the derivation and validation sets. Our study aimed to design a novel quantitative tool so that clinicians can predict the probability of MPE and BPE. Thus, our scoring system was based on

various clinical and laboratorial indexes, which are accessible in most hospitals, even primary hospitals. This present study integrated a total of thirty-four indexes, including not only primary indexes but also informative ratios reported in other studies, such as effusion/serum CEA and effusion NC/LC. A recent study revealed that the effusion NC/LC was lower in patients with TPE than in MPE patients and exhibited a relatively lower diagnostic significance [31]. Despite a significant difference in effusion NC/LC ($P < 0.0001$) in MPE and BPE in our training set, this difference changed insignificantly after the multivariate regression analysis (OR=0.753, 95%CI=0.543–1.044), $P = 0.0887$). Moreover, Hackner et al. [15] demonstrated that effusion/serum CEA was a useful marker in predicting MPE, with a sensitivity of 85% and specificity of 92%. Our study also found that effusion/serum CEA had higher diagnostic significance in the discrimination between MPE and BPE, as reflected by an AUC of 0.787. To the best of our knowledge, our study incorporated the most

laboratory parameters easily accessible in clinical practice to identify MPE from BPE with the largest sample to date.

The differential diagnosis between MPE and BPE can be challenging in a clinical setting because these conditions sometimes share similar clinical symptoms or biochemical indexes, thus leading to high difficulty. Therefore, many researchers have attempted to solve this clinical problem. Ren et al. [14] used four machine learning algorithms to build TPE diagnostic models based on 28 different features and obtained an acceptable diagnostic performance. However, this retrospective study did not construct a scoring system that could be applied in a clinical setting. In addition, a retrospective study [17] reported that a scoring model generated by logistic regression was successfully applied to identify TPE from MPE. However, no additional set was used to validate the scoring system, and this study did not include tumour biomarkers (neither serum CEA nor effusion CEA) in the scoring system. Our study not only considered serum and effusion CEA but also included other tumour markers, such as SCC and NSE-1. Our results showed that serum CEA, effusion CEA and effusion/serum CEA were significantly different between MPE and BPE according to multivariate regression analysis; thus, they were incorporated into the scoring system. Additionally, based on the diagnostic nomogram, we proposed a novel scoring system for the differential diagnosis of MPE and BPE. Very slight changes in the AUC (Fig. S1) were observed between the diagnostic nomogram and the novel scoring system in the training and validation sets, which suggests that the scoring system that was developed from the nomogram is not only convenient to use but also has robust diagnostic significance.

We recommend the wide application of the novel scoring model in most hospitals for a quick and presumptive discrimination between MPE and BPE, as this system is based on six easily accessible clinical variables. For PE patients with an unknown aetiology in a primary hospital and no hardware conditions for PET-CT or pleural biopsy, if the total score is over 15 points, then a high probability of malignancy exists. Our scoring model could provide an important reference for referrals to tertiary hospitals for further examinations, such as cell block, PET-CT, pleural biopsy or molecular tests, which are helpful for the selection of therapeutic regime and prediction of prognosis [32,33].

This present study is not exempt from limitations. First, this study was based on retrospective data, and no published data were used to assess the validity of the data, and thus the validity of the retrospective data was limited. Moreover, the variable of time to diagnosis was not included in the logistic regression because of the lack of data in the retrospective analysis. Next, although we verified the scoring system to discriminate between MPE and BPE in an independent validation set, the sample size in the external validation set was relatively small, and thus the external validation was only performed at a single medical centre. Finally, given that testing the fluid/serum levels of CEA is not routine in all parts of the world, our novel scoring system is limited in terms of the broader applicability in those countries. Therefore, multicentre validation of the scoring system with a large study population is urgently needed to obtain high-level evidence for its clinical application in the future.

In conclusion, the indexes of fever, ESR, effusion ADA, serum CEA, effusion CEA and effusion/serum CEA are significant for discriminating between MPE and BPE. This novel scoring system that was developed from a nomogram shows a favourable diagnostic performance and good calibration in the discrimination between MPE and BPE, especially in distinguishing between lung cancer with PE and TPE. This feasible and novel scoring system used to identify patients with MPE from those with BPE should be further validated in multicentre prospective studies in the near future.

Authors' contributions

Authors' contributions: WGD and YJ are the guarantors and take responsibility for the manuscript, including the data and analysis.

SFW, ST, YL and YL conceived and designed the study. NZ, YYG, JJX, YLM, SJZ, SWS, WG, HX, PM, XW and TTL acquired the data. YRD and ST analysed the data. SFW and ST drafted the manuscript. All authors approved the final version for submission.

Declarations of Competing Interests

All authors have no competing interests to report.

Acknowledgments

Not applicable.

Funding sources

This work was supported by the National Natural Science Foundation of China (no. 81572942 for Yang Jin, no. 81800094 for Juanjuan Xu). The funders had no role in the design or performance of this study; the collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2020.102924.

References

- [1] Feller-Kopman D, Light R. Pleural Disease. *N Engl J Med* 2018;378:740–51.
- [2] Murthy P, Ekeke CN, Russell KL, Butler SC, Wang Y, Luketich JD, et al. Making cold malignant pleural effusions hot: driving novel immunotherapies. *Oncimmunology* 2019;8:e1554969.
- [3] Ferreiro L, Toubes ME, San José ME, Suárez-Antelo J, Golpe A, Valdés L. Advances in pleural effusion diagnostics. *Expert Rev Respir Med* 2020;14:51–66.
- [4] Taghizadeh N, Fortin M, Tremblay A. US hospitalizations for malignant pleural effusions: data from the 2012 national inpatient sample. *Chest* 2017;151:845–4.
- [5] Wang G, Wang S, Yang X, Sun Q, Jiang G, Huang M, et al. Accuracy of Xpert MTB/RIF ultra for the diagnosis of pleural TB in a multicenter cohort study. *Chest* 2019.
- [6] Xu HY, Li CY, Su SS, Yang L, Ye M, Ye JR, et al. Diagnosis of tuberculous pleurisy with combination of adenosine deaminase and interferon-gamma immunospot assay in a tuberculosis-endemic population: a prospective cohort study. *Medicine (Baltimore)* 2017;96:e8412.
- [7] Ryu JS, Ryu HJ, Lee SN, Memon A, Lee SK, Nam HS, et al. Prognostic impact of minimal pleural effusion in non-small-cell lung cancer. *J Clin Oncol* 2014;32:960–7.
- [8] Clive AO, Kahan BC, Hooper CE, Bhatnagar R, Morley AJ, Zahan-Evans N, et al. Predicting survival in malignant pleural effusion: development and validation of the LENT prognostic score. *Thorax* 2014;69:1098–104.
- [9] Assawasaksakul T, Boonsarngsuk V, Incharoen P. A comparative study of conventional cytology and cell block method in the diagnosis of pleural effusion. *J Thorac Dis* 2017;9:3161–7.
- [10] Grosu HB, Kazzaz F, Vakili E, Molina S, Ost D. Sensitivity of initial thoracentesis for malignant pleural effusion stratified by tumor type in patients with strong evidence of metastatic disease. *Respiration* 2018;96:363–9.
- [11] Woo CG, Son SM, Han HS, Lee KH, Choe KH, An JY, et al. Diagnostic benefits of the combined use of liquid-based cytology, cell block, and carcinoembryonic antigen immunocytochemistry in malignant pleural effusion. *J Thorac Dis* 2018;10:4931–9.
- [12] Guldaval F, Anar C, Polat G, Gayaf M, Yavuz MY, Korkmaz A, et al. Contribution of cell block obtained by thoracentesis in the diagnosis of malignant pleural effusion. *J Cytol* 2019;36:205–8.
- [13] Hallifax RJ, Corcoran JP, Ahmed A, Nagendran M, Rostom H, Hassan N, et al. Physician-based ultrasound-guided biopsy for diagnosing pleural disease. *Chest* 2014;146:1001–6.
- [14] Ren Z, Hu Y, Xu L. Identifying tuberculous pleural effusion using artificial intelligence machine learning algorithms. *Respir. Res.* 2019;20(1):220.
- [15] Hackner K, Errhalt P, Handzhiev S. Ratio of carcinoembryonic antigen in pleural fluid and serum for the diagnosis of malignant pleural effusion. *Ther Adv Med Oncol* 2019;11:386246538.
- [16] Lee J, Yoo SS, Lee SY, Cha SI, Park JY, Kim CH. Pleural fluid adenosine deaminase/serum C-reactive protein ratio for the differentiation of tuberculous and paraneumonic effusions with neutrophilic predominance and high adenosine deaminase levels. *Infection* 2017;45:59–65.
- [17] Gonzalez A, Fielli M, Ceccato A, Luna C. Score for differentiating pleural tuberculosis from malignant effusion. *Med Sci (Basel)* 2019;7(3):36.

- [18] Valdes L, San-Jose E, Ferreiro L, Golpe A, Gonzalez-Barcala FJ, Toubes ME, et al. Predicting malignant and tuberculous pleural effusions through demographics and pleural fluid analysis of patients. *Clin Respir J* 2015;9:203–13.
- [19] Muruganandan S, Azzopardi M, Fitzgerald DB, Shrestha R, Kwan B, Lam D, et al. Aggressive versus symptom-guided drainage of malignant pleural effusion via indwelling pleural catheters (AMPLE-2): an open-label randomised trial. *Lancet Respir Med* 2018;6:671–80.
- [20] Guo M, Wu F, Hu G, Chen L, Xu J, Xu P, et al. Autologous tumor cell-derived micro-particle-based targeted chemotherapy in lung cancer patients with malignant pleural effusion. *Sci Transl Med* 2019;11.
- [21] Bielsa S, Salud A, Martinez M, Esquerda A, Martin A, Rodriguez-Panadero F, et al. Prognostic significance of pleural fluid data in patients with malignant effusion. *Eur J Intern Med* 2008;19:334–9.
- [22] Roberts ME, Neville E, Berrisford RG, Antunes G, Ali NJ. Management of a malignant pleural effusion: british thoracic society pleural disease guideline 2010. *Thorax* 2010;65(Suppl 2):i32–40.
- [23] Yang M, Tong Z, Wang Z, Zhang Y, Xu L, Wang X, et al. Development and validation of the PET-CT score for diagnosis of malignant pleural effusion. *Eur J Nucl Med Mol Imaging* 2019;46:1457–67.
- [24] Porcel JM, Pardina M, Bielsa S, González A, Light RW. Derivation and validation of a CT scan scoring system for discriminating malignant from benign pleural effusions. *Chest* 2015;147:513–9.
- [25] Liu K, Jin S, Song Z, Jiang L. High accuracy detection of malignant pleural effusion based on label-free surface-enhanced Raman spectroscopy and multivariate statistical analysis. *Spectrochim Acta A Mol Biomol Spectrosc* 2019;226:117632.
- [26] Yang Z, Gao Y, Fan X, Zhao X, Zhu S, Guo M, et al. A multivariate prediction model for high malignancy potential gastric GI stromal tumors before endoscopic resection. *Gastrointest Endosc* 2019.
- [27] Tian S, Cheng SB, Guo YY, Xie M, Zhan N, Zeng Z, et al. High efficient isolation of tumor cells by a three dimensional scaffold chip for diagnosis of malignant effusions. *ACS Applied Bio Materials* 2020;3:2177–84.
- [28] Wu S, Zheng J, Li Y, Wu Z, Shi S, Huang M, et al. Development and validation of an MRI-based radiomics signature for the preoperative prediction of lymph node metastasis in bladder cancer. *EBioMedicine* 2018;34:76–84.
- [29] He Y, Zhu Z, Chen Y, Chen F, Wang Y, Ouyang C, et al. Development and validation of a novel diagnostic nomogram to differentiate between intestinal tuberculosis and crohn's disease: a 6-year prospective multicenter study. *Am J Gastroenterol* 2019;114:490–9.
- [30] Zhang J, Zhao X, Zhao Y, Zhang J, Zhang Z, Wang J, et al. Value of pre-therapy (18) F-FDG PET/CT radiomics in predicting EGFR mutation status in patients with non-small cell lung cancer. *Eur J Nucl Med Mol Imaging* 2019.
- [31] Akturk UA, Ernam D, Akbay MO, Kocak ND, Ogur E, Irmak I. Role of the neutrophil-lymphocyte ratio in the differential diagnosis of exudative pleural effusion. *Clinics (Sao Paulo)* 2016;71:611–6.
- [32] Carter J, Miller JA, Feller-Kopman D, Ettinger D, Sidransky D, Maleki Z. Molecular profiling of malignant pleural effusion in metastatic non-small-cell lung carcinoma. the effect of preanalytical factors. *Ann Am Thorac Soc* 2017;14(7):1169–76.
- [33] Rodriguez EF, Shabihkhani M, Carter J, Maleki Z. Molecular alterations in patients with pulmonary adenocarcinoma presenting with malignant pleural effusion at the first diagnosis. *Acta Cytol* 2017;61(3):214–22.