

Development, assessment and validation of a novel prediction nomogram model for risk identification of tracheobronchial tuberculosis in patients with pulmonary tuberculosis

Qian Qiu ,¹ Siju Li,² Yong Chen,³ Xiaofeng Yan,⁴ Song Yang,⁴ Shi Qiu,⁵ Anzhou Peng,⁴ Yaokai Chen⁶

To cite: Qiu Q, Li S, Chen Y, et al. Development, assessment and validation of a novel prediction nomogram model for risk identification of tracheobronchial tuberculosis in patients with pulmonary tuberculosis. *BMJ Open Respir Res* 2023;**10**:e001781. doi:10.1136/bmjresp-2023-001781

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjresp-2023-001781>).

QQ and SL contributed equally.

Received 21 April 2023
Accepted 20 October 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Yaokai Chen;
yaokaichen@hotmail.com and

Dr Anzhou Peng;
penganzhou2010@163.com

ABSTRACT

Objective Tracheobronchial tuberculosis (TBTB), a specific subtype of pulmonary tuberculosis (PTB), can lead to bronchial stenosis or bronchial occlusion if not identified early. However, there is currently no available means for predicting the risk of associated TBTB in PTB patients. The objective of this study was to establish a risk prediction nomogram model for estimating the associated TBTB risk in every PTB patient.

Methods A retrospective cohort study was conducted with 2153 PTB patients. Optimised characteristics were selected using least absolute shrinkage and selection operator regression. Multivariate logistic regression was applied to build a predictive nomogram model. Discrimination, calibration and clinical usefulness of the prediction model were assessed using C-statistics, receiver operator characteristic curves, calibration plots and decision analysis. The developed model was validated both internally and externally.

Results Among all PTB patients who underwent bronchoscopies (n=2153), 40.36% (n=869) were diagnosed with TBTB. A nomogram model incorporating 11 predictors was developed and displayed good discrimination with a C-statistics of 0.782, a sensitivity of 0.661 and a specificity of 0.762 and good calibration with a calibration-in-the-large of 0.052 and a calibration slope of 0.957. Model's discrimination was favourable in both internal (C-statistics, 0.782) and external (C-statistics, 0.806) validation. External validation showed satisfactory accuracy (sensitivity, 0.690; specificity, 0.804) in independent cohort. Decision curve analysis showed that the model was clinically useful when intervention was decided on at the exacerbation possibility threshold of 2.3%–99.2%. A clinical impact curve demonstrated that our model predicted high-risk estimates and true positives.

Conclusion We developed a novel and convenient risk prediction nomogram model that enhances the risk assessment of associated TBTB in PTB patients. This nomogram can help identify high-risk PTB patients who may benefit from early bronchoscopy and aggressive treatment to prevent disease progression.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Tracheobronchial tuberculosis (TBTB) is often associated with pulmonary TB (PTB) disease and its late stages may result in bronchial stenosis or even bronchial occlusion. Early diagnosis and treatment are critical in preventing permanent lung damage. Yet, there are no means available for predicting associated TBTB risk in patients with PTB based on electronic clinical healthcare records, rather only on associated health conditions and epidemiological information.

WHAT THIS STUDY ADDS

⇒ We developed and validated a novel and simple nomogram based on retrospective, descriptive and non-interventional cohort investigation, incorporating 11 readily obtainable clinical parameters to aid in calculating the risk of associated TBTB in PTB patients. The nomogram was found to have favourable accuracy, reasonable discriminative ability and easy accessibility.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The prevalence of TB-induced tracheobronchial stenosis varies as a function of the prevalence of TB. However, due to a high degree of bronchostenosis beyond the initial period of TBTB, enhancing risk assessment of associated TBTB in PTB patients using an accessible quantitative tool, especially in low/middle-income countries, is necessary.

BACKGROUND

Despite the implementation of improved tuberculosis (TB) control programmes and strategies, approximately 10.6 million people worldwide contracted TB in 2021. Although this number has been declining slowly in recent years, this positive trend was reversed during the COVID-19 pandemic. Until the COVID-19 pandemic, TB was the leading



cause of death from a single infectious agent, ranking above HIV/AIDS.¹ TB is caused by the aerobic bacillus *Mycobacterium tuberculosis* (MTB) and is spread when patients expel bacteria into the air by, for example, coughing. The disease typically affects the lungs, causing pulmonary TB (PTB), but can also affect other sites, causing extrapulmonary TB (EPTB). Of the 6.4 million new cases recorded in 2021 and 5.3 million cases (83%) had PTB.² TB remains a major threat to public health and can lead to high rates of disability and mortality, which confers a heavy burden on both families and society at large.

Tracheobronchial TB (TBTB), which is characterised by a tuberculous infection of the trachea and/or bronchi, is a specific subtype of PTB and is closely associated with TB disease of the lungs.³ TBTB incidence ranges between 5.88% and 50% in PTB patients across different centres.⁴⁻⁶ Low patient adherence with chemotherapeutic regimens as well as increased bacterial drug resistance in patients with TBTB may result in bronchial stenosis or even bronchial occlusion, which may be caused by repetitively unhealed scar, which may result in partial or total pulmonary atelectasis, eventually destroying the corresponding lung.⁷ Incidence of stenosis may reach 68% within 6 months and is greater than 90% long term.⁸ This process may irreversibly impair lung physiology, resulting in respiratory failure and death. Early diagnosis and treatment of TBTB is, therefore, of uttermost importance to avoid permanent lung damage.

TBTB appears to have a preponderance in female in their second and third decades of life.^{3, 8} Cough is the most common symptom, followed by sputum production, weight loss, haemoptysis, chest pain and dyspnoea.^{3, 9} Clinical findings are heterogeneous and can include a focal wheeze and decreased air entry on auscultation.¹⁰ Because the signs and symptoms are non-specific, the diagnosis of TBTB should be made using a combination of a high index of clinical suspicion, clinical findings, radiology and sputum/tissue histopathological analysis. Thus, considering the demographic and clinical characteristics that tend to favour the occurrence of TBTB, it is vital to construct a comprehensive analytic model to accurately estimate the associated TBTB risk of every PTB patient. A predictive model is likely to help physicians perform targeted examinations in identified high-risk patients, and make expeditious and effective therapeutic decisions for these patients.

In recent years, the nomogram has been considered to be a viable and effective predictive tool for disease diagnostics and assessment of prognostic outcomes.^{11, 12} Several medical nomograms have been developed for the clinical diagnosis of TB and osteoarticular TB.^{13, 14} However, none has been developed for predicting associated TBTB risk in patients with PTB. In this study, we proposed to explore the clinical attributes of PTB patients with comorbid TBTB and assess possible predictors of TBTB to identify potential risk factors. In this study, we have established and validated a risk prediction

nomogram for TBTB, and assessed its calibration and discrimination to determine the model's validity. Finally, the model's prospective clinical value as a potential tool for reducing tracheobronchial complications of PTB and improving long-term outcomes in affected patients was assessed.

MATERIALS AND METHODS

Patients and study design

This was a non-interventional, retrospective cohort study. Confirmed PTB patients admitted to the Chongqing Public Health Medical Center from January 2018 to December 2019 were enrolled as our primary cohort. The diagnosis of PTB is definitively established by isolation of MTB from a bodily secretion or fluid (eg, culture of sputum, bronchoalveolar lavage or pleural fluid) or tissue (eg, pleural biopsy or lung biopsy).¹⁵ For external validation, PTB patients who were admitted to the same hospital between January 2020 and January 2021 were screened, using identical criteria. All patients underwent CT imaging and received bronchoscopic examination. A patient history of chronic diseases, TB medical history, hospitalisation history for TB, anti-TB treatment and TB-related complications were recorded and analysed.

Data collection

We collected data on epidemiological data including demographics (age, gender), diagnosis information, clinical presentation (symptoms and signs), disease progression (the duration of the illness, treatment process and treatment outcomes), laboratory tests (sputum culture, TB DNA testing, drug susceptibility testing, etc), radiological imaging (CT scans of the chest), treatment history, underlying chronic diseases, medical history and bronchoscopic examination during the hospital admission, from electronic medical records for inpatients by using data collection forms. Length of hospital stay was also recorded. A trained team of three physicians and researchers independently entered and cross-checked data in a computerised database. If any core data were missing, clarification was sought with the coordinators, who subsequently contacted clinicians responsible for the treatment of the patients. At the same time, medical histories and hospitalisation information were obtained by accessing paper-based medical records to enhance the completeness of information collection as much as possible.

Patient and public involvement

No patients were involved in the design or conduct of this study.

Patient identities and other private information were filtered and anonymised, and therefore, the requirement for informed consent was waived by the relevant authorities for this study. Access to study data was provided to the

authors by the Medical Affairs Administration Section of Chongqing Public Health Medical Center.

Diagnostic criteria for PTB and TBTB

PTB and TBTB were separately defined according to the listed criteria and 'Diagnosis and Treatment Guideline for Tracheobronchial Tuberculosis'.^{16,17} Briefly, TBTB was diagnosed based on visible lesions under bronchoscopy and either (1) positive acid-fast bacilli (AFB) in a sputum smear, brushing smear or bronchial alveolar lavage fluid, (2) positive MTB culture or (3) histopathological diagnosis of TB. Drug-resistant PTB patients, patients with a positive HIV test, patients with malignancy and predicted survival of less than 6 months, and patients with contraindications to bronchoscopy (such as severe hypoxaemia, respiratory failure, recent cardiovascular event or uncontrolled arrhythmia) were excluded.¹⁶ The exclusion criteria for drug-resistant TB in our study were based on the following criteria: (1) Patients with confirmed drug-resistant TB, including multidrug resistant TB and extensively drug-resistant TB, as determined by microbiological testing or clinical diagnosis, were excluded. (2) Patients with a history of previous treatment for drug-resistant TB or a documented history of resistance to first-line anti-TB drugs were excluded. (3) Patients with a documented history of non-adherence or failure of first-line anti-TB treatment were excluded.

Bronchoscopic examination

Bronchoscopy is a common diagnostic method used for the diagnosis and therapy of TB, especially as a relatively simple and effective method to directly observe the presence of airway luminal wall lesions. All clinical practices used in our study followed relevant guidelines for bronchoscopic examination and treatment for PTB.^{18–20} As an invasive procedure, bronchoscopy requires the skill of a well-trained pulmonologist and formal written consent from every patient before being performed. In the clinic, every patient who was examined via bronchoscopy was required to sign a disclaimer detailing the aims, methods, merits and risks of the procedure in detail. PTB patients underwent bronchoscopic examination based on the attending physician's medical advice and the patient's informed consent in the normal course of medical treatment. This study did not involve private patient information or any adjustment in clinical treatment. Also, the study guidelines did not permit the collection or assessment of any samples other than approved study data.

Based on the Chinese guidelines for classification of TBTB,¹⁷ bronchoscopic subtypes of TBTB are described as inflammatory infiltration, ulceration necrosis, granulation hyperplasia, scar stricture, tracheobronchomalacia and lymphatic fistula. Bronchial brushings or lavage were performed to identify AFB and transbronchial biopsies were conducted to confirm the pathological diagnosis of TBTB. The type and location of the lesion and bronchoscopic classification were recorded.

Statistical analysis

R (V.3.6.3) and GraphPad Prism (V.8.02) were both used for statistical analysis. R packages used in this study were 'epiDisplay', 'glmnet', 'Hmisc', 'rmda', 'rms', 'vcd', 'ggDCA', 'pROC' and 'ROCR'. Statistical significance levels were two sided. A $p < 0.05$ was considered statistically significant. In the demographic, clinical and laboratory data, continuous measurements that were normally distributed were expressed as the mean (\pm SD), and continuous measurements that did not have a normal distribution were expressed as the median (IQR), and were compared between two groups using an independent samples t-test or the Mann-Whitney U test. Categorical variables were presented as the amount (%), and compared with the χ^2 test or Fisher's exact test. The degree of association between two categorical variables was described using the phi coefficient, Cramer's V coefficient and Pearson's contingency coefficient.

Feature selection

Of the 44 clinical characteristics from 1894 patients in the entire cohort, 22 variables with limited association to TBTB incidence were excluded, such as symptoms: dizziness, headache, nasal congestion, hoarseness, sore throat, difficulty breathing, anorexia, nausea, vomiting, abdominal pain, diarrhoea, abdominal distension, body aches, weight loss, fatigue, etc, as well as various bacterial and molecular test results from sputum and BALF, bronchoscopy examination results (including affected sites, subtypes under microscopy, degree of tracheal or bronchial stenosis, extent of lesions and lesion types), and multiple treatment methods and outcomes under bronchoscopy. The remaining 22 variables include complete data for all patients. Second, variables of underlying diseases with an incidence rate of approximately less than 5% were excluded. Finally, 19 variables were included in the model analysis. The least absolute shrinkage and selection operator (LASSO) method was used to determine the optimal risk factor predictive features of TBTB patients.²¹ Features with non-zero coefficients using the LASSO regression model were selected. Additionally, part of the feature definitions was described in online supplemental appendix 1.

Prediction nomogram model development

A multivariable logistic regression analysis was performed to develop a predictive model, by integrating features selected from the LASSO regression model.²² The risk prediction model was developed with all potential predictors applied based on the multivariable logistic analysis. Using this model, the predicted patient exacerbation risk was calculated.²³ The risk score calculation formula was as follows: risk score = feature₁ × coef₁ + feature₂ × coef₂ + ... + feature_n × coef_n (coef: regression coefficients of features obtained from multivariable logistic analysis, n: total number of diagnostic-related features). Based on this formula, the risk score for each patient was



calculated. The receiver operator characteristic (ROC) curve was plotted and the Youden index (sensitivity+specificity-1) was calculated to determine the optimal cut-off value for the risk score. Using the cut-off value as the threshold, TBTB patients were classified into high-risk and low-risk groups.

Apparent performance of the nomogram model

To assess the model calibration, calibration curves were plotted, and the Hosmer-Lemeshow test was performed. Calibration-in-the-large was assessed by the regression curve intercept, while the calibration slope showing the degree of miscalibration was determined by the regression slope of the linear predictor.²⁴

ROC curves were computed to quantify the performance of the model concerning its discrimination, and the area under the curve (AUC, also known as C-statistics) was calculated. The performance of the model at different cut-off points was evaluated using various metrics, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy.

Validation of the nomogram model

Internal validation

Internal validation was achieved via the bootstrapping statistical technique (10 000 bootstrap resamples), to determine a relatively corrected C-statistics of the model-building process, including all previous potential predictors.

External validation

The externally validated nomogram model was assessed with an independent validation cohort. The logistic regression formula calculated from the primary cohort was applied to the external cohort, and total points for each patient were calculated. Logistic regression in the external cohort was then assessed using the total points as a factor. The ROC curve and C-statistics were then computed based on the regression analysis.

Clinical use

A decision curve analysis (DCA) and clinical impact curves were used to evaluate the clinical effectiveness of the model.²⁵ The x-axis of the DCA plot represents the threshold probability. In the risk assessment tool, the probability of patients being diagnosed with the disease is denoted as P_i . When P_i reaches a certain threshold (denoted as P_t), it is classified as positive, and treatment measures are taken. At this point, there will be benefits for patients who receive treatment but also harms for non-patients who receive treatment and costs for patients who are not treated. The y-axis represents the net benefit (NB), which is the difference between benefits and costs after treatment. The `plot_roc_components` function using 'rmda' R package plotted the probability

distribution of false positives and true positives for the ROC curve across a range of risk thresholds.

RESULTS

Diagnosis and incidence of TBTB

Of 7328 consecutive patients with TB hospitalised at Chongqing Public Health Medical Center during study period of the primary cohort, 2152 of 6190 PTB patients underwent bronchoscopy, and 1894 patients met inclusion criteria and were enrolled in this study. Among these 1894 patients, 753 were diagnosed with TBTB. The prevalence of PTB-associated TBTB in this population was 39.76% (753/1894), while it was calculated to be 44.79% (116/259) in the validation cohort. The overall prevalence of TBTB in PTB is 40.36% (869/2153). There was no statistically significant difference in TBTB prevalence between the two cohorts ($p=0.137$). Of the 753 patients diagnosed with TBTB in the primary cohort, 295 (39.2%) were sputum smear positive, 182 (24.2%) were MTB culture-positive, 403 (53.5%) were bronchial brush smear positive and 319 (42.4%) were bronchial brush culture positive.

Clinical characteristics of TBTB

In the comparison of baseline data between the primary and validation cohorts, we observed that there were minimal significant differences in most variables in online supplemental appendix 2. These included the prevalence of comorbid TBTB or EPTB, gender, age, smoking status, underlying diabetes and certain clinical symptoms and signs (such as cough, sputum, haemoptysis, fever and the presence of lung cavity). This lack of significant differences justified the utilisation of these cohorts as appropriate primary and validation sets for the model. However, statistically significant differences did emerge in other variables. These variables encompassed the presence of hypertension, other symptoms and signs (chest pain and chest discomfort, night sweats, tachypnoea and atelectasis), the duration of symptoms, results from pulmonary function tests, anti-TB treatment history and the number of previous hospitalisations.

PTB patient characteristics in the primary and validation cohorts were shown in [table 1](#). The primary cohort of TBTB patients (median age 33 years, range 16–77 years) consisted of 271 (35.99%, median age 39 years) males and 482 (64.01%, median age 31 years) females ([table 1](#)). Compared with PTB patients without TBTB, TBTB patients were more likely to be female, were more likely not to have EPTB, were more likely to be non-smokers, were more likely to have an anti-TB treatment history, and were more likely to have multiple previous hospitalisations (all $p<0.01$; [table 1](#)).

The common symptoms of TBTB included cough (88.6%), presence of sputum (47.4%), tachypnoea (19.7%), fever (12.6%), haemoptysis (11.4%), chest discomfort (8.4%) and chest pain (5.8%). Furthermore, the preceding first three symptoms accounted for the

Table 1 Demographic and clinical characteristics of patients in the primary and validation cohorts

Characteristics	Primary cohort			Validation cohort		
	PTB without TBTB	TBTB	P value	PTB without TBTB	TBTB	P value
n	1141	753		143	116	
EPTB, n (%)			<0.001			0.154
No	999 (87.6)	719 (95.5)		124 (86.7)	107 (92.2)	
Yes	142 (12.4)	34 (4.5)		19 (13.3)	9 (7.8)	
Gender, n (%)			<0.001			<0.001
Male	730 (64)	271 (36)		100 (69.9)	42 (36.2)	
Female	411 (36)	482 (64)		43 (30.1)	74 (63.8)	
Age, n (%)			0.207			0.488
≥60	125 (11)	77 (10.2)		16 (11.2)	15 (12.9)	
<45	716 (62.8)	502 (66.7)		79 (55.2)	70 (60.3)	
45–59	300 (26.3)	174 (23.1)		48 (33.6)	31 (26.7)	
Smoker, n (%)			<0.001			0.003
No	694 (60.8)	599 (79.5)		88 (61.5)	91 (78.4)	
Yes	447 (39.2)	154 (20.5)		55 (38.5)	25 (21.6)	
Underlying diseases, n (%)						
Diabetes	88 (7.7)	43 (5.7)	0.093	9 (6.3)	8 (6.9)	0.846
Hypertension	26 (2.3)	8 (1.1)	0.051	12 (8.4)	6 (5.2)	0.311
Anti-TB treatment history, n (%)			0.001			<0.001
0 times	449 (39.4)	242 (32.1)		85 (59.4)	44 (37.9)	
≥1times	692 (60.6)	511 (67.9)		58 (40.6)	72 (62.1)	
No of previous hospitalisations, n (%)			<0.001			<0.001
0	631 (55.3)	257 (34.1)		35 (24.5)	18 (15.5)	
1–2	497 (43.6)	421 (55.9)		99 (69.2)	71 (61.2)	
3–4	12 (1.1)	54 (7.2)		9 (6.3)	15 (12.9)	
≥5	1 (0.1)	21 (2.8)		0 (0)	12 (10.3)	
Symptoms, n (%)						
Cough	720 (63.1)	667 (88.6)	<0.001	78 (54.5)	103 (88.8)	<0.001
Sputum production	280 (24.5)	357 (47.4)	<0.001	35 (24.5)	56 (48.3)	<0.001
Tachypnoea	144 (12.6)	148 (19.7)	<0.001	25 (17.5)	28 (24.1)	0.187
Haemoptysis	177 (15.5)	86 (11.4)	0.012	15 (10.5)	12 (10.3)	0.97
Fever	183 (16)	95 (12.6)	0.039	23 (16.1)	13 (11.2)	0.259
Chest pain	112 (9.8)	44 (5.8)	0.002	25 (17.5)	11 (9.5)	0.064
Chest discomfort	83 (7.3)	63 (8.4)	0.383	20 (14)	16 (13.8)	0.964
Night sweat	52 (4.6)	27 (3.6)	0.301	15 (10.5)	7 (6)	0.201
Duration of symptoms, n (%)			0.306			0.639
≤4 weeks	323 (28.3)	197 (26.2)		49 (34.3)	43 (37.1)	
>4 weeks	818 (71.7)	556 (73.8)		94 (65.7)	73 (62.9)	
Pulmonary function test			0.699			0.02
Normal	336 (29.4)	228 (30.3)		80 (55.9)	48 (41.4)	
Abnormal	805 (70.6)	525 (69.7)		63 (44.1)	68 (58.6)	
Chest CT manifestations						
Lung cavity	290 (25.4)	178 (23.6)	0.38	45 (31.5)	18 (15.5)	0.003
Atelectasis	20 (1.8)	49 (6.5)	<0.001	7 (4.9)	10 (8.6)	0.229

EPTB, extrapulmonary TB; PTB, pulmonary TB; TB, tuberculosis; TBTB, tracheobronchial TB.

majority of symptoms in the TBTB patients, compared with patients without TBTB (all $p < 0.05$), whereas fever (12.6% vs 16%), haemoptysis (11.4% vs 15.5%) and chest pain (5.8% vs 9.8%) accounted for a much smaller proportion of symptoms.

According to chest CT results, 23.6% of TBTB patients had pulmonary cavitation, and only 6.5% had atelectasis; however, this accounted for a greater percentage of CT findings than in patients without TBTB. There were no mathematical differences in rates of underlying diseases (diabetes and hypertension) and results of pulmonary function testing between the two study populations.

Selection of features

Of the 44 clinical characteristics from 1894 patients in the entire cohort, 22 characteristics were chosen based on the criteria for variable inclusion. Three variables, that is, cardiovascular disease, hypertension and chronic pulmonary disease, were excluded because of the low proportion present (0.63%, 1.8% and 5.07%, respectively). To determine the critical but simple predictors used to assess the risk of TBTB accompanied by PTB, PTB, all 1894 patients were included for further screening using the LASSO method in the primary cohort. As a result, 17 potential predictors (figure 1A,B) that had non-zero coefficients in the LASSO regression model were selected out of the 19 features tested. The selected features included EPTB, gender, age, smoking, diabetes, cough, sputum, tachypnoea, haemoptysis, fever, chest pain, chest discomfort, pulmonary function test, lung cavity, atelectasis, anti-TB treatment history and the number of previous hospitalisations.

Multivariate analysis for TBTB

During multivariate logistic regression analysis, most selected features were identified as the related risk factors of patients with comorbid TBTB, except for age, haemoptysis, fever, chest discomfort, lung cavity and anti-TB treatment history (online supplemental appendix 3). No multicollinearity issues between different variables were encountered during our study computations. In this regard, our large sample size potentially mitigated the risk of multicollinearity problems. We, thus, concluded that 11 potential risk factors of TBTB harboured stable and meaningful estimation of β -coefficients in our study. There were no interaction terms among these variables. Compared with male patients, female PTB patients were associated with a higher risk of concomitant TBTB (OR 2.148, 95% CI 1.68 to 2.754). The greater the number of previous hospitalisations a PTB patient has, the greater the risk of secondarily acquired associated TBTB, and in particular, the risk of concomitant TBTB in patients with 5 or more previous hospitalisations for TB was significantly higher than those who had not been hospitalised for TB (OR 53.162, 95% CI 8.408 to 1120.401) (figure 2, online supplemental appendix 3).

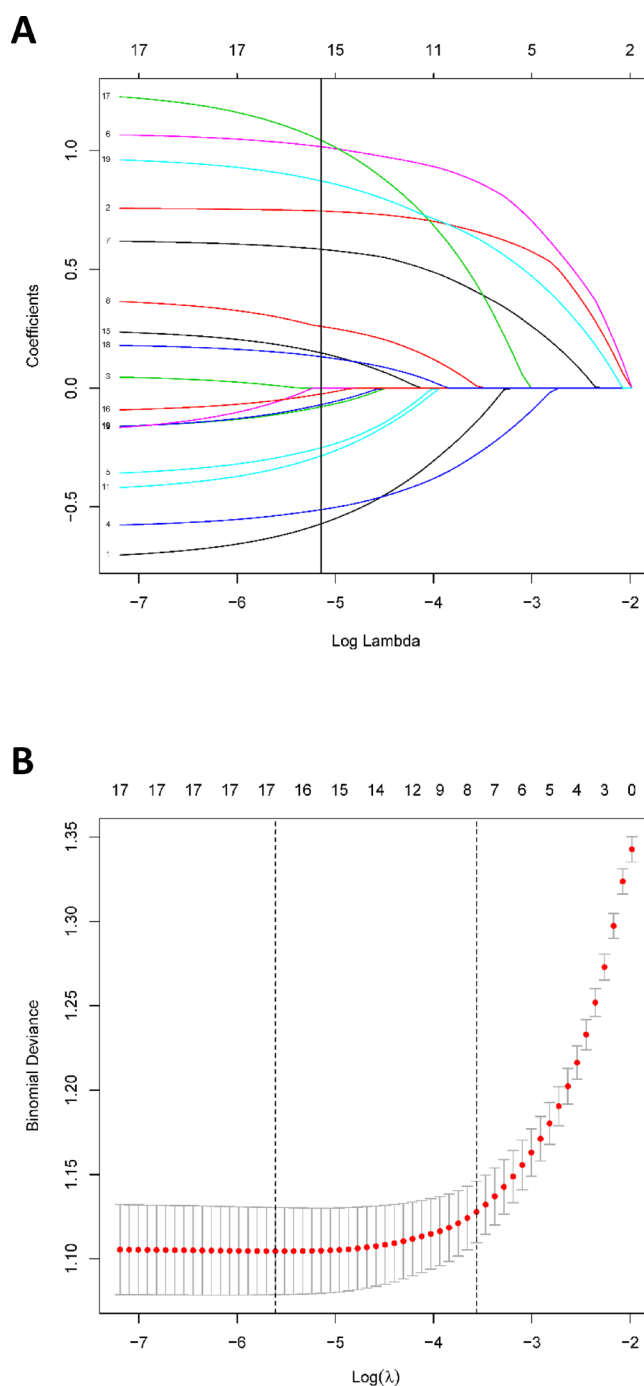


Figure 1 Selection of demographic, clinical and laboratory result features using the LASSO model. (A) Fivefold cross-validation with minimum criteria was used for optimal parameter (lambda) selection of the LASSO model. A partial likelihood binomial deviance curve is plotted against log (lambda). Vertical lines are drawn at the best values with 1-SE criteria. (B) LASSO coefficient profiles of all 40 characteristics were plotted against the log (lambda) sequence. Seven characteristics are shown with non-zero coefficients at the value chosen by fivefold cross-validation, marked by the vertical line. LASSO, least absolute shrinkage and selection operator.

Based on the 11 variables and their regression coefficients, we calculated the risk scores for each PTB sample. By plotting the ROC curve, we determined the optimal

Variable	N	Odds ratio (95% CI)	p
EPTB	No 1718	Reference	
	Yes 176	0.49 (0.31, 0.75)	0.002
Gender	Male 1001	Reference	
	Female 893	2.15 (1.68, 2.75)	<0.001
Smoker	No 1293	Reference	
	Yes 601	0.55 (0.42, 0.73)	<0.001
Diabetes	No 1763	Reference	
	Yes 131	0.66 (0.42, 1.01)	0.060
Cough	No 507	Reference	
	Yes 1387	2.90 (2.16, 3.92)	<0.001
Sputum production	No 1257	Reference	
	Yes 637	1.92 (1.51, 2.43)	<0.001
Tachypnea	No 1602	Reference	
	Yes 292	1.41 (1.06, 1.89)	0.019
Chest pain	No 1738	Reference	
	Yes 156	0.66 (0.43, 1.00)	0.052
Pulmonary function test	Normal 564	Reference	
	Abnormal 1330	1.26 (0.99, 1.61)	0.059
Atelectasis	No 1825	Reference	
	Yes 89	3.46 (1.93, 6.43)	<0.001
Number of previous hospitalizations	0 888	Reference	
	1-2 918	2.42 (1.93, 3.04)	<0.001
	3-4 66	8.92 (4.63, 18.48)	<0.001
	≥5 22	53.16 (8.41, 1120.40)	<0.001

Figure 2 Forest plot showing the results of multivariate analysis for associated TBTB in PTB patients. EPTB, extrapulmonary tuberculosis; PTB, pulmonary tuberculosis; TBTB, tracheobronchial tuberculosis.

cut-off value for risk scores to be -0.288 . Using this value, we divided the patients into a high-risk group (N=770) and a low-risk group (N=1124). Among them, 64.68% of patients in the high-risk group combined with TBTB, while 22.69% of patients in the low-risk group combined with TBTB. Notably, the proportion of PTB patients with comorbid TBTB was higher in the high-risk group than in the low-risk group ($\chi^2=336.37$, $p<0.0001$). The risk grouping of PTB patients showed a significant positive association with comorbid TBTB, with a moderate correlation (phi coefficient: 0.421; contingency coefficient: 0.388; Cramer's V: 0.421). These results indicate

that PTB patients in the high-risk group scores are more likely to combine with TBTB.

Development of the TBTB risk prediction model

A risk prediction nomogram model incorporating the aforementioned predictive variables was developed, with a ranked risk of between 0.1 and 0.9 (figure 3). Among all the included variables, not having EPTB had a risk score of 18, female gender=19, non-smoker=15, no diabetes=11, cough=27, sputum production=16, tachypnoea=9, no chest pain=10, abnormal pulmonary

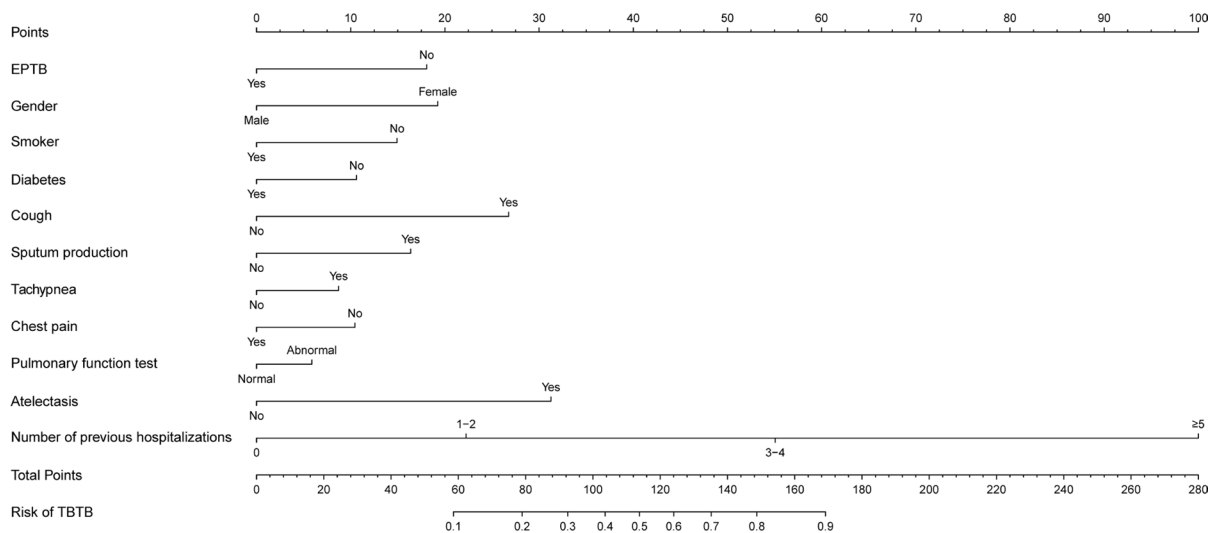


Figure 3 Risk prediction nomogram for associated TBTB in PTB patients. Points derived from the listed characteristics are added together to obtain 'total points', and the predicted risk of exacerbation is the corresponding value of 'risk of exacerbation'. EPTB, extrapulmonary tuberculosis; PTB, pulmonary tuberculosis; TBTB, tracheobronchial tuberculosis.

function test=6, atelectasis=31 and the number of previous hospitalisations for TB (1–2 times=22; 3–4 times=55; more than 5 times=100) (online supplemental appendix 4). A total score was obtained by calculating the sum of all the individual scores based on the patient's clinical risk variables. The calculated risk of TBTB is shown by the corresponding value on the 'total points' axis (figure 3).

According to this nomogram, the total scores calculated in this model were also categorised into three levels of risk based on their probability of associated TBTB. Those with total scores ranging from 0 to 59 had a less than 10% probability of exacerbation and were considered low risk, while those with scores between 79 and 104 had a 20%–40% probability of exacerbation, and were considered intermediate risk. Those with scores higher than 114 had a more than 50% probability of exacerbation and were considered high risk.

Apparent performance of the TBTB risk nomogram

The TBTB risk nomogram model was shown to have a good correlation between estimated and actual progression in the primary cohort. The calibration plots demonstrated a close-to-ideal calibration slope of 0.957 (95% CI 0.261 to 1.704), and an estimated calibration-in-the-large of 0.052 (95% CI -0.359 to 0.47). Thus, the model was well-calibrated for the primary cohort (figure 4A). The Hosmer-Lemeshow test was non-significant ($p=0.956$), demonstrating a perfect fit. To evaluate the accuracy of our model, an ROC analysis of the patients was performed. The obtained AUC, using a cut-off value of -0.288, was 0.782 (95% CI 0.761 to 0.803), with a sensitivity of 0.661 and a specificity of 0.762. The NPV was calculated to be 0.773, and the PPV was 0.647, with an overall accuracy of 0.722, demonstrating reasonable accuracy (figure 4B). These results demonstrated that the exacerbation

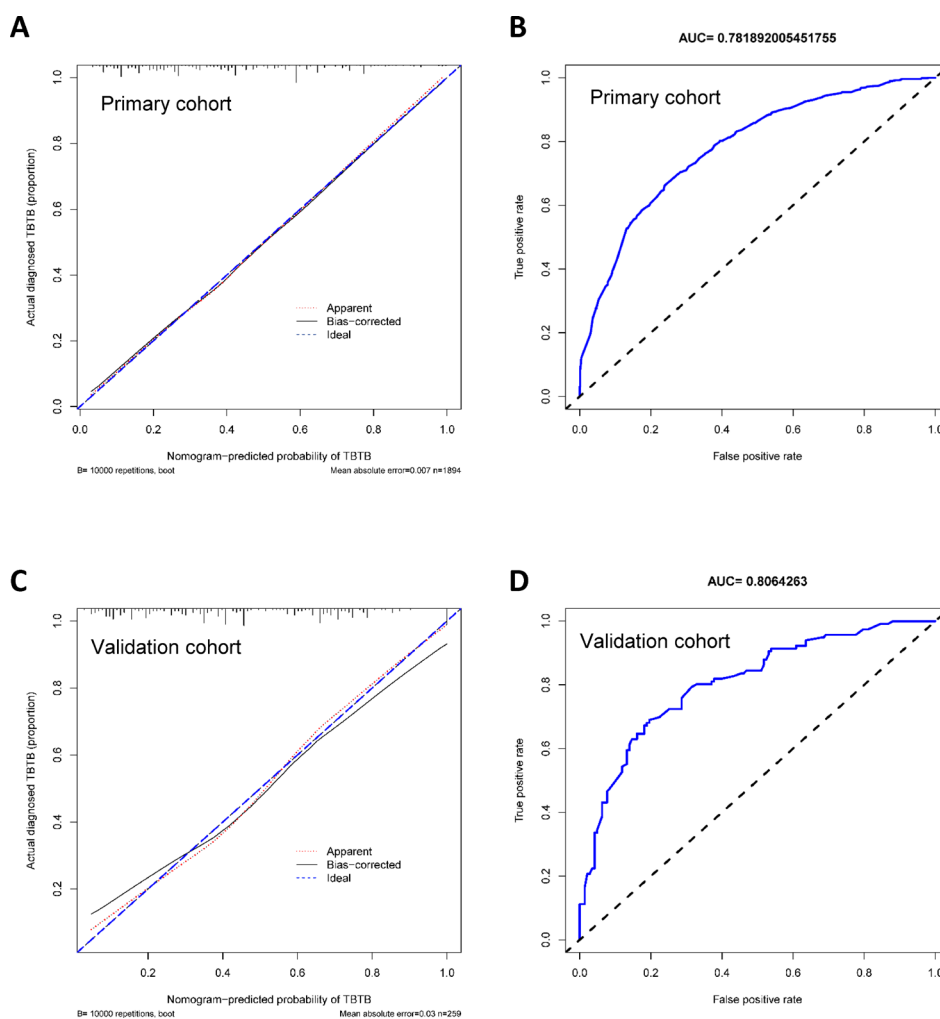


Figure 4 Apparent performance of risk prediction nomogram for associated TBTB in PTB patients. (A) Calibration curve of risk prediction nomogram. Predicted TBTB risk is shown on the x-axis, and diagnosed TBTB is shown on the y-axis. Perfect prediction by an ideal model is denoted by the diagonal dotted line, and the performance of the nomogram is denoted by the solid line, with better prediction shown by a closer fit. (B) The receiver operating characteristic (ROC) curve for the risk prediction nomogram predicting the 11 clinical features. (C) ROC curve of the validation cohort. AUC, area under the curve; PTB, pulmonary tuberculosis; TBTB, tracheobronchial tuberculosis.

nomogram had high prediction efficacy and reasonable discriminative ability in our primary cohort.

Validation of the risk prediction model

Internal validation

The model was subjected to bootstrapping statistical validation to further test its performance, yielding a relatively corrected C-statistics of 0.782 (95% CI 0.760 to 0.802), which indicates a good discriminative and predictive capability.

External validation

For all patients in the validation cohort, data for all 11 risk variables to be used in our nomogram model were available. Our model was also relatively well calibrated, and the Hosmer-Lemeshow test was non-significant ($p=0.732$) in the validation cohort (figure 4C). The ROC curve for the model, with an AUC of 0.806 (sensitivity, 0.690; specificity, 0.804) and an accuracy of 0.753, demonstrated good accuracy for the validation cohort (figure 4D).

Clinical use

The DCA results demonstrated that if the patient and doctor threshold probability was $>2.3\%$ and $<99.2\%$, respectively, using this model for TBTB risk prediction would provide more benefit than intervening for all patients or intervening for none. Within this wide range, the NB was comparable with several overlaps (figure 5A). For example, if we choose to predict the concomitant TBTB with a 40% threshold probability and treatment, then for every 100 patients using our model, 17 patients would benefit without harming anyone else. Furthermore, the clinical impact curve that was used to predict the risk stratification of 1000 associated TBTB patients visually displayed those who were presumed to be at high risk, and true positives within the high-risk threshold range of 3%–100% (figure 5B). For example, 456 out of 1000 patients would be deemed high risk if a 40% risk threshold was used, with about 280 of these being true severe or critical cases. Additionally, by plotting the components of the ROC curve, including the true positive rates and false positive rates, at different high-risk thresholds, we illustrated the relationships among the probability distribution of false and true positive rates, the risk threshold, and the cost–benefit ratio in both the primary and validation cohorts. These relationships were depicted in figure 5C,D. More details about the true positive rates and false positive rates by high-risk thresholds were shown in online supplemental appendix 5. By plotting the probability distribution of false positives and true positives in the ROC curve, clinicians can adjust the diagnostic results based on specific risk assessment thresholds to achieve more accurate patient classification and prediction. Selecting an appropriate threshold allows healthcare professionals to balance the trade-offs between false positives and true positives according to

specific clinical needs and treatment objectives, enabling them to make optimal diagnostic decisions.

DISCUSSION

In our study, we performed a comprehensive demographic and clinical characteristics-based disease profiling analysis in PTB patients with or without associated TBTB and developed a nomogram model for predicting concomitant TBTB. This nomogram contains 11 clinical attributes, including female gender, non-coexisting EPTB, non-smoker, non-coexisting diabetes, cough, sputum production, tachypnoea, no chest pain, abnormal pulmonary function test, atelectasis and multiple hospitalisations for TB, which performed well with reasonable discriminatory ability, and clinical usefulness in both the primary and validation cohorts.

Among our enrolled 2153 PTB patients, the number of the validation cohort collected from January 2020 to January 2021 decreased sharply (about 70%) from the primary cohort enrolled from January 2018 to December 2019. Given the impact of the COVID-19 outbreak on the continuity of hospital-based TB services in the survey of China,²⁶ the dramatic decline in enrolled TB hospitalisations in the validation cohort from January 2020 to January 2021 during the pandemic is almost certainly the result of changes in care seeking and access attributable to COVID-19, compared with the same period of 2019 in primary cohort. Most TB services, including diagnosis inpatient and outpatient care, decreased substantially during the COVID-19 emergency response phase (January to March 2020). There were three main drivers for these changes: (1) TB hospitals were temporarily converted to designated COVID-19 hospitals to handle the expected pandemic surge. For example, our hospital had been designated as a COVID-19 hospital in Chongqing and shifted at least some fraction of designated (range, 40%–100%) TB beds for COVID-19 care in 2020, and also dispatched professional TB staff for COVID-19 service (30%–100%). (2) TB hospitals reduced the number of consultations and hospitalisations to reduce the risk of nosocomial transmission of COVID-19. Our hospital set stricter indications for TB hospitalisation than had previously been used. Only those patients with severe TB, such as patients with haemoptysis, with massive pleural effusion, or those with drug-resistant TB, would be admitted to the hospital. (3) Consistent with WHO guidelines,²⁷ hospitals reduced the usage of bronchoscopy for TB patients. The rate of reduction reached a median of 67% (24%–100%) during the emergency response phase and by 27.5% (6%–90%) in the mitigation phase (April 2020) in 13 TB hospitals from thirteen provinces in different parts of mainland China.²⁶ Further, concerns about SARS-CoV-2 transmission in health facilities and on public transportation may have prevented individuals from seeking TB diagnosis or care.²⁶

In our comparative analysis of baseline data between the primary and validation cohorts, we observed that

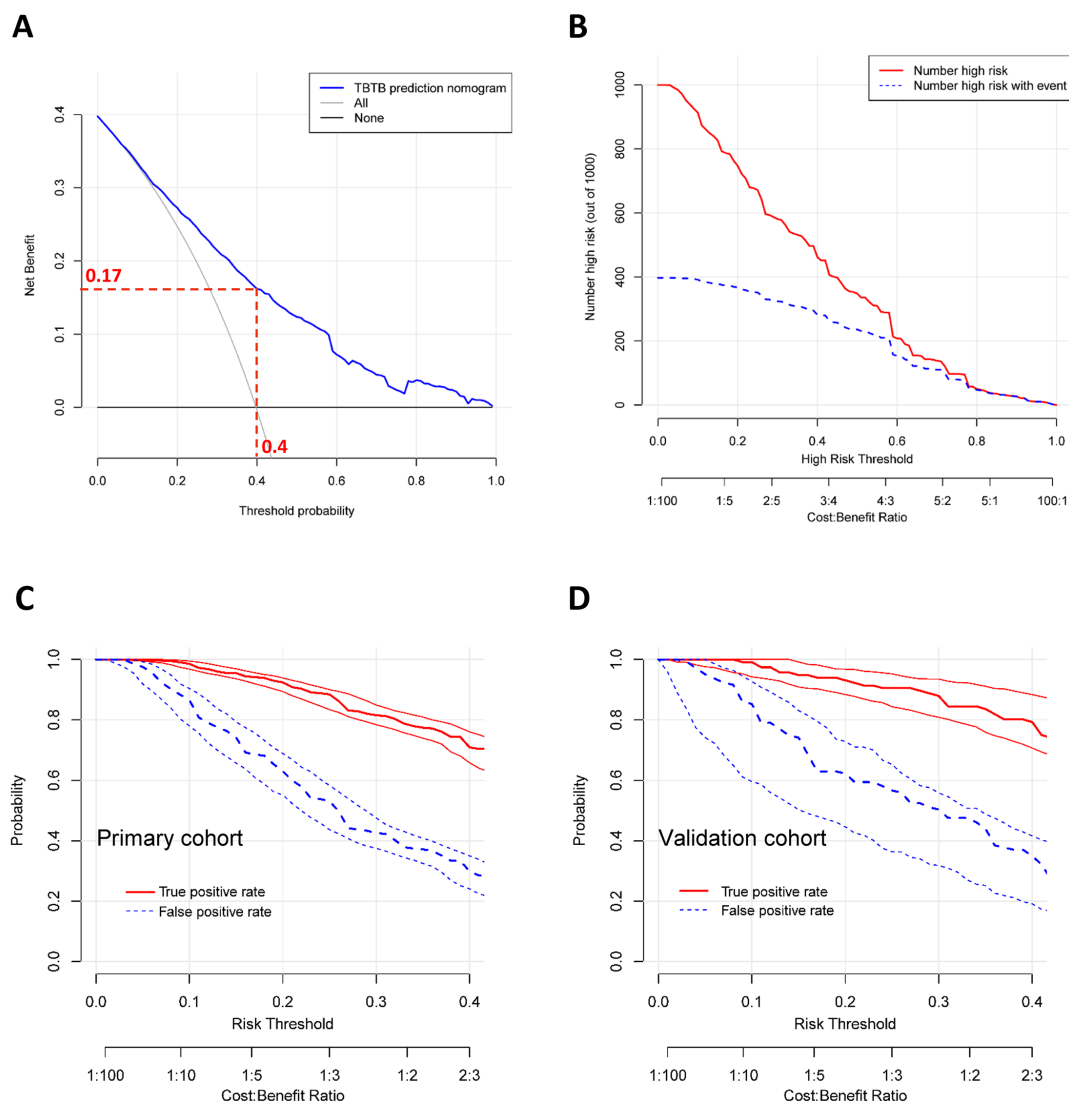


Figure 5 Decision curve analysis and clinical impact curve of the nomogram for associated TBTB in PTB patients. (A) Decision curve analysis. The thick blue line represents the prediction nomogram. The thin solid lines represent the assumption that all or no PTB patients progress to associated TBTB condition. The clinic net benefit was calculated for the risk prediction nomogram with risk threshold, and the red dashed lines represent a net benefit of 0.17 (x-axis) with a threshold probability of 0.4 (y-axis). (B) The clinical impact curve. The red solid line (number of high-risk patients) denotes how many of every 1000 patients would be deemed high risk for each risk threshold, and true positives are denoted by the blue dashed line. The correspondence between the high-risk threshold and cost–benefit ratio is represented by the two horizontal axes. (C) The probability distribution plot of false and true positive rates of ROC curve with the risk threshold and the cost–benefit ratio in the primary cohort. (D) The probability distribution plot of false and true positive rates of ROC curve with the risk threshold and the cost–benefit ratio in the validation cohort. PTB, pulmonary tuberculosis; ROC, receiver operator characteristic; TBTB, tracheobronchial tuberculosis.

there were no statistically significant differences in the incidence rates of concomitant TBTB or EPTB, gender, age, smoking status, the presence of diabetes, cough, sputum production, haemoptysis, fever and the presence of lung cavity. This absence of significant differences supports the use of these cohorts as suitable primary and validation cohorts for the model. Additionally, the decision to employ these datasets for model development is further reinforced by the lack of statistical disparities in these key variables between the two datasets. However, our analysis did reveal significant differences in several

variables when comparing baseline data between the primary and validation cohorts. Specifically, the validation cohort exhibited a higher prevalence of hypertension, the presence of chest pain, chest discomfort, night sweat or tachypnoea, the occurrence of atelectasis, a shorter duration of symptoms (≤ 4 weeks), normal pulmonary function test results, no prior anti-TB treatment history, and a greater number of previous hospitalisations (≥ 1) compared with the primary cohort. These baseline differences are likely influenced by the unique circumstances of the COVID-19 pandemic, the hospital's

stricter admission criteria for TB, and changes in testing protocols (eg, reduced pulmonary function tests). Nevertheless, in subsequent analyses when using the validation set, our model continued to demonstrate similarly strong performance as in the primary cohort. Thus, while these differences exist, the nomogram model remains valuable for predicting TBTB risk. However, healthcare providers should be aware of these variations when applying the model, considering the evolving clinical landscape and changing healthcare practices.

Bacteriological examination of sputum smear, that is, AFB staining and microscopy is a common TBTB test with low diagnostic yield,²⁸ which is consistent with findings in our study. In our primary cohort, 39.2% of patients were sputum smear-positive, and 24.2% of patients were sputum MTB culture-positive. The AFB smear and culture yield from BAL is known to be higher than that of sputum examination. Sputum analysis combined with specimens obtained via bronchoscopy in TBTB has a variable diagnostic yield of from 17% to 79%.^{8,29} In our study, 53.5% were bronchial brush smear positive, and 42.4% were bronchial brush culture positive. However, this relatively unimpressive diagnostic performance is far from sufficient for accurate TBTB diagnosis.

Several prediction models for predicting suspected PTB infection based on nosocomial populations have been published recently and may improve the diagnosis of PTB.^{30,31} In addition, Wang *et al*³² have created a CT-based predictive nomogram model to predict the risk of primary progressive PTB in children. A nomogram using body mass index, fasting blood glucose and triglyceride levels was used to identify high-risk patients that could form part of the target population for screening for diabetes mellitus.³³ Several TBTB risk factors have been reported in the past, such as female gender, relatively young age (in the second or third decade of life), clinical symptoms and atelectasis, among others.^{10,16,34} However, no specific TBTB risk prediction model exists. In our study, a nomogram for predicting TBTB was developed based on 11 readily accessible features of PTB patients. The variables included in our nomogram were filtered by LASSO regression analysis, considered superior for selecting potential predictors than univariate analysis.³⁵ Furthermore, we evaluated the clinical significance of these predictors. There is a consensus that TBTB appears to have a preponderance in females. It is hypothesised that females tend to expectorate less frequently than men due to social norms, thus leading to endobronchial stasis of secretions and potential susceptibility to subsequent MTB infection.³ Also, female bronchi are generally structurally narrower than those of men, and this has been considered to make females theoretically more susceptible to TBTB.^{16,28} It is known that frequent smokers are less likely to develop TBTB, as such, females are more likely to develop TBTB due to the lower prevalence of smoking.³⁶ Further longitudinal observational studies are required to extricate and define the specific risk differences between females and males in this respect.

The most common symptoms of TBTB are cough, sputum production and dyspnoea, which are presumed to be related to the prevailing endobronchial inflammation in the trachea and bronchi of patients with TBTB.³⁷ Chest pain is often a symptom when lesions develop or extend to the lungs, pleura, blood vessels and nerves of the pulmonary system, and only present in 15%–25% of TBTB patients,²⁸ thus not a good indicator of TBTB caused by bronchial or tracheal MTB. Our results confirmed that the absence of chest pain in a specific individual was likely to be associated with a relatively increased risk of TBTB. Pulmonary function tests tend to detect TBTB only after the disease progresses to the stage of tracheobronchial stenosis, with significant central airway narrowing. If a patient develops tracheobronchial stenosis, then, depending on the specific airway involved, the patient is very likely to progress to post-obstructive pneumonia or atelectasis in that specific region of the lung.³ Our results indicated that abnormal pulmonary function test results and atelectasis were also risk factors for predicting associated TBTB. Overall, the 11 risk variables used in our nomogram are readily available in a clinical setting. The nomogram showed reasonable discriminatory ability and good calibration, and DCA evaluation demonstrated its clinical usefulness. Since bronchoscopy is not yet fully or freely available in some TB-endemic areas, and many TB patients are sputum-free or have sputum-negative TB, our relatively cost-free nomogram is likely to be useful in screening for TBTB in these MTB-infected patients.

Using statistically derived risk factors, this study has established a prediction model for TBTB patients with relatively high precision. Both internal and external validation calculations showed favourable calibration power and reasonable discrimination, and internally and externally validated consistent C-statistics demonstrated that this model can be applied to a larger sample size with good expected accuracy. PTB patients with a calculated risk (via our nomogram) of TBTB of over 50% are very likely to have associated TBTB, and may, therefore, benefit from early bronchoscopy to detect endobronchial MTB involvement, and timely local therapy, which may include intratracheal instillation, aerosol therapy, surgery or bronchoscopic interventional procedures.^{38,39}

This study has several limitations. First, our model demonstrates moderate discrimination, with high specificity and low sensitivity for the early diagnosis of TBTB in PTB patients. Additional potential risk factors for concomitant TBTB need to be elucidated and added to our model, which could potentially increase the overall accuracy of TBTB screening. Second, as this was a retrospective population study, selection bias may be inherent in our study and findings. Third, the data are derived from an exclusively Southwest China population. There are known regional variations in the prevalence of TBTB; therefore, whether our nomogram applies to other regions or ethnic groups requires further verification in multicentre studies. Further large-scale, prospective non-randomised studies should be conducted to identify



accurate risk factors for TBTB in PTB patients. Additionally, another important limitation of our study is the potential selection bias associated with bronchoscopy, which may limit the generalisability of our findings to all PTB patients. The application of our nomogram to patients who did not undergo bronchoscopy requires further validation to assess its accuracy in this subpopulation.

CONCLUSION

This study developed and validated a novel and simple nomogram, incorporating 11 readily obtainable clinical features to aid in calculating the risk of associated TBTB in PTB patients with favourable accuracy, reasonable discriminative ability and easy accessibility. With the estimation of an individual's TBTB risk, necessary bronchoscopy can be expeditiously performed, appropriate diagnostic and therapeutic measures can be rapidly taken, risk of progression to tracheobronchial stenosis can be effectively reduced, and TBTB patient morbidity and mortality rates can be significantly improved.

Author affiliations

¹Post-Doctoral Research Center, Chongqing Public Health Medical Center, Chongqing, China

²Emergency Department, Chongqing Public Health Medical Center, Chongqing, China

³Department of Endocrinology, First Affiliated Hospital of Anhui Medical University, Hefei, China

⁴Division of Tuberculosis, Chongqing Public Health Medical Center, Chongqing, China

⁵Department of Nutrition, The Seventh Medical Center of Chinese PLA General Hospital, Beijing, China

⁶Division of Infectious Diseases, Chongqing Public Health Medical Center, Chongqing, China

Acknowledgements We would like to thank Jie Qiu of the Department of Engineering Physics, Tsinghua University, for providing programming assistance. We would also like to thank Dr. Vijay Harypursat of the Division of Infectious Diseases, Chongqing Public Health Medical Center, for English language revision, copy-editing and language enhancement to the English text of this manuscript.

Contributors QQ and AP: conceptualisation. QQ, AP, SL and YC: data collection. QQ: software. QQ and SQ: methodology. SQ: validation. QQ: statistical analysis. AP and SL: investigation. XF, YC, SY and MH: resources. AP: data curation. QQ and AP: writing-original draft preparation. QQ and YC: writing-review and editing. QQ: visualisation. XY and JX: supervision. YC and JX: project administration. QQ: funding acquisition, guarantor. All authors read and approved the final manuscript.

Funding This study was funded by grants from the Chongqing Natural Science Foundation, China (cstc2019jcyj-msxmX0028, cstc2021jcyj-msxmX0449), from the Joint Scientific Research Foundation by Chongqing Municipal Health Commission and Chongqing Municipal Bureau of Science & Technology, China (2018MSXM013, 2019QNXM038, 2022WSJK006) and the Youth Innovation Fund of Chongqing Public Health Medical Center, China (2019QNXYM01).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval All study procedures were approved by the Ethics Committee of Chongqing Public Health Medical Center (project decision number: 2020-064-02-KY).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. The majority of the data generated or analysed during this study are included in the published article. Data unavailable in the study may be obtained from the corresponding authors on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Qian Qiu <http://orcid.org/0000-0002-8281-6526>

REFERENCES

- 1 Bagcochi S. WHO's global tuberculosis report 2022. *Lancet Microbe* 2023;4.
- 2 World Health Organization. Global tuberculosis report 2022. Geneva: World Health Organization; 2022.
- 3 Pathak V, Shepherd RW, Shojaaee S. Tracheobronchial tuberculosis. *J Thorac Dis* 2016;8:3818–25.
- 4 Chung HS, Lee JH. Bronchoscopic assessment of the evolution of endobronchial tuberculosis. *Chest* 2000;117:385–92.
- 5 Kurasawa T, Kuze F, Kawai M, et al. Diagnosis and management of endobronchial tuberculosis. *Intern Med* 1992;31:593–8.
- 6 Lee JY, Yi CA, Kim TS, et al. CT scan features as predictors of patient outcome after bronchial intervention in endobronchial TB. *Chest* 2010;138:380–5.
- 7 Li Z, Mao G, Gui Q, et al. Bronchoplasty for treating the whole lung atelectasis caused by endobronchial tuberculosis in main bronchus. *J Thorac Dis* 2018;10:4000–5.
- 8 Kashyap S, Solanki A. Challenges in endobronchial tuberculosis: from diagnosis to management. *Pulm Med* 2014;2014:594806.
- 9 Sahin F, Yildiz P. Characteristics of endobronchial tuberculosis patients with negative sputum acid-fast bacillus. *J Thorac Dis* 2013;5:764–70.
- 10 Lee JH, Park SS, Lee DH, et al. Endobronchial tuberculosis. clinical and bronchoscopic features in 121 cases. *Chest* 1992;102:990–4.
- 11 Kawai K, Sunami E, Yamaguchi H, et al. Nomograms for colorectal cancer: A systematic review. *World J Gastroenterol* 2015;21:11877–86.
- 12 Wang S, Yang L, Ci B, et al. Development and validation of a nomogram prognostic model for SCLC patients. *J Thorac Oncol* 2018;13:1338–48.
- 13 Liang T, Chen J, Xu G, et al. Immune status changing helps diagnose osteoarticular tuberculosis. *PLoS One* 2021;16:e0252875.
- 14 Gao S-H, Chen C-G, Zhuang C-B, et al. Integrating serum microRNAs and electronic health records improved the diagnosis of tuberculosis. *J Clin Lab Anal* 2021;35:e23871.
- 15 Pai M, Nicol MP, Boehme CC. Tuberculosis diagnostics: state of the art and future directions. *Microbiol Spectr* 2016;4.
- 16 Jung S-S, Park H-S, Kim J-O, et al. Incidence and clinical predictors of endobronchial tuberculosis in patients with pulmonary tuberculosis. *Respirology* 2015;20:488–95.
- 17 Association CM. Diagnosis and treatment guideline for tracheobronchial tuberculosis. *Chin J Tuberc Respir Dis* 2012;35:581–7.
- 18 Xue Q, Wang N, Xue X, et al. Endobronchial tuberculosis: an overview. *Eur J Clin Microbiol Infect Dis* 2011;30:1039–44.
- 19 Zhou Y, Lin F, Cui Z, et al. Correlation between either Cupriavidus or Porphyromonas and primary pulmonary tuberculosis found by analysing the microbiota in patients' bronchoalveolar lavage fluid. *PLoS ONE* 2015;10:e0124194.
- 20 Su Z, Cheng Y, Wu Z, et al. Incidence and predictors of tracheobronchial tuberculosis in pulmonary tuberculosis: A multicentre, large-scale and prospective study in Southern China. *Respiration* 2019;97:153–9.
- 21 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33:1–22.
- 22 Kidd AC, McGettrick M, Tsim S, et al. Survival prediction in mesothelioma using a Scalable lasso regression model: instructions

- for use and initial performance using clinical predictors. *BMJ Open Respir Res* 2018;5:e000240.
- 23 Tan W, Xie X, Huang Z, *et al*. Construction of an immune-related genes Nomogram for the preoperative prediction of Axillary lymph node metastasis in triple-negative breast cancer. *Artif Cells Nanomed Biotechnol* 2020;48:288–97.
 - 24 Lee HJ, Lee A, Huang HH, *et al*. External validation of the updated Leibovich Prognostic models for clear cell and papillary renal cell carcinoma in an Asian population. *Urol Oncol* 2019;37:356.
 - 25 Hu ZG, Li WX, Ruan YS, *et al*. Incidence trends and risk prediction Nomogram of Metachronous second primary lung cancer in lung cancer survivors. *PLoS One* 2018;13:e0209002.
 - 26 Shen X, Sha W, Yang C, *et al*. Continuity of TB services during the COVID-19 pandemic in China. *Int J Tuberc Lung Dis* 2021;25:81–3.
 - 27 World Health Organization. *WHO guidelines on tuberculosis infection prevention and control: 2019 update*. Geneva: World Health Organization, 2019.
 - 28 Shahzad T, Irfan M. Endobronchial tuberculosis-a review. *J Thorac Dis* 2016;8:3797–802.
 - 29 Siow WT, Lee P. Tracheobronchial tuberculosis: a clinical review. *J Thorac Dis* 2017;9:E71–7.
 - 30 Rakoczy KS, Cohen SH, Nguyen HH. Derivation and validation of a clinical prediction score for isolation of Inpatients with suspected pulmonary tuberculosis. *Infect Control Hosp Epidemiol* 2008;29:927–32.
 - 31 Pegues CF, Johnson DC, Pegues DA, *et al*. Implementation and evaluation of an algorithm for isolation of patients with suspected pulmonary tuberculosis. *Infect Control Hosp Epidemiol* 1996;17:412–8.
 - 32 Wang B, Li M, Ma H, *et al*. Computed tomography-based predictive Nomogram for differentiating primary progressive pulmonary tuberculosis from community-acquired pneumonia in children. *BMC Med Imaging* 2019;19:63.
 - 33 Ji Y, Cao H, Liu Q, *et al*. Screening for pulmonary tuberculosis in high-risk groups of diabetic patients. *Int J Infect Dis* 2020;93:84–9.
 - 34 Peng A-Z, Yang A, Li S-J, *et al*. Incidence, laboratory diagnosis and predictors of Tracheobronchial tuberculosis in patients with pulmonary tuberculosis in Chongqing, China. *Exp Ther Med* 2020;20:174.
 - 35 Balachandran VP, Gonen M, Smith JJ, *et al*. Nomograms in oncology: more than meets the eye. *Lancet Oncol* 2015;16:e173–80.
 - 36 Rikimaru T, Kinoshita M, Yano H, *et al*. Diagnostic features and therapeutic outcome of erosive and ulcerous Endobronchial tuberculosis. *Int J Tuberc Lung Dis* 1998;2:558–62.
 - 37 Yanardag H, Tetikkurt C, Tetikkurt S, *et al*. Computed tomography and Bronchoscopy in Endobronchial tuberculosis. *Can Respir J* 2003;10:445–8.
 - 38 Lee P. Endobronchial tuberculosis. *Indian J Tuberc* 2015;62:7–12.
 - 39 Zhao L, Guan R, Yu H, *et al*. Endobronchial Anthracofibrosis: possibly a new indication of aggressive treatment for active Endobronchial tuberculosis. *J Thorac Dis* 2018;10:E511–5.

Appendix 1: Feature definitions

Features	Definitions
Cough	Cough is a reflex that keeps your throat and airways clear.
Sputum	Sputum is a thick type of mucus made in lungs. If you have an infection or chronic illness affecting the lungs or airways, it can make you cough up sputum.
Achypnea	A normal breathing rate for an adult at rest is 16 to 20 breaths per minute. Achypnea is the term that your health care provider uses to describe your breathing if it is too fast, especially if you have fast, shallow breathing from a lung disease or other medical cause.
Hemoptysis	Hemoptysis is the spitting up of blood or bloody mucus from the lungs and respiratory tract.
Fever	A fever is a body temperature that is higher than normal. A normal temperature can vary from person to person, but it is usually around 98.6 °F (37 °C).
Chest pain	Chest pain is pain that you feel anywhere along the front of your body between your neck and upper abdomen. Chest pain appears in many forms, ranging from a sharp stab to a dull ache. Sometimes chest pain feels crushing or burning.
Chest discomfort	Pressure, tightness, or other discomfort originating in or radiating to the chest.
Pulmonary function test	Pulmonary function tests (PFTs) are noninvasive tests that show how well the lungs are working. The tests measure lung volume, capacity, rates of flow, and gas exchange. Normal values for PFTs vary from person to person. The amount of air inhaled and exhaled in your test results are compared to the average for someone of the same age, height, sex, and race.
Lung cavity	Lung cavity is an abnormal, thick-walled, air-filled space within the lung. Diagnosis of a lung cavity is made with a chest X-ray or CT scan of the chest.
Atelectasis	Atelectasis is the collapse or closure of a lung resulting in reduced or absent gas exchange. It is usually unilateral, affecting part or all of one lung. Diagnosis of atelectasis is made with a chest X-ray or CT scan of the chest.

Appendix 2: Baseline characteristics of PTB patients in the primary and validation cohorts

Characteristics	Primary Cohort	Validation Cohort	p-value
n	1894	259	
TBTB, n (%)	753 (39.8)	116 (44.8)	0.122
EPTB, n (%)	176 (9.3)	28 (10.8)	0.434
Gender, n (%)			0.55
Male	1001 (52.9)	142 (54.8)	
Female	893 (47.1)	117 (45.2)	
Age, n (%)			0.097
≥ 60	202 (10.7)	31 (12)	
< 45	1218 (64.3)	149 (57.5)	
45-59	474 (25)	79 (30.5)	
Smoker, n (%)			0.784
No	1293 (68.3)	179 (69.1)	
Yes	601 (31.7)	80 (30.9)	
Underlying diseases, n (%)			
Diabetes	131 (6.9)	17 (6.6)	0.833
Hypertension	34 (1.8)	18 (6.9)	< 0.001
Anti-TB treatment history, n (%)			< 0.001
0 times	691 (36.5)	129 (49.8)	
≥ 1times	1203 (63.5)	130 (50.2)	
Number of previous hospitalizations, n (%)			< 0.001
0	888 (46.9)	53 (20.5)	
1-2	918 (48.5)	170 (65.6)	
3-4	66 (3.5)	24 (9.3)	
≥ 5	22 (1.2)	12 (4.6)	
Symptoms, n (%)			
Cough	1387 (73.2)	181 (69.9)	0.256
Sputum production	637 (33.6)	91 (35.1)	0.632
Tachypnea	292 (15.4)	53 (20.5)	0.038
Hemoptysis	263 (13.9)	27 (10.4)	0.126
Fever	278 (14.7)	36 (13.9)	0.739
Chest pain	156 (8.2)	36 (13.9)	0.003
Chest discomfort	146 (7.7)	36 (13.9)	< 0.001
Night sweat	79 (4.2)	22 (8.5)	0.002
Duration of symptoms, n (%)			0.007
≤ 4 weeks	520 (27.5)	92 (35.5)	
> 4 weeks	1374 (72.5)	167 (64.5)	
Pulmonary function test			< 0.001
Normal	564 (29.8)	128 (49.4)	
Abnormal	1330 (70.2)	131 (50.6)	
Chest CT manifestations			
Lung cavity	468 (24.7)	63 (24.3)	0.893
Atelectasis	69 (3.6)	17 (6.6)	0.024

Abbreviations: PTB, pulmonary tuberculosis; TBTB, tracheobronchial tuberculosis; EPTB, extrapulmonary tuberculosis.

Appendix 3: Prediction factors for associated TBTB in 1894 patients with PTB

Intercept and variable	Prediction model		
	β	Odds ratio (95% CI)	<i>p</i> -value
Intercept	-2.376	0.093 (0.062-0.137)	< 0.001
EPTB	-0.718	0.488 (0.308-0.754)	0.002
Female	0.765	2.148 (1.68-2.754)	< 0.001
Smoker	-0.593	0.553 (0.418-0.73)	< 0.001
Diabetes	-0.422	0.656 (0.419-1.012)	0.06
Cough	1.064	2.897 (2.157-3.919)	< 0.001
Sputum production	0.65	1.916 (1.511-2.435)	< 0.001
Tachypnea	0.346	1.414 (1.058-1.89)	0.019
Chest pain	-0.415	0.66 (0.43-0.998)	0.052
Pulmonary function test			
Normal		Reference	
Abnormal	0.234	1.263 (0.992-1.611)	0.059
Atelectasis	1.242	3.463 (1.925-6.427)	< 0.001
Number of previous hospitalizations			
0		Reference	
1-2	0.884	2.42 (1.929-3.045)	< 0.001
3-4	2.188	8.918 (4.631-18.478)	< 0.001
≥ 5	3.973	53.162 (8.408-1120.401)	< 0.001

β is the regression coefficient. Abbreviations: PTB, pulmonary tuberculosis; TBTB, tracheobronchial tuberculosis; EPTB, extrapulmonary tuberculosis; CI, confidence interval.

Appendix 4: Scores of clinical variables in each subgroup

Variable	Points	Variable	Points
EPTB		Tachypnea	
No	18	No	0
Yes	0	Yes	9
Gender		Chest pain	
Male	0	No	10
Female	19	Yes	0
Smoker		Pulmonary function test	
No	15	Normal	0
Yes	0	Abnormal	6
Diabetes		Atelectasis	
No	11	No	0
Yes	0	Yes	31
Cough		Number of previous hospitalizations	
No	0	0	0
Yes	27	1-2	22
Sputum production		3-4	55
No	0	≥ 5	100
Yes	16	Total scores	280

Abbreviations: EPTB, extrapulmonary tuberculosis.

