IASLC

Comprehensive Analysis of Clinical Logistic and Machine Learning-Based Models for the Evaluation of Pulmonary Nodules

Check for updates

Kai Zhang, MD,[a] Zihan Wei, MD,[a,b] Yuntao Nie, MD,[a] Haifeng Shen, MD,[a] Xin Wang, MD,[a,b] Jun Wang, MD,[a] Fan Yang, MD,[a] Kezhong Chen, MD[a,*]

[a]Department of Thoracic Surgery, Peking University People's Hospital, Beijing, People's Republic of China
[b]Peking University Health Science Center, Beijing, People's Republic of China

ABSTRACT

Introduction: Over the years, multiple models have been developed for the evaluation of pulmonary nodules (PNs). This study aimed to comprehensively investigate clinical models for estimating the malignancy probability in patients with PNs.

Methods: PubMed, EMBASE, Cochrane Library, and Web of Science were searched for studies reporting mathematical models for PN evaluation until March 2020. Eligible models were summarized, and network meta-analysis was performed on externally validated models (PROSPERO database CRD42020154731). The cut-off value of 40% was used to separate patients into high prevalence (HP) and low prevalence (LP), and a subgroup analysis was performed.

Results: A total of 23 original models were proposed in 42 included articles. Age and nodule size were most often used in the models, whereas results of positron emission tomography-computed tomography were used when collected. The Mayo model was validated in 28 studies. The area under the curve values of four most often used models (PKU, Brock, Mayo, VA) were 0.830, 0.785, 0.743, and 0.750, respectively. High-prevalence group (HP) models had better results in HP patients with a pooled sensitivity and specificity of 0.83 (95% confidence interval [CI]: 0.78–0.88) and 0.71 (95% CI: 0.71–0.79), whereas LP models only achieved pooled sensitivity and specificity of 0.70 (95% CI: 0.60–0.79) and 0.70 (95% CI: 0.62–0.77). For LP patients, the pooled sensitivity and specificity decreased from 0.68 (95% CI: 0.57–0.78) and 0.93 (95% CI: 0.87–0.97) to 0.57 (95% CI: 0.21–0.88) and 0.82 (95% CI: 0.65–0.92) when the model changed from LP to HP models. Compared with the clinical models, artificial intelligence-based models have promising preliminary results.

Conclusions: Mathematical models can facilitate the evaluation of lung nodules. Nevertheless, suitable model should be used on appropriate cohorts to achieve an accurate result.

Introduction

A pulmonary nodule (PN) is defined as an approximately round lesion surrounded by pulmonary parenchyma that is less than 3 cm in diameter. PNs have become increasingly common with the increased use of computed tomography (CT).[1–3] Although most nodules

---

are benign, a proportion of nodules are lung cancers, which is the leading cause of cancer-related death worldwide.[4] It is considered that the incidence of cancer in patients with solitary PNs ranges from 3.2% to 4.5%.[5,6] Therefore, the main goal for PN management is to identify patients with malignant nodules and administer proper treatment.

Current guidelines for the management of PNs recommend a systematic approach to PN assessment on the basis of clinical and radiographic characteristics.[7–9] The evaluation could be carried out either by experienced clinicians or by mathematical models developed to quantify the probability of malignancy of PNs. For patients with a high risk of malignant PNs, more aggressive interventions such as surgical intervention and CT biopsy are recommended, whereas serial high-resolution CT on a regular basis is recommended for PNs with a low risk of malignancy.[10]

Over the years, multiple models have been developed for the evaluation of PNs. Nevertheless, owing to various results and a lack of comparison, a consensus has not been made on the diagnostic value of these models. Moreover, with the development of deep learning, artificial intelligence (AI)-based models have been developed, and few articles have compared them with mathematical models. To perform a comprehensive analysis, we reviewed current clinical mathematical models that evaluate the probability of the malignancy of PNs and conducted a network analysis of the diagnostic accuracy of most often used models. We also summarized AI-based models that reported area under the curve[11] (AUC) values and compared them with those of the mathematical models.

Materials and Methods
Search Strategy
First, we searched the Medical Subject Headings term database of the National Center for Biotechnology Information for all possible expressions for "lung cancer" and proposed possible expressions for "prediction model." Then, we used the combination of the expressions to search the PubMed, EMBASE, Cochrane Library, and Web of Science databases up to March 30, 2020, without language limitations. The specific search strategy is listed as follows: ("Clinical Model" or "Clinical Prediction Model" or "Mathematical Model" or "Mathematical Prediction Model" or "Prediction Model" or "Gurney Model" or "Mayo Clinic Model" or "Herder Model" or "VA Model" or "PKU Model" or "Brock Model" or "TREAT Model" or "Bayesian Inference Malignancy Calculator" or "BIMC") and ("Pulmonary Neoplasms" or "Lung Neoplasm" or "Pulmonary Neoplasm" or "Lung Cancer" or "Pulmonary Cancer" or "Pulmonary Cancers"

or "Cancer of the Lung" or "Cancer of Lung" or "Pulmonary Nodule" or "Lung Nodule").

Titles and abstracts were used to identify papers related to prediction models for the cancer probability assessment of PNs. Full texts were then retrieved to extract data for calculation. This analysis was performed according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement.[12] The study design was registered in the PROSPERO database (CRD42020154731).

Selection Criteria
Reviews, case studies, editorials, meeting abstracts, and search results that were not related to any search criteria were excluded. All articles that proposed or validated prediction models for cancer probability assessment were included, and further screening was conducted after reading the full-text articles. The exclusion criteria for full-text screening were as follows: (1) the models were not built for predicting the cancer probability of lung nodules; (2) the models were not built with mathematical methods; (3) the models did not take clinical information into consideration; and (4) insufficient data for analysis.

Data Extraction
The following basic information was extracted: first author's name, year of publication, nation, size of the study, characteristics of the patients included in the study, prevalence of malignancy among the patients, average nodule size, models compared, and the result of comparisons (including the AUC, sensitivity, and specificity of each model compared). For papers that did not report the sensitivity and specificity of the models compared from publicized materials, we sent an e-mail to request the data. Then, we calculated the number of true positives, false negatives, false positives, and true negatives from the data acquired. The AUC values of other PN evaluation methods, including biomarkers, imaging, and physician assessments, were collected in the process for further analysis. Two authors (KZ and ZW) determined the study eligibility and extracted data independently, and any discrepancies between the two authors were resolved by discussion with a third author (KC).

Evaluation of AI-Based Models
Recently, AI-based models were reported to have a fairly good performance in PN evaluation.[13,14] Therefore, we also analyzed AI-based models in this study. The AUC values reported by AI-based models were collected, even when these articles were excluded from the major network meta-analysis. Although there has been no

article on March 30, 2020, that has compared mathematical models with AI-based models directly, the AUC values of AI-based models were summarized and the trend was analyzed.

Statistical Analysis

All models compared in the studies were included, and the variables used in the models were reviewed. The AUC values were compared by depicting a network plot. Most often used and externally validated models (Brock, Mayo, PKU, VA) were selected for a network meta-analysis; the summary receiver operating characteristic (SROC) curve was plotted with the method proposed by Reitsma et al.[15]; and the area under the SROC curve (AUSROC) was calculated. Sensitivity and specificity of each model were also pooled using analysis of variance model,[16] and diagnostic OR and superiority index were calculated.

In this article, we considered a model as most often used if it has been used in at least five independent cohorts. We noticed that the malignant rate of PNs in the included articles fell into the following two distributions: >40% and <25% (Supplementary Fig. 1); therefore, we used the cut-off value of 40% to separate patients into high prevalence (HP) and low prevalence (LP) and performed a subgroup analysis. Accordingly, models developed using HP nodules (malignant rate >40%) were defined as HP models and models using LP nodules (malignant rate <25%) were defined as LP models. During the analysis, when we encountered studies from the same medical center, we included the study using data from multiple hospitals to prevent duplicated patients.

All analyses were performed by R Software (R version 3.6.1 [2019-07-05], The R Foundation for Statistical Computing, with packages "mada" and "meta4diag").

Quality of Evidence

Quality Assessment of Diagnostic Accuracy Studies 2 is a tool designed by the Quality Assessment of Diagnostic Accuracy Studies 2 group for the evaluation of the quality of diagnostic accuracy studies.[17] The tool comprises the following four domains: patient selection, index test, reference standard, and flow and timing. The methodological quality of the eligible studies was evaluated by this tool by two reviewers (KZ and ZW; Supplementary Table 1).

Results

Our search resulted in 1816 articles, and after assessment, 42 articles were eligible for the study (Fig. 1A). Further searches through the reference list did not reveal additional relevant articles. The status of the data collection is summarized in Supplementary Figure 2.

The characteristics of all 42 articles were summarized in Table 1. A total of 23 original models were proposed by these 42 articles (Supplementary Table 2), among which 10 models were externally validated. Most articles used logistic regression to generate a new model, and a few used Bayesian analysis. The variables collected to propose a new model were summarized in Figure 1B, and the variables used in the final model are summarized in Figure 1C. Among the 23 articles that proposed a new model, 22 articles collected data on the age and 21 collected data on the nodule size of the patient, and most of them used these variables in the final model. Nevertheless, although sex was also most often collected, it was seldom used in the final model. Moreover, although there were only a few models that collected the level of uptake or maximum standardized uptake value from positron emission tomography (PET)-CT results, all of them used the result in the final model. The characteristics of the CT images of PNs, including lobulation, calcification, cavitation, border, and pleural retraction sign, were also widely used in the final models.

For further analysis, we included models that had been validated by at least two external sources for a descriptive analysis of AUC values and seven models qualified for the analysis. A detailed comparison is found in Figure 2A (AUC; Supplementary Table 3). Among the models that were compared more than 10 times, the PKU model had a better AUC value in 26 of 34 of the comparisons. Although only a few articles compared the BIMC model with the other models, the BIMC model yielded an excellent AUC value and wined in all the comparisons. Then, we selected the most often used models for a network meta-analysis. SROC curves were plotted for most often used models (used in at least five independent cohorts) (Fig. 2B), and the PKU model yielded the best AUC. The AUSROC values for the PKU, Brock, VA, and Mayo models were 0.830, 0.785, 0.750, and 0.743, respectively. Diagnostic OR and superior index were also calculated, which revealed similar tendencies (Fig. 2B). The pooled sensitivity and specificity of all externally validated models with diagnostic values provided are compared in Figure 2C, in which most often used models display better and more balanced performance in sensitivity and specificity.

Because the prevalence of lung cancer differs considerably in HP and LP patients, further analysis was performed after separating patients into the HP group and LP group. The PKU model remained the best model among the four most often used models for HP patients (AUSROC 0.826; Fig. 3A) and had a balanced performance in diagnostic values (Fig. 3B). For LP patients, the
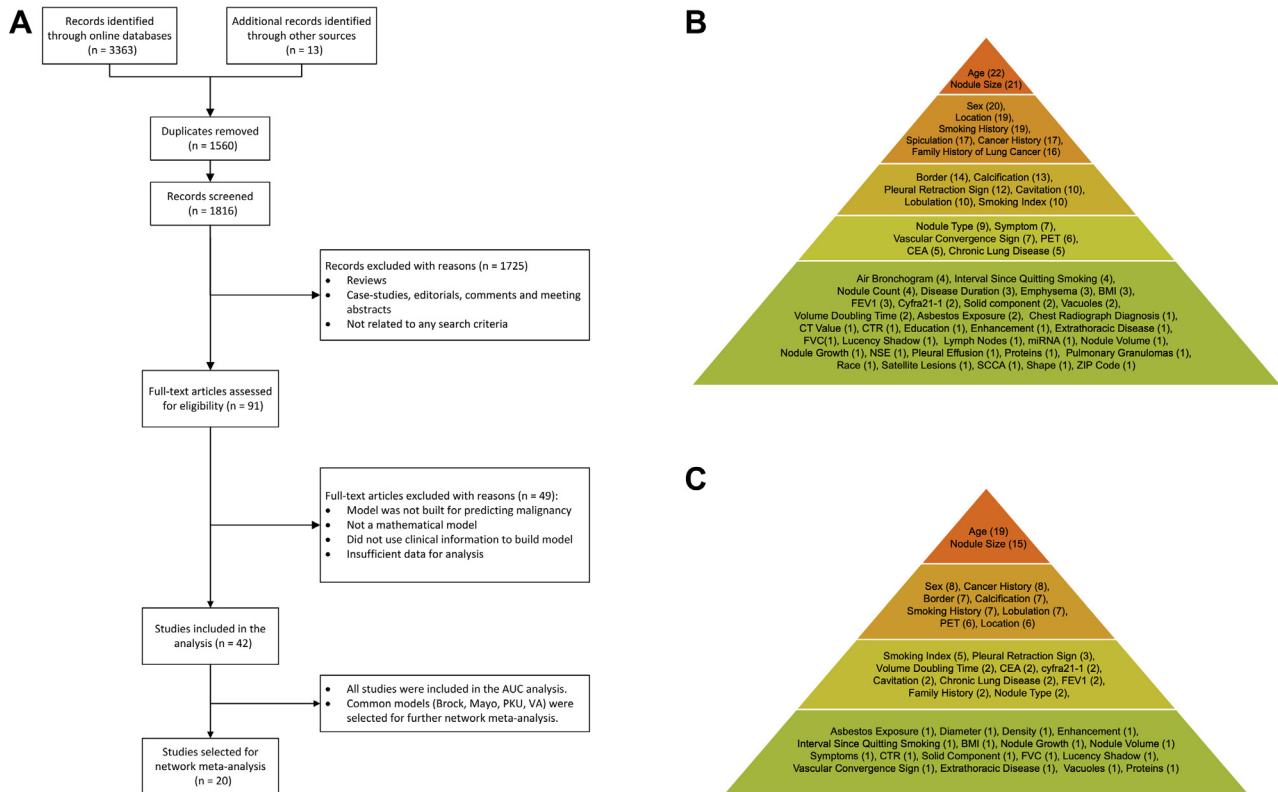
Figure 1. Process of study selection and summarization of all variables collected and used in eligible models. (*A*) PRISMA flow diagram of the study selection process. (*B*) All variables collected by eligible models. The variables are summarized in a pyramid chart and separated into five levels. Variables with a higher frequency occupy a higher level. The frequency is labeled after the variable names. (*C*) All variables used by the models. The variables are summarized in a pyramid chart and separated into four levels. Variables with a higher frequency occupy a higher level. The frequency is labeled after the variable names. BMI, body mass index; CEA, carcinoembryonic antigen; CT, computed tomography; CTR, consolidation/tumor ratio; FEV1, forced expiratory volume in the first second; FVC, forced vital capacity; miRNA, microRNA; NSE, neuron-specific enolase; PET, positron emission tomography; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; SCCA, squamous cell carcinoma antigen.

Mayo model yielded the best AUSROC (0.928; Fig. 3*C*) with a smaller confidence region of its summary point. Therefore, according to available data, the Mayo model is suitable for cancer risk prediction in LP patients, with applicable diagnostic value (Fig. 3*D*).

A subgroup analysis was conducted to investigate the effect of the prevalence of malignancy in different cohorts on the models. As found in Figure 4, HP models had a better result in predicting the cancer probability of PNs in HP patients, with a pooled sensitivity and specificity of 0.83 (95% confidence interval [CI]: 0.78–0.88; Fig. 4*A*) and 0.71 (95% CI: 0.63–0.79; Fig. 4*B*) compared with the LP models, which had a pooled sensitivity and specificity of 0.70 (95% CI: 0.60–0.79; Fig. 4*G*) and 0.70 (95% CI: 0.62–0.77; Fig. 4*H*). For LP patients, we observed that the pooled sensitivity and specificity decreased from 0.68 (95% CI: 0.57–0.78; Fig. 4*E*) and 0.93 (95% CI: 0.87–0.97; Fig. 4*F*) to 0.57 (95% CI: 0.21–0.88; Fig. 4*C*) and 0.82 (95% CI: 0.65–0.92; Fig. 4*D*) when the model was changed from LP models to HP

models. Overall, the pooled sensitivity and specificity of the HP models were 0.82 (95% CI: 0.77–0.87; Supplementary Fig. 3*A*) and 0.72 (95% CI: 0.65–0.79; Supplementary Fig. 3*B*), and the pooled sensitivity and specificity of the LP models were 0.70 (95% CI: 0.61–0.77; Supplementary Fig. 3*C*) and 0.76 (95% CI: 0.68–0.83; Supplementary Fig. 3*D*), respectively.

To explore the influence of PET-CT on the diagnostic performance of the models, we performed subgroup analysis on the PET-CT results. Models using PET-CT results as a variable had a high pooled sensitivity of 0.88 (95% CI: 0.77–0.95; Supplementary Fig. 4*A*) compared with 0.73 (95% CI: 0.68–0.77; Supplementary Fig. 4*C*) for models that did not use PET-CT. Nevertheless, the pooled specificity seemed to be lower in models with PET-CT results of 0.71 (95% CI: 0.49–0.89; Supplementary Fig. 4*B*) compared with 0.76 (95% CI: 0.71–0.80; Supplementary Fig. 4*D*) for models without PET-CT.

Among some of the included articles, the diagnostic value of the model was also compared with that of

Table 1. Characteristics of Eligible Studies

ID	Article	Country	Models	Study Population	Subgroup	Sample Size	Prevalence of Malignancy, %	Average Nodule Size (mm)
1	Gurney et al.[18,a]	United States	Gurney[b]	Pathologically confirmed SPNs	HP	66	67	B 15 M 29
2	Swensen et al.[19,a]	United States	Mayo[b]	Pathologically confirmed SPNs	LP	629	23	B 11.6 M 17.8
3	Herder et al.[20,a]	Netherland	Herder,[b] Mayo[b]	SPN without benign calcifications, referred for PET scan	HP	106	57	—
4	Gould et al.[21,a]	United States	VA[b]	PNs measured between 7 and 30 mm on CT	HP	375	54	17.03
5	Schultz et al.[22]	United States	Mayo,[b] VA[b]	SPNs confirmed by pathology or 2-y follow-up, age between 39 and 87 y	HP	151	44	15
6	Li et al.[23,a]	People's Republic of China	PKU,[b] Mayo,[b] VA[b]	Pathologically confirmed SPNs after surgery	HP	371	53	19.8
7	Tian et al.[24,a]	People's Republic of China	R Tian et al.	SPNs with PET result	HP	105	71	12.8
8	McWilliams et al.[25,a]	Canada	Brock[b]	PNs from current or former smokers, ages between 50 and 75 y	LP LP	7008 (PanCan) 5021 (BCCA)	1 1	4.3 3.7
9	Xiao et al.[26]	People's Republic of China	Mayo,[b] VA,[b] PKU[b]	Pathologically confirmed SPNs after surgery	HP	107	73	19.3
10	Deppen et al.[27,a]	United States	TREAT, Mayo[b]	Nodules from VUMC cohort and VA cohort	HP HP	492 (VUMC) 226 (VA)	72 93	28 29
11	Zhang et al.[28]	People's Republic of China	PKU,[b] Mayo,[b] VA[b]	Nodule count < 5, mGGO, and solid, no metastasis	HP	154	81	—
12	Al-Ameri et al.[29]	United Kingdom	Herder,[a] Mayo,[a] VA,[a] Brock[a]	PNs confirmed by pathology or 2-y follow-up, without pure GGO	HP	244	41	—
13	Vachani et al.[30,a]	United States	A Vachani et al.	PNs confirmed by pathology or 2-y follow-up, age >40 y	HP	141	55	13
14	Soardi et al.[31,a]	Italy	BIMC, Gurney[b]	SPNs with PET result, without calcification	HP	343	58	14.9
15	Yang et al.[32,a]	People's Republic of China	J Yang et al., PKU,[b] Mayo,[b] VA[b]	Pathologically confirmed SPNs after surgery	HP	252	67	17
16	Zhang et al.[33,a]	People's Republic of China	GMUFH,[a] Mayo,[b] VA,[b] Brock,[b] PKU[b]	Pathologically confirmed SPNs	HP	120	60	—
17	Chen et al. 2016[34,a]	People's Republic of China	J Chen et al., PKU,[b] Mayo,[b] VA[b]	Pathologically confirmed SPNs	HP HP	200 89 (Validation)	68 79	17.41 18.91
18	Perandini et al.[35]	Italy	Herder,[b] BIMC	SPNs with PET result, without calcification	HP	180	54	17.8
19	Perandini et al.[36]	Italy	Mayo,[b] Gurney,[b] PKU,[b] BIMC	Pathologically confirmed SPNs	HP	285	55	15.36
20	Soardi et al.[37]	Italy	BIMC, Mayo[b]	SPNs from three medical centers	HP	200	54	15.89

(continued)

Table 1. Continued

ID	Article	Country	Models	Study Population	Subgroup	Sample Size	Prevalence of Malignancy, %	Average Nodule Size (mm)
21	Chen et al.[38]	People's Republic of China	Mayo,[b] PKU[b]	Pathologically confirmed PNs after surgery	HP	41	76	—
22	Yang et al.[39,a]	People's Republic of China	Li Yang et al., VA,[b] Mayo[b]	SPN referred to CT-guided biopsy	HP HP	1078[4] 344 (Validation)	67 69	18.43 18.16
23	Tanner et al.[40]	United States	Mayo,[b] VA[b]	SPN with progression in 60 d, age > 40 y	HP	337	47	15.8
24	W Yu (2017)[41,a]	People's Republic of China	W Yu et al.	Pathologically confirmed GGO	HP HP	362[4] 206 (Validation)	67 70	1.6 1.5
25	Lin et al.[42]	People's Republic of China	Mayo[b]	PNs from current or former smokers, ages between 55 and 74 y	HP HP	135 (JPHTCM) 126 (BVAMC)	51 50	15.14 14.365
26	She et al.[43,a]	People's Republic of China	Y She et al., VA,[b] Mayo,[b] PKU,[b] Brock[b]	Pathologically confirmed solid SPNs after surgery	HP HP	899[4] 899 (Validation)	67 66	17.3 17.3
27	Yang et al.[44]	Korea	Mayo,[b] VA,[b] Brock,[b] Herder[b]	Nodule count < 5, mGGO, and solid, no metastasis	HP	242	77	20
28	Kim et al.[45]	Korea	Brock	Single subsolid nodules confirmed as AAH or AIS or MIA or IPA	HP HP	101 (GGO) 309 (mGGO)	58 91	B 11.1 M 14.2 B 13.6 M 17.6
29	Wang et al.[46,c]	People's Republic of China	ZU,[b] Mayo,[b] VA[b]	SPNs with PET result	HP	177	67	18.89
30	Nair et al.[47]	United States	Brock,[b] Mayo,[b] VA[b]	Nodules from NLST	LP LP	2196 (Set 1) 6568 (Set 2)	9 3	12.1 7.6
31	Ying et al.[48,c]	People's Republic of China	Ying et al., Mayo[b]	Pathologically confirmed microsized SPN (<10 mm)	HP HP	102[4] 10 (Validation)	76 60	— —
32	Winter et al.[49]	United States	A Winter et al., Brock[b]	Nodules from NLST	LP	7879	3	6.89
33	Xiao et al.[50,a]	People's Republic of China	CJFH, Mayo,[b] VA,[b] Brock,[b] PKU,[b] GMUFH[b]	Pathologically confirmed nonsolid SPNs after surgery	HP	362	87	17.6
34	Kim et al.[51,a]	Korea	H Kim et al., Brock[b]	Pathologically confirmed subsolid nodules after surgery	HP HP	321[4] 106 (Validation)	72 72	15.7 15.8
35	Uthoff et al.[52]	United States	Mayo,[b] VA,[b] Brock,[b] PKU[b]	SPNs, age between 40 and 87 y	LP	317	22	B 9.2 M 16.3
36	Xi et al.[53,a]	People's Republic of China	K Xi et al.	Pathologically confirmed SPNs	HP HP	40 52	70 75	B 19 M 25.1 B 14.0 M 18.3
37	Hammer et al.[54]	United States	Brock[b]	GGO and PSN from NLST	LP	434	6	—
38	Marcus et al.[55,a]	United Kingdom	UKLS[b]	Nodules from UKLS trial	LP	1013	5	—
39	Cui et al.[56,a]	People's Republic of China	Mayo,[b] Brock,[b] VA[b]	SPNs confirmed by pathology or 2-y follow-up	HP	277	73	17

Table 1. Continued

ID	Article	Country	Models	Study Population	Subgroup	Sample Size	Prevalence of Malignancy, %	Average Nodule Size (mm)
40	Guo et al.[57]	People's Republic of China	GLCI, Mayo,[b] PKU,[b] Herder,[b] ZU [b]	SPNs with PET result	HP HP	312[4] 159 (Validation)	69 80	18.6 —
41	González Maldonado et al.[11]	Germany	Brock,[b] UKLS,[b] Mayo,[b] VA,[b] PKU[b]	Nodules from LUSI trial	LP	3903	2	B 4.0 M 9.4
42	Li et al.[58]	People's Republic of China	Brock,[b] Mayo,[b] VA,[b] PKU[b]	Pathologically confirmed PNs after surgery	HP	496	86	—

[a]Models first established in the article.
[b]Externally validated model.
B stands for benign and M stands for malignant.

AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; BCCA, British Columbia Cancer Agency; BIMC, Bayesian inference malignancy calculator; BVAMC, Baltimore VA Medical Center; CJFH, China-Japan Friendship Hospital; CT, computed tomography; GGO, ground-glass opacity; GLCI, Guangdong Lung Cancer Institute; GMUFH, The First Affiliated Hospital Of Guangzhou Medical University; HP, high prevalence; ID, identification; IPA, invasive pulmonary adenocarcinoma; JPHTCM, Jiangsu Province Hospital of Traditional Chinese Medicine; LP, low prevalence; LUSI, German Lung Cancer Screening Intervention; mGGO, mixed ground-glass opacity; MIA, minimally invasive adenocarcinoma; NLST, National Lung Screening Trial; PanCan, Pan-Canadian Early Detection of Lung Cancer; PET, positron emission tomography; PKU, Peking University; PN, pulmonary nodule; PSN, part part-solid nodule; SPN, solidary pulmonary nodule; TREAT, thoracic research evaluation and treatment; UKLS, UK Lung Cancer Screening; VA, Department of Veterans Affairs; VUMC, Vanderbilt University Medical Center; ZU, Zhejiang University.

various biomarkers, imaging methods, and physician assessment. The AUC values of each method were collected and analyzed. The average AUC value of the models was higher than that of the other methods, although no significant difference was observed (Supplementary Fig. 5*A*).

When evaluating AI-based models, 11 articles were included, and five of them were developed using HP patients whereas others were developed using LP patients from screening projects. The AUC values of the AI-based models were compared with those of the models, biomarkers, imaging, and physicians (Supplementary Fig. 5*B*). In recent 5 years, the AUC of the AI-based models had raised from an average of 0.831 ($\pm$0.071) in 2017 to 0.919 in 2020, whereas the AUC of the mathematical models seems to bear a better robustness. Further regression of the AUC values of the AI-based models revealed that the AUC values of the AI-based models increased with the development of AI throughout the years ($p = 0.074$; Supplementary Fig. 5*C*), whereas mathematical models did not. Though the development of AI-based models seemed not statistically different, the trend can also be validated by studies in the same data set.[59] Nevertheless, this might indicate that the performance of well-trained AI models might exceed that of the current methods in PN evaluation in the future. External validation is still needed for the AI-based models.

Discussion

With the increasing use of CT in lung cancer screening, it has become increasingly considerable to estimate the cancer probability accurately during the management of PNs for both inpatients in the surgical department and outpatients who participate in CT screening. In view of this, we summarized all clinical mathematical models for the evaluation of PNs and conducted a network meta-analysis for the first time. To ensure objectivity and fairness, we contacted the authors of all published articles that lacked the desired data (Supplementary Fig. 2).

As the first probability model that used logistic regression, the Mayo model has become the most externally validated among all models (Table 2). It is built from a retrospective data set of 419 patients with more than 20 variables taken into consideration. Owing to the large number of variables collected, the Mayo model has remained a rather accurate model throughout the years. Among most often used models, the PKU model yields the best AUC. It is the first model built with the Chinese population and is the only eastern population model that has been validated with the western
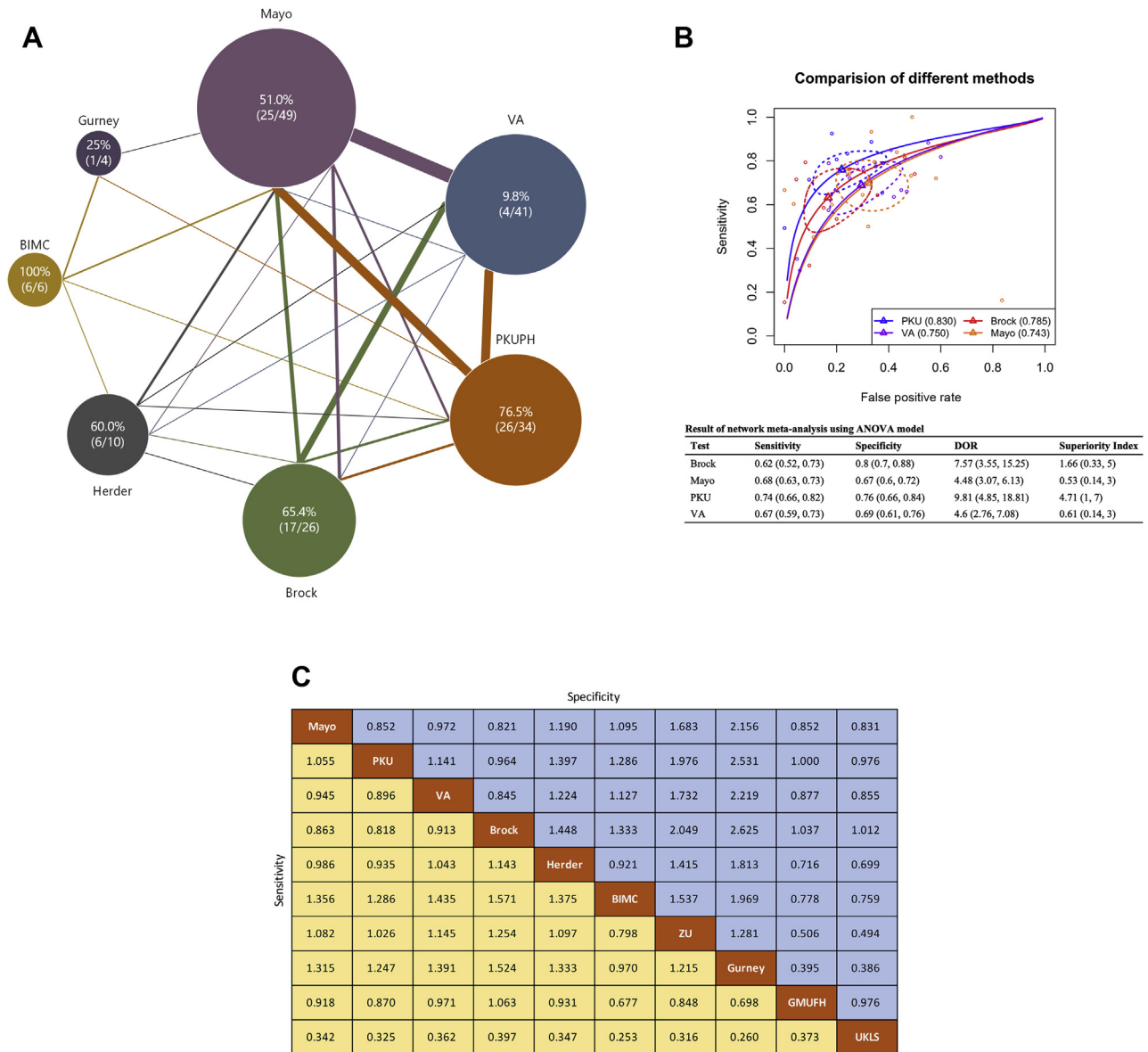
Figure 2. Comparisons of the AUC values, SROC curves, and diagnostic values among the models. (*A*) AUC comparison of seven models validated by at least two external sources. Each circular node represents a validated model. The area of the node is proportional to the total number of comparisons in eligible studies. The ratio of the times of better performance to the total number of comparisons is listed inside the node. Each line represents a type of head-to-head comparison, and the color of the line is identical to that of the winning model. The width of the lines is proportional to the number of head-to-head comparisons. (*B*) SROC curves of models with sufficient external validation (at least five independent cohorts). The solid line depicts the SROC curve plotted by the method proposed by Reitsma et al.,[10] and individual observations are marked with round points. The summary point is marked with a triangle point on the SROC curve, and its 95% confidence region is plotted with a dotted line. Different colors are assigned to each model. AUC values are listed in parentheses after the model names in the figure legend. Result of network meta-analysis using ANOVA model is listed below. (*C*) Comparison of the pooled sensitivity and specificity of the validated models. The value in each cell is defined as the pooled sensitivity or specificity of the model in the same row divided by the pooled sensitivity or specificity of the model in the same column. Cells with the model name are marked in orange, and cells containing the sensitivity and specificity values are marked in yellow and blue, respectively. ANOVA, analysis of variance; AUC, area under the curve; BIMC, Bayesian Inference Malignancy Calculator; DOR, duration of response; SROC, summary receiver operating characteristic.

population. Compared with the Mayo model, all patients enrolled in the PKU model had a defined pathologic diagnosis and comprehensive radiographical characteristics.

Owing to the variation in research cohort, models proposed in past studies can be separated into two categories. The first category is models on the basis of the population who underwent lung cancer screening.
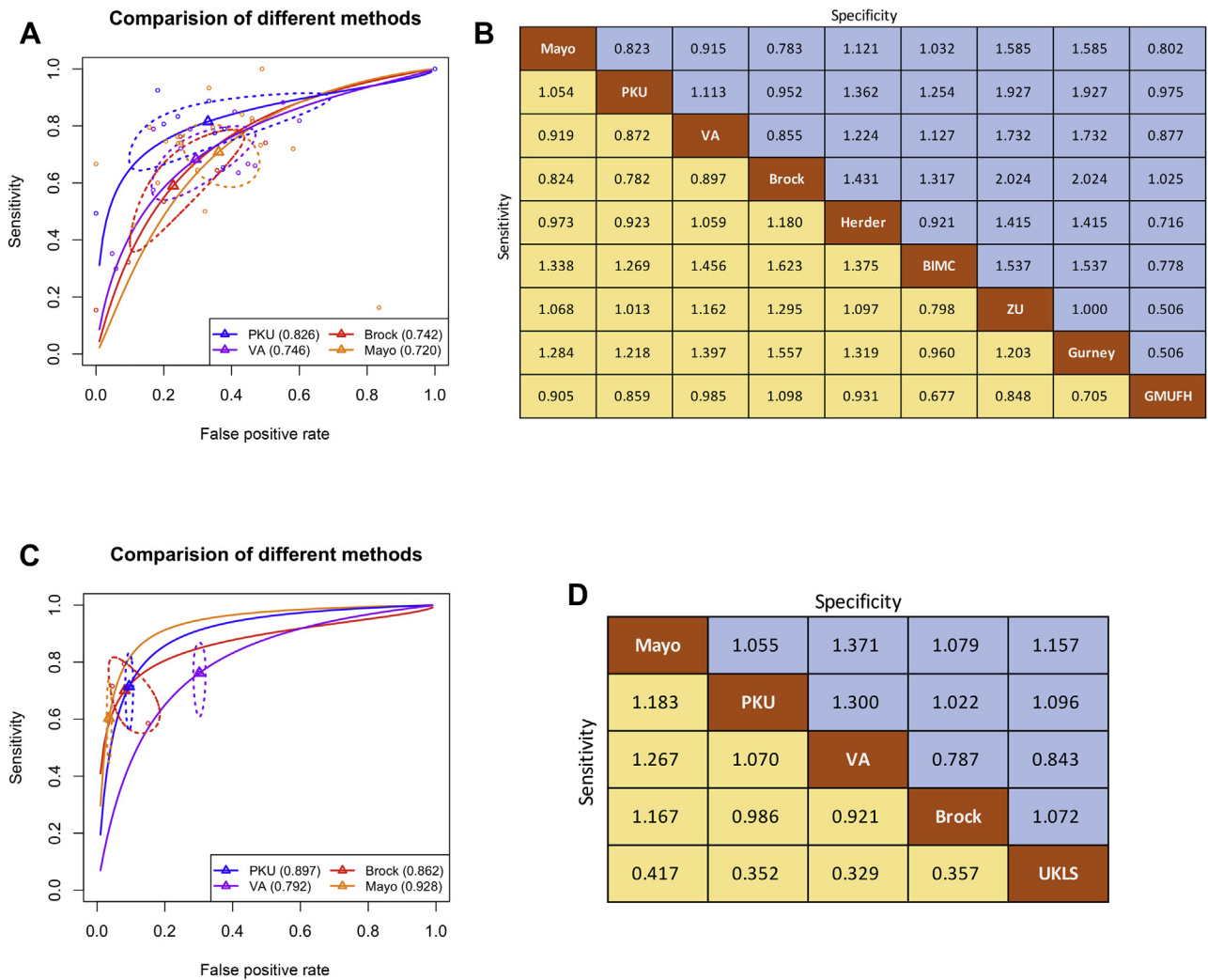
Figure 3. Subgroup analysis based on patient characteristics. (*A*) SROC curves of models with sufficient external validation (at least five independent cohorts) used in HP patients. (*B*) Comparison of the pooled sensitivity and specificity of the validated models in HP patients. (*C*) SROC curves of models with sufficient external validation used in LP patients. (*D*) Comparison of the pooled sensitivity and specificity of the validated models in LP patients. In the SROC plots, the solid line depicts the SROC curve plotted, and individual observations are marked with a round point. The summary point is marked with a triangle point on the SROC curve, and its 95% confidence region is plotted with a dotted line. Different colors are assigned to each model. AUC values are listed in parentheses after the model names in the figure legend. In the comparison of the diagnostic value, the value in each cell is defined as the pooled sensitivity/specificity of the model in the same row divided by the pooled sensitivity/specificity of the model in the same column. Cells with the model name are marked in orange, and cells containing the sensitivity and specificity values are marked in yellow and blue, respectively. AUC, area under the curve; HP, high prevalence; LP, low prevalence; SROC, summary receiver operating characteristic.

The characteristic of this type of model is that benign nodules account for most of the PNs enrolled in model development. The other category is models on the basis of patients treated in clinic or surgery. The characteristic of this type of model is that eligible patients for model development have already undergone preliminary screening, during which only people with HP nodules are admitted for further treatment. Therefore, the malignant rate differs considerably in these two categories. As found in Supplementary Figure 1, the malignant rate in the first category is below 25%, whereas in the other category, this rate is usually above 40%. As a result, the

effectiveness of the models established by different populations differs, and it is not fair to compare different models using the same population. The original cut-off value may no longer be suitable if the models are not used on the targeted populations (Supplementary Table 4). Nevertheless, previous studies often failed to compare these two types of models in different populations. For the first time, we distinguished between the HP group and the LP group on the basis of the probability of malignancy. Furthermore, subgroup analysis revealed that regardless of the HP model or LP model, sensitivity and specificity dropped as long as they were
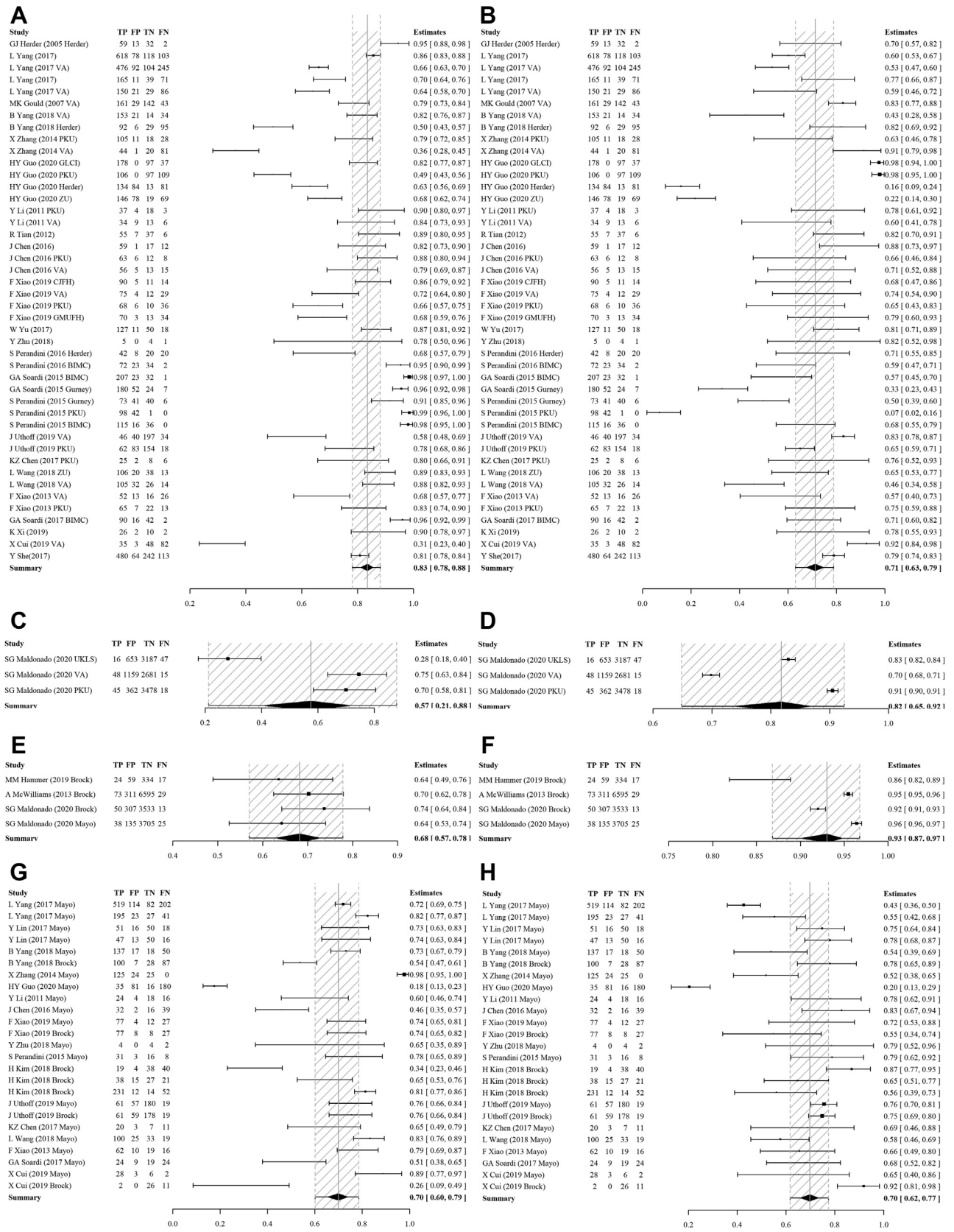
Figure 4. Subgroup analysis of the effect of study population on the models. (*A*) Forest plot of the pooled sensitivity when the HP model is used on HP patients. (*B*) Forest plot of the pooled specificity when the HP model is used on HP patients. (*C*) Forest plot of the pooled sensitivity when the HP model is used on LP patients. (*D*) Forest plot of the pooled specificity when the HP

not used on the targeted populations. Thus, we recommend that the more suitable model should be used for the appropriate cohorts to achieve the best result.

Nevertheless, there are few limitations in our analysis. First, although AUC is the most important indicator of model accuracy, AUC alone cannot comprehensively describe a model. For example, performance calibration is also important for clinical use of a model, but not enough data were provided for a subgroup analysis in our research. Moreover, most of the values of sensitivity and specificity are acquired by Youden index. Although the Youden index provides the highest overall accuracy, in some cases, one would prefer additional sensitivity at the loss of some specificity or vice versa, which makes the Youden index not suitable for some clinical scenario. Another limitation lies in the cohort included in the analysis, as the prevalence was much higher than some recent screening cohorts, which may lead to bias when used in these cohorts. Moreover, owing to the lack of data, some results bear a large 95% CI, which makes the conclusions not so determinate, and more cohorts are needed (especially outpatient cohorts) for further validation to achieve a more accurate result in the comparisons of models.

It is noteworthy that PET-CT results are included in the final model as long as they are collected, suggesting that positive PET-CT results are a strong indicator for malignancy. Nevertheless, problems remain for PEC-CT, which are as follows: (1) although PET-CT improves the sensitivity, it may also have a false-positive result for inflammation, tuberculosis, and so on; (2) PET-CT is only recommended for solid nodules instead of ground-glass opacities, making the clinical application of the model restricted to solid nodules; and (3) because of the high cost of PET-CT in some countries, the result is not available in all situations, which is also a limitation for clinical application. These limitations are also the reason why PET-CT is viewed as a preclinical evaluation instead of as a standard procedure for lung cancer diagnosis by most researchers.

A few studies evaluated both models and physician assessments. We analyzed these articles and found that the models have a better result than the clinicians, but there were no significant differences (Supplementary Fig. 5A). It is important to note that in these studies, the physicians were experienced and familiar with the models, which might lead to bias. The greatest strength of the models is that they are stable and easy to widely use. In fact, many doctors in small hospitals or rural

hospitals do not have sufficient experience in differentiating benign and malignant PNs. We believe that the models are more accurate than these doctors and thus are of value in clinical application. Another advantage of prediction models would be the objectivity. Physician judgment of the same nodule may vary in different scenarios (different environment, emotional state, physical state, etc.), but the prediction result of a model stays the same, making the model's assessment much more objective and repeatable. We believe that a better model could aid more in clinical work both for experienced clinicians and younger clinicians and could make a more objective conclusion for patients.

Although there are guidelines recommending using models for PN evaluation, the clinical applications of these prediction models are still limited. An important reason for this is that the mathematical formula is not practical for clinical practice, and it is time consuming to calculate the cancer probability of each nodule encountered. In fact, models can be exported to an easy-to-read form, such as nomograms or web calculators. Especially for the latter, with only the clinical information of the patient as input, the malignant risk of a PN can be conveniently calculated in less than 10 seconds. Nevertheless, only a few models had a web version built at the time of publication. An example is given in Supplementary Figure 6, illustrating the typical calculation process with the mathematical formula, nomogram, and web calculator, revealing that the nomogram and web calculator are clearer and easier for clinical use. In addition, decision-making in clinical practice cannot exclusively depend on the risk probability of the model. On one hand, clinical treatment is also affected by many other factors, such as patient preference. On the other hand, the model is merely a generally comprehensive analysis. Owing to the heterogeneity of patients' nodules, the model might not be accurate for individuals. The role of the model is to only provide a relatively reliable reference for clinical judgment, but it cannot completely replace the clinicians to make the final decision.

In recent years, AI has started to play a role in the cancer risk prediction of PNs, and AI-based prediction models have been compared with traditional mathematical models. Therefore, we summarized all reported computer-aided diagnosis (CADx) systems with their AUC values (Supplementary Table 5). As we were preparing this meta-analysis, Baldwin et al.[10] reported the result of a comparison of CADx system and the Brock model, revealing that the CADx achieved a better AUC.

model is used on LP patients. (*E*) Forest plot of the pooled sensitivity when the LP model is used on LP patients. (*F*) Forest plot of the pooled specificity when the LP model is used on LP patients. (*G*) Forest plot of the pooled sensitivity when the LP model is used on HP patients. (*H*) Forest plot of the pooled specificity when the LP model is used on HP patients. FN, false negative; FP, false positive; HP, high prevalence; LP, low prevalence; TN, true negative; TP, true positive.

Table 2. Summarization of Highlights of Different Models

First Established Model	Gurney et al.[1] (First Model Using Bayesian Analysis)	Mayo (1997) (First Model Using Logistic Regression)
Model with the largest sample size	Brock (7008 nodules)	
Most verified model	Mayo (compared in 28 articles)	
Best performing model	BIMC (among all validated models)	PKUPH (among all models validated by ≥5 cohorts)
Model with the most variables collected	Mayo (23 variables)	
Models with external validation when established	Brock, TREAT	
Models compared with physicians	Gurney, Mayo, VA, Brock	
Models with a nomogram or a web calculator	Y She et al., Herder, BIMC, GLCI	
Sample with highest and lowest cancer rates	Highest: TREAT	Lowest: Brock
Models with highest and lowest cut-off values (mentioned in original article)	Highest: CJFH (0.794)	Lowest: W Yu et al. (0.3649)
Model that has been compared with AI models	Brock (compared with AI based on CNN in David Baldwin et al., AI had better result in HP patients)	

AI, artificial intelligence; BIMC, Bayesian inference malignancy calculator; CJFH, China-Japan Friendship Hospital; CNN, convolutional neural networks; GLCI, Guangdong Lung Cancer Institute; HP, high prevalence; PKUPH, Peking University People's Hospital; TREAT, thoracic research evaluation and treatment; VA, Department of Veterans Affairs.

Nevertheless, the analysis was conducted on a HP cohort, whereas the accuracy of the Brock model might be underestimated as it was developed on the basis of LP cohorts. According to our result, it is more equitable in future studies to use identical background models for the comparison to evaluate whether such AI models have transcended the mathematical model. Despite the outstanding AUC values reported, problems remain for the CADx systems. First, because researchers seldom provide a model for external validation, there is a lack of prospective studies to validate its efficacy in different populations. In addition, because of the lack of clinical information in open data sets, all current CADx systems can only predict malignant risk with radiographical characteristics and cannot take clinical features into consideration as clinicians or models do. In some cases, such as the evaluation of PNs in patients with a history of cancer, the judgment of CADx may bear notable bias. Therefore, further exploration and improvement are still needed for the CADx systems. Mathematical models and machine learning models are both statistical models in some ways. Deep learning, the typical method used in nodule detection, uses simple functions such as the sigmoid function or the rectified linear unit function as the activation function inside individual neurons. The utilization of large amounts of neurons results in a multilayer network, whereas its purpose is still to separate nodules into benign and malignant nodules. Therefore, it is safe to say that this network places individual observations into a higher dimension and finds a function to fit the observations, which is basically the idea behind mathematical models. In fact, the discovery of new regression functions is how humans fit the observations, whereas the training of neural networks is how machines fit them. In some ways, AI is an extension of mathematical models. According to our result, it is possible that with the continuous training of AI, its diagnostic efficiency may be further improved, and eventually exceed the prediction accuracy of mathematical models in the future. Nevertheless, so far, there is no enough evidence to prove that the accuracy of AI can be improved in the future and more comparisons between AI-based models and mathematical models in the same population are still needed. Therefore, for now, the widely validated mathematical models are still the most convenient and relatively accurate way to assist PN management.

In conclusion, we systematically reviewed and analyzed a variety of prediction models of PNs. The Mayo model is the most widely used and validated model, whereas the PKU model yields the best AUC among the most often used models. Because of the discrepant development cohorts among the models, it is vital that the most suitable model is used on the appropriate cohorts, and mixing models might lead to decreased accuracy. Nomograms or web calculators are intuitive and preferred by clinicians, but their clinical application needs to be further investigated.

CRediT Authorship Contribution Statement

Kai Zhang: Methodology, Formal analysis, Data curation, Writing - original draft, Writing - review & editing.

Zihan Wei: Methodology, Formal analysis, Data curation, Writing - original draft.

Yuntao Nie: Formal analysis, Data curation, Validation.

Haifeng Shen: Formal analysis, Data curation.

Xin Wang: Data curation.

Jun Wang: Conceptualization, Supervision.

Fan Yang: Methodology, Supervision.

Kezhong Chen: Conceptualization, Methodology, Supervision, Writing - review & editing.

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *JTO Clinical and Research Reports* at www.jtocrr.org and at https://doi.org/10.1016/j.jtocrr.2022.100299.

Acknowledgments

References

1. de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med*. 2020;382:503-513.

2. Field JK, Duffy SW, Baldwin DR, et al. The UK Lung Cancer Screening Trial: a pilot randomised controlled trial of low-dose computed tomography screening for the early detection of lung cancer. *Health Technol Assess*. 2016;20:1-146.

3. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395-409.

4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68:7-30.

5. Yousaf-Khan U, van der Aalst C, de Jong PA, et al. Final screening round of the NELSON lung cancer screening trial: the effect of a 2.5-year screening interval. *Thorax*. 2017;72:48-56.

6. Infante M, Sestini S, Galeone C, et al. Lung cancer screening with low-dose spiral computed tomography: evidence from a pooled analysis of two Italian randomized trials. *Eur J Cancer Prev*. 2017;26:324-329.

7. Callister ME, Baldwin DR, Akram AR, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax*. 2015;70(suppl 2):ii1-ii54.

8. Gould MK, Donington J, Lynch WR, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer. 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2013;143(suppl):e93S-e120S.

9. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology*. 2017;284:228-243.

10. Ost D, Fein AM, Feinsilver SH. Clinical practice. The solitary pulmonary nodule. *N Engl J Med*. 2003;348:2535-2542.

11. González Maldonado S, Delorme S, Hüsing A, et al. Evaluation of prediction models for identifying malignancy in pulmonary nodules detected via low-dose computed tomography. *JAMA Netw Open*. 2020;3: e1921221.

12. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6:e1000097.

13. Baldwin DR, Gustafson J, Pickup L, et al. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. *Thorax*. 2020;75:306-312.

14. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58:982-990.

15. Chen K, Nie Y, Park S, et al. Development and validation of machine learning-based model for the prediction of malignancy in multiple pulmonary nodules: analysis from multicentric cohorts. *Clin Cancer Res*. 2021;27:2255-2265.

16. Nyaga VN, Aerts M, Arbyn M. ANOVA model for network meta-analysis of diagnostic test accuracy data. *Stat Methods Med Res*. 2018;27:1766-1784.

17. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529-536.

18. Gurney JW, Lyddon DM, McKay JA. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part II. Application. *Radiology*. 1993;186:415-422.

19. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med*. 1997;157:849-855.

20. Herder GJ, van Tinteren H, Golding RP, et al. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. *Chest*. 2005;128:2490-2496.

21. Gould MK, Ananth L, Barnett PG, Veterans Affairs SNAP Cooperative Study Group. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest*. 2007;131:383-388.

22. Schultz EM, Sanders GD, Trotter PR, et al. Validation of two models to estimate the probability of malignancy in patients with solitary pulmonary nodules. *Thorax*. 2008;63:335-341.

23. Li Y, Chen KZ, Wang J. Development and validation of a clinical prediction model to estimate the probability of malignancy in solitary pulmonary nodules in Chinese people. *Clin Lung Cancer*. 2011;12:313-319.

24. Tian R, Su MG, Tian Y, Li FL, Kuang AR. [Development of a predicting model to estimate the probability of malignancy of solitary pulmonary nodules]. *Sichuan Da Xue Xue Bao Yi Xue Ban*. 2012;43:404-408.

25. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med*. 2013;369:910-919.

26. Xiao F, Liu D, Guo Y, et al. Novel and convenient method to evaluate the character of solitary pulmonary nodule-comparison of three mathematical prediction models and further stratification of risk factors. *PLoS One*. 2013;8:e78271.

27. Deppen SA, Blume JD, Aldrich MC, et al. Predicting lung cancer prior to surgical resection in patients with lung nodules. *J Thorac Oncol*. 2014;9:1477-1484.

28. Zhang X, Yan HH, Lin JT, et al. Comparison of three mathematical prediction models in patients with a solitary pulmonary nodule. *Chin J Cancer Res*. 2014;26:647-652.

29. Al-Ameri A, Malhotra P, Thygesen H, et al. Risk of malignancy in pulmonary nodules: A validation study of four prediction models. *Lung Cancer*. 2015;89:27-30.

30. Vachani A, Pass HI, Rom WN, et al. Validation of a multiprotein plasma classifier to identify benign lung nodules. *J Thorac Oncol*. 2015;10:629-637.

31. Soardi GA, Perandini S, Motton M, Montemezzi S. Assessing probability of malignancy in solid solitary pulmonary nodules with a new Bayesian calculator: improving diagnostic accuracy by means of expanded and updated features. *Eur Radiol*. 2015;25:155-162.

32. Yang J, Xue-Li S, Guo-Xiang L, Qiang L, Yi-Ping H, Han Y. Establishment and validation of mathematics model for differentiating benign and malignant solitary pulmonary nodules. *Acad J Second Mil Med Univ*. 2015;36:407-412.

33. Zhang M, Zhuo N, Guo Z, et al. Establishment of a mathematic model for predicting malignancy in solitary pulmonary nodules. *J Thorac Dis*. 2015;7:1833-1841.

34. Chen J, Xiao-Dan Y, Jian-Ding Y, Zheng-Rong Z. Establishment and verification of a mathematical model for predicting malignancy of solitary pulmonary nodules. *Chin Comput Med Imaging*. 2016;22:573-577.

35. Perandini S, Soardi GA, Larici AR, et al. Multicenter external validation of two malignancy risk prediction models in patients undergoing 18F-FDG-PET for solitary pulmonary nodule evaluation. *Eur Radiol*. 2017;27:2042-2046.

36. Perandini S, Soardi GA, Motton M, Rossi A, Signorini M, Montemezzi S. Solid pulmonary nodule risk assessment and decision analysis: comparison of four prediction models in 285 cases. *Eur Radiol*. 2016;26:3071-3076.

37. Soardi GA, Perandini S, Larici AR, et al. Multicentre external validation of the BIMC model for solid solitary pulmonary nodule malignancy prediction. *Eur Radiol*. 2017;27:1929-1933.

38. Chen K, Zhang J, Guan T, et al. Comparison of plasma to tissue DNA mutations in surgical patients with non-small cell lung cancer. *J Thorac Cardiovasc Surg*. 2017;154:1123-1131.e2.

39. Yang L, Zhang Q, Bai L, et al. Assessment of the cancer risk factors of solitary pulmonary nodules. *Oncotarget*. 2017;8:29318-29327.

40. Tanner NT, Porter A, Gould MK, Li XJ, Vachani A, Silvestri GA. Physician assessment of pretest probability of malignancy and adherence with guidelines for pulmonary nodule evaluation. *Chest*. 2017;152:263-270.

41. Wei Y, Tianxiang C, Liyun X, Zhaoyu W, Hanbo C, Yongkui Z. Establishment of the diagnostic model in solitary pulmonary nodule appearing as ground-glass nodule. *Chin J Med Imaging*. 2017;25:435-440.

42. Lin Y, Leng Q, Jiang Z, et al. A classifier integrating plasma biomarkers and radiological characteristics for distinguishing malignant from benign pulmonary nodules. *Int J Cancer*. 2017;141:1240-1248.

43. She Y, Zhao L, Dai C, et al. Development and validation of a nomogram to estimate the pretest probability of cancer in Chinese patients with solid solitary pulmonary nodules: a multi-institutional study. *J Surg Oncol*. 2017;116:756-762.

44. Yang B, Jhun BW, Shin SH, et al. Comparison of four models predicting the malignancy of pulmonary nodules: A single-center study of Korean adults. *PLoS One*. 2018;13:e0201242.

45. Kim H, Park CM, Jeon S, et al. Validation of prediction models for risk stratification of incidentally detected pulmonary subsolid nodules: a retrospective cohort study in a Korean tertiary medical centre. *BMJ Open*. 2018;8:e019996.

46. Wang L, Chen Y, Tang K, Lin J, Zhang H. The value of ^{18}F-FDG PET/CT Mathematical Prediction Model in Diagnosis of Solitary Pulmonary Nodules. *BioMed Res Int*. 2018;2018:9453967.

47. Nair VS, Sundaram V, Desai M, Gould MK. Accuracy of models to identify lung nodule cancer risk in the National Lung Screening Trial. *Am J Respir Crit Care Med*. 2018;197:1220-1223.

48. Ying ZHU, Panfeng XU, Yake YAO, Jianfang P, Jianying ZHOU. Independent predictors and prediction model of malignant micro-sized solitary pulmonary nodules. *Chin J Clin Oncol*. 2018;45:497-502.

49. Winter A, Aberle DR, Hsu W. External validation and recalibration of the Brock model to predict probability of cancer in pulmonary nodules using NLST data. *Thorax*. 2019;74:551-563.

50. Xiao F, Yu Q, Zhang Z, Liu D, Liang C. [Establishment and verification of A novel predictive model of Malignancy for Non-solid Pulmonary Nodules]. *Zhongguo Fei Ai Za Zhi*. 2019;22:26-33.

51. Kim H, Goo JM, Park CM. A simple prediction model using size measures for discrimination of invasive adenocarcinomas among incidental pulmonary subsolid nodules considered for resection. *Eur Radiol*. 2019;29:1674-1683.

52. Uthoff J, Koehn N, Larson J, et al. Post-imaging pulmonary nodule mathematical prediction models: are they clinically relevant? *Eur Radiol*. 2019;29:5367-5377.

53. Xi K, Wang W, Wen Y, et al. Combining plasma miRNAs and computed tomography features to differentiate the nature of pulmonary nodules. *Front Oncol*. 2019;9:975.

54. Hammer MM, Palazzo LL, Kong CY, Hunsaker AR. Cancer risk in subsolid nodules in the National Lung Screening Trial. *Radiology*. 2019;293:441-448.

55. Marcus MW, Duffy SW, Devaraj A, et al. Probability of cancer in lung nodules using sequential volumetric screening up to 12 months: the UKLS trial. *Thorax*. 2019;74:761-767.
56. Cui X, Heuvelmans MA, Han D, et al. Comparison of Veterans Affairs, Mayo, Brock classification models and radiologist diagnosis for classifying the malignancy of pulmonary nodules in Chinese clinical population. *Transl Lung Cancer Res*. 2019;8:605-613.
57. Guo HY, Lin JT, Huang HH, et al. Development and validation of a ^{18}F-FDG PET/CT-based clinical prediction model for estimating malignancy in solid pulmonary nodules based on a population with high prevalence of malignancy. *Clin Lung Cancer*. 2020;21:47-55.
58. Li Y, Hu H, Wu Z, et al. Evaluation of models for predicting the probability of malignancy in patients with pulmonary nodules. *Biosci Rep*. 2020;40:BSR20193875.
59. Schreuder A, Scholten ET, van Ginneken B, Jacobs C. Artificial intelligence for detection and characterization of pulmonary nodules in lung cancer CT screening: ready for practice? *Transl Lung Cancer Res*. 2021;10:2378-2388.