BMJ Open

Validation of prediction models for risk stratification of incidentally detected pulmonary subsolid nodules: a retrospective cohort study in a Korean tertiary medical centre

Hyungjin Kim,[1] Chang Min Park,[1,2,3] Sunkyung Jeon,[1] Jong Hyuk Lee,[1] Su Yeon Ahn,[1] Roh-Eul Yoo,[1,4] Hyun-ju Lim,[1] Juil Park,[1] Woo Hyeon Lim,[1] Eui Jin Hwang,[1] Sang Min Lee,[5] Jin Mo Goo[1,3,2]

Check for updates

[1]Department of Radiology, Seoul National University College of Medicine, Seoul, Korea
[2]Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Korea
[3]Seoul National University Cancer Research Institute, Seoul, Korea
[4]Department of Radiology, National Cancer Center, Goyang, Korea
[5]Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Seoul, Korea

Correspondence to
Dr Chang Min Park;
cmpark.morphius@gmail.com

ABSTRACT

Objectives To validate the performances of two prediction models (Brock and Lee models) for the differentiation of minimally invasive adenocarcinoma (MIA) and invasive pulmonary adenocarcinoma (IPA) from preinvasive lesions among subsolid nodules (SSNs).

Design A retrospective cohort study.

Setting A tertiary university hospital in South Korea.

Participants 410 patients with 410 incidentally detected SSNs who underwent surgical resection for the pulmonary adenocarcinoma spectrum between 2011 and 2015.

Primary and secondary outcome measures Using clinical and radiological variables, the predicted probability of MIA/IPA was calculated from pre-existing logistic models (Brock and Lee models). Areas under the receiver operating characteristic curve (AUCs) were calculated and compared between models. Performance metrics including sensitivity, specificity, accuracy, positive predictive value (PPV) and negative predictive value (NPV) were also obtained.

Results For pure ground-glass nodules (n=101), the AUC of the Brock model in differentiating MIA/IPA (59/101) from preinvasive lesions (42/101) was 0.671. Sensitivity, specificity, accuracy, PPV and NPV based on the optimal cut-off value were 64.4%, 64.3%, 64.4%, 71.7% and 56.3%, respectively. Sensitivity, specificity, accuracy, PPV and NPV according to the Lee criteria were 76.3%, 42.9%, 62.4%, 65.2% and 56.3%, respectively. AUC was not obtained for the Lee model as a single cut-off of nodule size (≥10 mm) was suggested by this model for the assessment of pure ground-glass nodules. For part-solid nodules (n=309; 26 preinvasive lesions and 283 MIA/IPAs), the AUC was 0.746 for the Brock model and 0.771 for the Lee model (p=0.574). Sensitivity, specificity, accuracy, PPV and NPV were 82.3%, 53.8%, 79.9%, 95.1% and 21.9%, respectively, for the Brock model and 77.0%, 69.2%, 76.4%, 96.5% and 21.7%, respectively, for the Lee model.

Conclusions The performance of prediction models for the incidentally detected SSNs in differentiating MIA/IPA from preinvasive lesions might be suboptimal. Thus, an alternative risk calculation model is required for the incidentally detected SSNs.

Strengths and limitations of this study

► This is the first study to externally validate the performance of pre-existing risk prediction models for the incidentally detected pulmonary subsolid nodules.
► This study performed head-to-head comparisons between the prediction models for the risk stratification of subsolid nodules.
► The main limitation of this study is that it only analysed surgically resected lung nodules, thus inducing selection bias.
► Study population was small to conduct separate analyses for the pure ground-glass nodules and part-solid nodules.

INTRODUCTION

Pulmonary subsolid nodules (SSNs) represent a histological spectrum of adenocarcinoma, and its preinvasive precursors, including atypical adenomatous hyperplasia (AAH) and adenocarcinoma-in-situ (AIS).[1] SSNs are common findings at chest CT which have been increasingly detected in CT screening studies.[2 3] Indeed, according to one prospective screening study, 4.2% of the participants had at least one pure ground-glass nodule (pGGN) and 5.0% had at least one part-solid nodule (PSN) at baseline rounds of screening.[2]

With this prevalence in mind, numerous studies have justifiably focused on the differentiation of invasive adenocarcinomas from preinvasive lesions[4–13] as invasive adenocarcinoma requires surgical resection with conventional lobectomy and lymph node dissection[14] whereas preinvasive lesions can be followed up conservatively with annual CT surveillance or resected at a lesser extent

(sublobar resection).[15] Thus, the discrimination of invasive adenocarcinoma has been a major topic of interest for many radiologists and clinicians to date.

In a quest to obtain quantitative risk-prediction tools for pulmonary nodules, McWilliams *et al*[16] developed a prediction model (Brock model) using various clinical and radiological features. The Brock model demonstrated higher accuracy in determining the likelihood of malignancy in pulmonary nodules compared with other existing models[17] and was also externally validated in three independent screening populations.[18–20] Nevertheless, in the context that a substantial percentage of persistent SSNs may belong to the adenocarcinoma spectrum, the performance of the established model in differentiating invasive adenocarcinoma should also be validated in order to encourage the use of the model in routine practice, as suggested by the British Thoracic Society (BTS).[21]

Lee *et al*[7] also developed a prediction model (Lee model) using simple size metrics and morphological features for the differentiation of invasive adenocarcinomas appearing as SSNs. The model accuracy was reported to be excellent for the identification of invasive adenocarcinomas. However, it has also not been tested or validated.

Therefore, we aimed to validate the performances of the two prediction models (Brock and Lee models) for the differentiation of minimally invasive adenocarcinomas (MIAs) and invasive pulmonary adenocarcinomas (IPAs) from preinvasive lesions among SSNs. The purpose of our study was to evaluate the feasibility of the two models in the risk stratification of persistent SSNs.

METHODS
Study population
We retrospectively reviewed the electronic medical records of our hospital and found 1915 patients who had undergone surgical resection for lung cancer between 2011 and 2015. Among the 1915 patients, we identified 1073 patients whose pathological diagnoses belonged to the pulmonary adenocarcinoma spectrum including AAH, AIS, MIA and IPA.[1 22] Thereafter, we reviewed the thin-section CT images of the patients to include only those with SSNs (reviewers: JSK, JHL, SYA, REY, HL and HK); 548 patients whose lung cancers appeared as solid nodules on CT scans were excluded. We also excluded 76 patients with nodules smaller than 5 mm or larger than 3 cm and 39 patients in whom data regarding the family history of lung cancer were not available. Consequently, 410 patients were included in this study. Among these patients, 18 patients had two nodules and one patient had three nodules. A single nodule was selected randomly for these 19 patients in order to remove within-subject correlation. Therefore, a total of 410 nodules from 410 patients were analysed in the present study (figure 1). There were 174 men and 236 women (median, 61 years; IQR: 54–69 years). As for the nodule type, there were 101 pGGNs and 309 PSNs. IPAs were found in 290 nodules
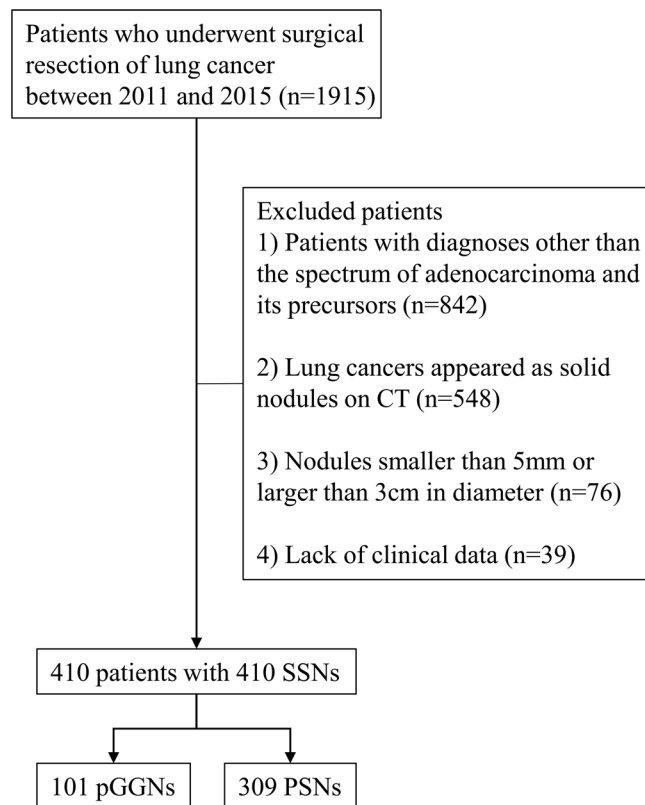

Figure 1 Flow chart of patient inclusion and exclusion. pGGN, pure ground-glass nodule; PSN, part-solid nodule; SSN, subsolid nodule.

followed by MIA in 52 nodules, AIS in 51 nodules and AAH in 17 nodules. Median nodule size was 15.8 mm (IQR: 11.8–20.9 mm) (table 1).

Data collection
Patient characteristics including demographic data were collected from the electronic medical records of Seoul National University Hospital. Patient age, sex, pathological diagnosis, family history of lung cancer and nodule location (lobe) were recorded. The thin-section CT images were also reviewed to obtain radiological information of nodules (nodule type, nodule size, solid portion size, solid proportion, lobulation, spiculation and nodule count per scan) and the background lung parenchyma (presence of visually detected emphysema). These features were used as input variables for logistic regression analysis at the Brock model[16] and Lee model.[7] Nodule size and solid portion size were measured as the maximum transverse diameter (mm) using an electronic calliper. Solid proportion (%) was calculated as the solid portion size divided by the nodule size. Nodule count was defined as the total number of non-calcified nodules at least 1 mm in diameter.[16] Image review was conducted by three radiologists (JP, WHL and HK), and each nodule was analysed once by one of these radiologists. Details regarding the CT scanning protocols are described in the online supplementary material.

Table 1 Demographic data of the entire study population

Characteristics	Value
Age (year)*	61 (54–69)
Sex	
Male	174 (42.4)
Female	236 (57.6)
Pathology	
AAH	17 (4.1)
AIS	51 (12.4)
MIA	52 (12.7)
IPA	290 (70.7)
Nodule type	
Pure GGN	101 (24.6)
Part-solid nodule	309 (75.4)
Family history of lung cancer	
Yes	16 (3.9)
No	394 (96.1)
Emphysema	
Yes	44 (10.7)
No	366 (89.3)
Nodule size (mm)*	15.8 (11.8–20.9)
Solid portion size (mm)*	5.7 (1.6–11.2)
Solid proportion (%)*	41.0 (11.4–62.0)
Location	
Upper lobe	249 (60.7)
Other lobes	161 (39.3)
Lobulation	
Yes	138 (33.7)
No	272 (66.3)
Spiculation	
Yes	134 (32.6)
No	276 (67.3)
Nodule count per scan*	3 (1–5)

Unless otherwise specified, data are numbers of patients (with percentages in parentheses).
*Data are median (with IQR in parentheses).
AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma-in-situ; GGN, ground-glass nodule; IPA, invasive pulmonary adenocarcinoma; MIA, minimally invasive adenocarcinoma.

Measurement variability

We previously analysed and reported the measurement variability of SSNs and solid portion size using two same-day repeat CT scans.[23] Measurement variability range for the maximum transverse diameter of SSNs on lung window CT images was ±2.2 mm. For the solid portion, it was ±3.7 mm. Inter-reader agreement (κ) of nodule type ranged from 0.80 to 0.96. Therefore, we did not re-evaluate the measurement variability or inter-reader agreement of nodule type in this study.

Statistical analysis

To investigate whether the variables incorporated in the established models (Brock and Lee models) were significantly different between preinvasive (AAH and AIS) and invasive lesions (MIA and IPA), we first performed a univariate analysis. Categorical variables were analysed using the Pearson χ^2 test or Fisher's exact test, and continuous variables were analysed using the independent t-test or Mann-Whitney U test, as appropriate.

We then calculated the predicted probability from each logistic regression model. For the Brock model, a full model with spiculation was used with input variables of age, sex, family history of lung cancer, emphysema, nodule size, nodule type, nodule location, nodule count per scan and spiculation.[16] Regression coefficients and the model constant were available from the original paper.[16] Nodule size was subjected to power transformation prior to entry as described previously.[16] Age and nodule count were centred at a mean of 62 years and 4, respectively.[16] We recorded the predicted probability of each nodule which was a continuous value from 0 to 1 (0 to 100%). For the Lee model, two different methods were used for analysis. For pGGNs, a single cut-off of nodule size (≥10 mm) was used to discriminate invasive lesions as stated by Lee *et al*.[7] In the case of PSNs, four variables (nodule size, solid proportion, lobulation and spiculation) were substituted into the following regression formula.[7]

$$\text{Logit (Probability)} = 0.396 - 0.200 \times [\text{nodule size (mm)}] - 0.048 \times [\text{solid proportion (\%)}] + 1.049 \times (\text{nonlobulation}) + 3.288 \times (\text{nonspiculation})$$

This logistic regression formula was originally made to predict a preinvasive lesion. Therefore, predicted probability for an invasive lesion was calculated as '1 - probability of being a preinvasive lesion'. Predicted probability was obtained only for PSNs in terms of the Lee model. No preprocessing of variables was performed.

With the predicted probability obtained through each model, receiver operating characteristic curve (ROC) analysis was performed to investigate the discriminative performance of the prediction models in diagnosing invasive lesions. Areas under the ROC curve (AUCs) were obtained and an optimal cut-off value based on the Youden Index was recorded. We calculated the sensitivity, specificity, accuracy, positive predictive value (PPV) and negative predictive value (NPV) of each model with the optimal cut-off. For the Brock model, two different cut-offs were applied to the calculation: (1) a threshold of 10% risk of malignancy as suggested by the BTS[21] and (2) an optimal cut-off based on the Youden Index. ROC analysis was performed for each nodule type separately and then for the entire SSNs in the case of the Brock model. In terms of the Lee model, ROC analysis was performed only for PSNs.

AUCs were compared between the models based on DeLong's method.[24] As the predicted probability of pGGNs was not available for the Lee model, AUC comparison was conducted only for PSNs. Diagnostic

accuracy was also compared between the models using the McNemar test.

Lastly, calibration of the models was assessed using the Hosmer-Lemeshow test for the 10 probability groups (deciles). All statistical analyses were performed using two commercial software programs (MedCalc V.12.3.0; MedCalc Software, Mariakerke, Belgium and SPSS V.19.0; IBM SPSS Statistics) and R software V.3.1.0 (http://www.R-project.org; PredictABEL package). A p value <0.05 was considered to indicate statistical significance.

Patient and public involvement

Patients or public were not involved in the development of the research question and outcome measures. No patients were involved in the study design or conduct of the study. Dissemination of the study results to the study participants was not practical given the retrospective nature of our study. Lastly, there were no patient advisers.

RESULTS
Pathological diagnoses of pGGNs and PSNs

Among 101 pGGNs, 42 were preinvasive and 59 were invasive lesions. As for the 309 PSNs, 26 were preinvasive and 283 were invasive lesions.

Comparisons between preinvasive and invasive lesions

For pGGNs, a family history of lung cancer was more frequently observed in patients with invasive lesions (6/59) than in those with preinvasive lesions (0/42; p=0.040). Invasive lesions (14.2±5.4 mm) were also significantly larger than preinvasive lesions (11.1±4.1 mm; p=0.002). In addition, patients with invasive lesions had a smaller nodule count per scan (invasive vs preinvasive lesions: median, 2 vs 4 nodules per scan; p=0.006). There were no significant differences in age, sex, presence of emphysema, nodule location, lobulation and spiculation (table 2).

For PSNs, nodule size, solid portion size and solid proportion were significantly larger in invasive lesions (invasive vs preinvasive lesions: median nodule size, 17.6 mm vs 13.6 mm, p<0.001; median solid portion size, 8.4 mm vs 4.6 mm, p<0.001; median solid proportion, 52.8% vs 36.8%, p=0.032). Lobulation and spiculation were more frequently observed in invasive lesions (invasive vs preinvasive lesions: lobulation, 118/283 vs 5/26, p=0.025; spiculation, 122/283 vs 4/26, p=0.006). There were no significant differences in age, sex, family history of lung cancer, presence of emphysema, nodule location and nodule count per scan (table 3).

SSN risk stratification using the Brock and Lee models

For pGGNs, the AUC of the Brock model's predicted probability for differentiating invasive lesions from preinvasive lesions was 0.671 (95% CI: 0.571 to 0.762) (table 4). A cut-off of 10%, suggested by the BTS, yielded a sensitivity, specificity, accuracy, PPV and NPV of 32.2%, 90.5%, 56.4%, 82.6% and 48.7%, respectively. Another cut-off of

Table 2 Comparison of clinical and radiological characteristics between differing pathological diagnoses in patients with pure GGNs

Characteristics	AAH and AIS (n=42)	MIA and IPA (n=59)	P values
Age (year)*	57±10	57±10	0.839
Sex			
Male	16 (38.1)	29 (49.2)	0.270
Female	26 (61.9)	30 (50.8)	
Family history of lung cancer			
Yes	0 (0)	6 (10.2)	0.040
No	42 (100)	53 (89.8)	
Emphysema			
Yes	5 (11.9)	5 (8.5)	0.738
No	37 (88.1)	54 (91.5)	
Nodule size (mm)*	11.1±4.1	14.2±5.4	0.002
Location			
Upper lobe	25 (59.5)	31 (52.5)	0.487
Other lobes	17 (40.5)	28 (47.5)	
Lobulation			
Yes	3 (7.1)	12 (20.3)	0.090
No	39 (92.9)	47 (79.7)	
Spiculation			
Yes	1 (2.4)	7 (11.9)	0.135
No	41 (97.6)	52 (88.1)	
Nodule count per scan†	4 (2–6)	2 (1–3)	0.006

Unless otherwise specified, data are numbers of patients (with percentages in parentheses).
*Data are mean±SD.
†Data are median (with IQR in parentheses).
AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma-in-situ; GGN, ground-glass nodule; IPA, invasive pulmonary adenocarcinoma; MIA, minimally invasive adenocarcinoma.

4.29%, an optimal threshold based on the Youden Index, provided a sensitivity, specificity, accuracy, PPV and NPV of 64.4%, 64.3%, 64.4%, 71.7% and 56.3%, respectively. Brock model for pGGNs showed poor calibration (p<0.001). A nodule size cut-off (10 mm) suggested by Lee *et al*[7] was also applied to our study population. The resultant sensitivity, specificity, accuracy, PPV and NPV were 76.3%, 42.9%, 62.4%, 65.2% and 56.3%, respectively. There were no significant differences in diagnostic accuracy between the Brock model and Lee criteria (Brock model cut-off 10% vs nodule size cut-off 10 mm, p=0.461; Brock model cut-off 4.29% vs nodule size cut-off 10 mm, p=0.832).

As for PSNs, the AUC of the Brock model was 0.746 (95% CI: 0.694 to 0.794) for the discrimination of invasive lesions from preinvasive lesions (table 5). A cut-off of 10% yielded a sensitivity, specificity, accuracy, PPV and NPV of 82.3%, 53.8%, 79.9%, 95.1% and 21.9%, respectively. The

Table 3 Comparison of clinical and radiological characteristics between differing pathological diagnoses in patients with part-solid nodules

Characteristics	AAH and AIS (n=26)	MIA and IPA (n=283)	P values
Age (year)*	62 (54–68)	63 (56–70)	0.257
Sex			
Male	7 (26.9)	122 (43.1)	0.109
Female	19 (73.1)	161 (56.9)	
Family history of lung cancer			
Yes	1 (3.8)	9 (3.2)	0.590
No	25 (96.2)	274 (96.8)	
Emphysema			
Yes	2 (7.7)	32 (11.3)	0.752
No	24 (92.3)	251 (88.7)	
Nodule size (mm)*	13.6 (9.8–16.6)	17.6 (13.3–22.4)	<0.001
Solid portion size (mm)*	4.6 (3.4–5.8)	8.4 (5.2–13.6)	<0.001
Solid proportion (%)*	36.8 (25.1–63.2)	52.8 (33.4–66.0)	0.032
Location			
Upper lobe	12 (46.2)	181 (64.0)	0.073
Other lobes	14 (53.8)	102 (36.0)	
Lobulation			
Yes	5 (19.2)	118 (41.7)	0.025
No	21 (80.8)	165 (58.3)	
Spiculation			
Yes	4 (15.4)	122 (43.1)	0.006
No	22 (84.6)	161 (56.9)	
Nodule count per scan*	3 (2–8)	3 (1–4)	0.132

Unless otherwise specified, data are numbers of patients (with percentages in parentheses).
*Data are median (with IQR in parentheses).
AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma-in-situ; IPA, invasive pulmonary adenocarcinoma; MIA, minimally invasive adenocarcinoma.

optimal cut-off based on the Youden Index was 10.11% which was very close to the suggested threshold by the BTS. Therefore, performance metrics were not calculated separately. AUC of the Lee model was 0.771 (95% CI: 0.720 to 0.817), and an optimal cut-off of 66.68%

provided a sensitivity, specificity, accuracy, PPV and NPV of 77.0%, 69.2%, 76.4%, 96.5% and 21.7%, respectively. AUCs and diagnostic accuracies were not significantly different between the two models (p=0.574 and p=0.169, respectively). In addition, both models exhibited poor calibration (p<0.001).

With respect to the pooled analysis for the entire SSNs, the AUC of Brock model was 0.810 (95% CI: 0.769 to 0.846). Calibration was also poor in this model (p<0.001).

DISCUSSION

In this study, we revealed that AUCs for the differentiation of invasive lesions among PSNs using the established risk prediction models ranged from 0.746 to 0.771 with no significant differences between the two models. Diagnostic accuracies based on the optimal cut-offs were 79.9% for the Brock model and 76.4% for the Lee model. For pGGNs, the diagnostic accuracy was 56.4%–64.4% for the Brock model depending on the cut-off values used and 62.4% for the Lee criteria. For the entire SSNs, the Brock model showed AUC of 0.810.

McWilliams et al[16] originally developed a lung cancer prediction model (Brock model) using participants enrolled in a lung cancer screening study. Thus, the Brock model initially targeted pulmonary nodules detected on first screening CT. Incidentally detected nodules and surgical candidates were not the original target lesions of this model. However, at present, the BTS recommends using the same diagnostic approach for nodules detected incidentally as those detected through screening.[21] BTS also recommends using the Brock model for the risk calculation of both solid nodules and SSNs.[21] A cut-off of 10% predicted probability for malignancy is suggested in order to differentiate high-risk SSNs for the performance of biopsy or surgical resection.[21] This quantitative diagnostic approach is to discern malignant SSNs with an appropriate false-positive rate. However, it must be noted that most persistent SSNs belong to one of the four categories of the adenocarcinoma spectrum: AAH, AIS, MIA and IPA. Therefore, the potential of the risk-prediction model in discriminating lesions with invasive components (MIA and IPA) should also be tested using pathological

Table 4 Performance of each prediction model in diagnosing minimally invasive adenocarcinoma and invasive adenocarcinoma for pure ground-glass nodules

	Sensitivity	Specificity	Accuracy	PPV	NPV	AUC
Brock model 1*	32.2	90.5	56.4	82.6	48.7	0.671 (0.571–0.762)
Brock model 2†	64.4	64.3	64.4	71.7	56.3	
Lee model (nodule size‡)	76.3	42.9	62.4	65.2	56.3	–

Data are in percentages except for AUC. There were 42 preinvasive lesions and 59 invasive lesions (minimally invasive adenocarcinomas and invasive adenocarcinomas).
*A cut-off of 10% predicted probability was used as suggested by the British Thoracic Society.
†A cut-off of 4.29% predicted probability, the optimal threshold based on the Youden Index, was used.
‡A cut-off of 10 mm nodule size was used.
AUC, area under the receiver operating characteristic curve; NPV, negative predictive value; PPV, positive predictive value.

Table 5 Performance of each prediction model in diagnosing minimally invasive adenocarcinoma and invasive adenocarcinoma for part-solid nodules

	Sensitivity	Specificity	Accuracy	PPV	NPV	AUC
Brock model*	82.3	53.8	79.9	95.1	21.9	0.746 (0.694–0.794)
Lee model†	77.0	69.2	76.4	96.5	21.7	0.771 (0.720–0.817)

Data are in percentages except for AUC. There were 26 preinvasive lesions and 283 invasive lesions (minimally invasive adenocarcinomas and invasive adenocarcinomas).
*A cut-off of 10% predicted probability was used which was suggested by the British Thoracic Society and was optimal at the same time.
†An optimal cut-off of 66.68% was adopted based on the Youden Index.
AUC, area under the receiver operating characteristic curve; NPV, negative predictive value; PPV, positive predictive value.

diagnosis as a reference standard. Indeed, as clinical management strategies differ substantially between preinvasive and invasive lesions, if it can be feasible to predict the invasiveness of SSNs, clinical planning whether to perform annual CT surveillance, limited resection or conventional lobectomy can be facilitated.

The performance of the prediction models for the risk stratification of SSNs was not optimal according to our study results. For PSNs, AUCs of the two models ranged between 0.746 and 0.771 with diagnostic accuracies close to 80%. The performance of the prediction model for PSNs in the original paper by Lee et al[7] was 0.905 (AUC). The study population of the present study was similar to that of the study by Lee et al.[7] An important reason for the performance drop would be the spectrum effect which is a common cause of model performance heterogeneity.[25] A variation in the assessment of CT morphological features (lobulation and spiculation) would be another potential cause. Past research on distinguishing invasive adenocarcinomas appearing as SSNs have reported that logistic regression models built with size metrics, morphological features or texture features showed AUCs ranging from 0.79 to 0.98.[4–6 9 11 13] However, these models were not tested for an independent cohort or validated externally.

Another important finding of our study was that the PPV for the differentiation of invasive lesions among PSNs was very high for both models, over 95%. In other words, the probability of being an invasive lesion was over 95% for nodules predicted as being invasive through these models. A concern, however, is the high false-negative rate of these models. PSNs predicted as preinvasive lesions, which have a low calculated risk, should be managed according to their solid portion size, if they are persistent lesions.[26] A few studies have shown that the solid portions in PSNs are well correlated to the pathological invasive component.[12 27 28] Fleischner Society guideline recommends that PSNs with solid components ≥6mm should be monitored with CT scans at 3–6 months interval.[26] PSNs with solid portions larger than 8mm should be biopsied or surgically resected in consideration of invasive adenocarcinomas.[26] BTS also recommends that the solid component size should be considered to further refine the estimate of malignancy risk.[21] In addition, growing solid component is also a sign of an invasive adenocarcinoma as described in both guidelines.[21 26]

The diagnostic accuracies of both the Brock model and Lee criteria were even lower for pGGNs. Among multiple clinical and radiological characteristics investigated in our study, only three variables (family history of lung cancer, nodule size and nodule count per scan) were significantly different between preinvasive and invasive lesions. This implies the need for other useful features for the development of new better prediction models. Features such as nodule volume, mass or radiomic features may provide additional clues for their differentiation.[4 29] In addition, changes in nodule characteristics at follow-up CT scans, such as an increase in nodule size, attenuation or new development of a solid portion, may also be valuable for the discrimination.[14] Alternatively, computational classification analysis, including deep learning algorithms, which do not require hand-crafted features and can be self-trained directly from raw image pixels, may be another solution for the diagnosis of pGGNs.[30]

The Brock model has been externally validated for the cohorts of the Danish Lung Cancer Screening Trial[19] and National Lung Screening Trial[18]; AUCs for the discrimination of malignant from benign nodules ranged from 0.834 for the former and 0.963 for the latter. AUCs for the validation cohort of the original paper, the British Columbia Cancer Agency chemoprevention trial cohort was 0.970.[16] In addition, for an Australian lung cancer screening cohort, Zhao et al[20] tested the utility of the Brock model for the baseline evaluation of 52 SSNs and demonstrated that the AUC was 0.89. To the contrary, however, the model performance evaluated in our study was lower than those reported in the literature. The main reason for such a discrepancy may be that we included patients who underwent surgical resection of SSNs unlike previous studies. Thus, the proportion of preinvasive lesions was small (16.6%), and a major portion of our study population consisted of invasive lesions (83.4%). Such high prevalence of invasive lesions would have affected our study results. Nevertheless, SSNs of interest in daily clinical practice may be closer to those in our study. In routine practice, transient SSNs, which are definitely benign, do not require risk calculation as they are easily confirmed through follow-up CT scans at short-term intervals.[26] In addition, small SSNs <6mm are usually preinvasive and do not require CT surveillance.

On the other hand, particular concern should be given to persistent SSNs ≥6 mm, especially to those with solid components. As the role of biopsy or positron emission tomography is limited for SSNs,[15] we supposed that risk prediction models may provide value for more appropriate management planning. In this context, we applied prediction models to surgically resected SSNs for the validation of their clinical utility.

There were several limitations to our study. First, our study was not conducted for a screening cohort as described earlier in this manuscript. The prevalence of preinvasive lesions was low compared with that of the screening setting. Thus, the performance measures in this study should be carefully interpreted with respect to the target population which were the incidentally detected surgical candidates. Second, our retrospective study included a small number of patients, and analyses were conducted separately for pGGNs and PSNs. Separate analysis of pGGNs and PSNs has resulted in a slight underestimation of the performance of Brock model. Third, optimal cut-offs for the models were not obtained from ROC analyses of the original study populations from which the models were derived. Fourth, radiological nodule information was extracted from our heterogeneous CT dataset, in which CT acquisition parameters such as radiation dosage, slice thickness or contrast-enhancement were not uniform across the study population. However, these factors would have had little effect on the variables we used. In addition, all CT scans had thin-section images (slice thickness ≤1.5 mm). Fifth, nodule size and solid portion size were measured as the longest transverse diameter in accordance with the definition of lesion size and solid proportion in the original papers. However, recent analyses have revealed that the usage of average diameter as an input variable may enhance the model performance.[31]

In conclusion, the performance of the Brock model and Lee model for the differentiation of invasive lesions among SSNs was suboptimal. In particular, both models showed lower performance for pGGNs compared with that for PSNs. Thus, an alternative approach such as computer-aided classification should be developed for the preoperative diagnosis of invasive lesions among SSNs.

REFERENCES

1. Austin JH, Garg K, Aberle D, et al. Radiologic implications of the 2011 classification of adenocarcinoma of the lung. *Radiology* 2013;266:62–71.
2. Yankelevitz DF, Yip R, Smith JP, et al. CT Screening for Lung Cancer: Nonsolid Nodules in Baseline and Annual Repeat Rounds. *Radiology* 2015;277:555–64.
3. Scholten ET, de Jong PA, de Hoop B, et al. Towards a close computed tomography monitoring approach for screen detected subsolid pulmonary nodules? *Eur Respir J* 2015;45:765–73.
4. Chae HD, Park CM, Park SJ, et al. Computerized texture analysis of persistent part-solid ground-glass nodules: differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas. *Radiology* 2014;273:285–93.
5. Ding H, Shi J, Zhou X, et al. Value of CT characteristics in predicting invasiveness of adenocarcinoma presented as pulmonary ground-glass nodules. *Thorac Cardiovasc Surg* 2017;65:136–41.
6. Jin C, Cao J, Cai Y, et al. A nomogram for predicting the risk of invasive pulmonary adenocarcinoma for patients with solitary peripheral subsolid nodules. *J Thorac Cardiovasc Surg* 2017;153:462–9.
7. Lee SM, Park CM, Goo JM, et al. Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using CT features. *Radiology* 2013;268:265–73.
8. Li Q, Fan L, Cao ET, et al. Quantitative CT analysis of pulmonary pure ground-glass nodule predicts histological invasiveness. *Eur J Radiol* 2017;89:67–71.
9. Liang J, Xu XQ, Xu H, et al. Using the CT features to differentiate invasive pulmonary adenocarcinoma from pre-invasive lesion appearing as pure or mixed ground-glass nodules. *Br J Radiol* 2015;88:20140811.
10. Moon Y, Sung SW, Lee KY, et al. Pure ground-glass opacity on chest computed tomography: predictive factors for invasive adenocarcinoma. *J Thorac Dis* 2016;8:1561–70.
11. Son JY, Lee HY, Kim JH, et al. Quantitative CT analysis of pulmonary ground-glass opacity nodules for distinguishing invasive adenocarcinoma from non-invasive or minimally invasive adenocarcinoma: the added value of using iodine mapping. *Eur Radiol* 2016;26:43–54.
12. Yanagawa M, Johkoh T, Noguchi M, et al. Radiological prediction of tumor invasiveness of lung adenocarcinoma on thin-section CT. *Medicine* 2017;96:e6331.
13. Zhang Y, Shen Y, Qiang JW, et al. HRCT features distinguishing pre-invasive from invasive pulmonary adenocarcinomas appearing as ground-glass nodules. *Eur Radiol* 2016;26:2921–8.
14. Kim H, Park CM, Koh JM, et al. Pulmonary subsolid nodules: what radiologists need to know about the imaging features and management strategy. *Diagn Interv Radiol* 2014;20:47–57.
15. Naidich DP, Bankier AA, MacMahon H, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 2013;266:304–17.
16. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013;369:910–9.
17. 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.

18. White CS, Dharaiya E, Campbell E, *et al*. The vancouver lung cancer risk prediction model: assessment by using a subset of the national lung screening trial cohort. *Radiology* 2017;283:264–72.

19. Winkler Wille MM, van Riel SJ, Saghir Z, *et al*. Predictive accuracy of the pancan lung cancer risk prediction model -external validation based on CT from the Danish Lung Cancer Screening Trial. *Eur Radiol* 2015;25:3093–9.

20. Zhao H, Marshall HM, Yang IA, *et al*. Screen-detected subsolid pulmonary nodules: long-term follow-up and application of the PanCan lung cancer risk prediction model. *Br J Radiol* 2016;89:20160016.

21. 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.

22. Travis WD, Brambilla E, Noguchi M, *et al*. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* 2011;6:244–85.

23. Kim H, Park CM, Song YS, *et al*. Measurement variability of persistent pulmonary subsolid nodules on same-day repeat CT: what is the threshold to determine true nodule growth during follow-up? *PLoS One* 2016;11:e0148853.

24. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.

25. Riley RD, Ensor J, Snell KI, *et al*. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 2016;353:i3140.

26. 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;284:228–43.

27. Ko JP, Suh J, Ibidapo O, *et al*. Lung adenocarcinoma: correlation of quantitative CT findings with pathologic findings. *Radiology* 2016;280:931–9.

28. Lee KH, Goo JM, Park SJ, *et al*. Correlation between the size of the solid component on thin-section CT and the invasive component on pathology in small lung adenocarcinomas manifesting as ground-glass nodules. *J Thorac Oncol* 2014;9:74–82.

29. Song YS, Park CM, Park SJ, *et al*. Volume and mass doubling times of persistent pulmonary subsolid nodules detected in patients without known malignancy. *Radiology* 2014;273:276–84.

30. Esteva A, Kuprel B, Novoa RA, *et al*. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.

31. van Riel SJ, Ciompi F, Jacobs C, *et al*. Malignancy risk estimation of screen-detected nodules at baseline CT: comparison of the PanCan model, Lung-RADS and NCCN guidelines. *Eur Radiol* 2017;27:4019–29.