



Development and Validation of a Risk Assessment Model for Pulmonary Nodules Using Plasma Proteins and Clinical Factors

Anil Vachani, MD, MSCE; Stephen Lam, MD; Pierre P. Massion, MD; James K. Brown, MD; Michael Beggs, PhD; Amanda L. Fish, MBA; Luis Carbonell, MBA; Shan X. Wang, PhD; and Peter J. Mazzone, MD, MPH

BACKGROUND: Deficiencies in risk assessment for patients with pulmonary nodules (PNs) contribute to unnecessary invasive testing and delays in diagnosis.

RESEARCH QUESTION: What is the accuracy of a novel PN risk model that includes plasma proteins and clinical factors? How does the accuracy compare with that of an established risk model?

STUDY DESIGN AND METHODS: Based on technology using magnetic nanosensors, assays were developed with seven plasma proteins. In a training cohort (n = 429), machine learning approaches were used to identify an optimal algorithm that subsequently was evaluated in a validation cohort (n = 489), and its performance was compared with the Mayo Clinic model.

RESULTS: In the training set, we identified a support vector machine algorithm that included the seven plasma proteins and six clinical factors that demonstrated an area under the receiver operating characteristic curve of 0.87 and met other selection criteria. The resulting risk reclassification model (RRM) was used to recategorize patients with a pretest risk of between 10% and 84%, and its performance was assessed across five risk strata (low, $\leq 10\%$; moderate, 10%-34%; intermediate, 35%-70%; high, 71%-84%; very high, $> 85\%$). Stratification by the RRM decreased the proportion of intermediate-risk patients from 26.7% to 10.8% ($P < .001$) and increased the low-risk and high-risk strata from 16.8% to 21.9% ($P < .001$) and from 3.7% to 12.1% ($P < .001$), respectively. Among patients classified as low risk by the RRM and Mayo Clinic model, the corresponding true-negative to false-negative ratios were 16.8 and 19.5, respectively. Among patients classified as very high risk by the RRM and Mayo Clinic model, the corresponding true-positive to false-positive ratios were 28.5 and 17.0, respectively. Compared with the Mayo Clinic model, the RRM provided higher specificity at the low-risk threshold and higher sensitivity at the very high-risk threshold.

INTERPRETATION: The RRM accurately reclassified some patients into low-risk and very high-risk categories, suggesting the potential to improve PN risk assessment.

CHEST 2023; 163(4):966-976

KEY WORDS: biomarkers; diagnosis; lung cancer; pulmonary nodules, prediction

ABBREVIATIONS: AUC = area under the receiver operating characteristic curve; IPN = indeterminate pulmonary nodule; LR = low risk; MR = moderate risk; NPV = negative predictive value; PN = pulmonary nodule; PPV = positive predictive value; RRM = risk reclassification model; SVM = support vector machine; VHR = very high risk

AFFILIATIONS: From the Pulmonary, Allergy, and Critical Care Division (A. V.), Department of Medicine, University of Pennsylvania, the Corporal Michael J. Crescenz VA Medical Center (A. V.), Department of Medicine, Philadelphia, PA, the Division of Allergy, Pulmonary and Critical Care Medicine (P. P. M.), Vanderbilt University, Nashville, TN, the Division of Pulmonary, Critical Care,

Take-home Points

Research Question: What is the accuracy of a novel pulmonary nodule (PN) risk model that includes plasma proteins and clinical factors? How does the accuracy compare with that of an established risk model?

Results: A support vector machine (SVM) algorithm that included seven plasma proteins and six clinical factors decreased the proportion of intermediate-risk patients and increased the proportion within low-risk (< 10%) and high-risk (\geq 70%) strata, respectively. Compared with the Mayo Clinic model, the SVM algorithm provided higher specificity at the low-risk (< 10%) threshold and higher sensitivity at the very high-risk (\geq 85%) threshold.

Interpretation: An SVM algorithm accurately reclassified some patients into low- and very high-risk categories, suggesting the potential to improve PN risk assessment.

Pulmonary nodules (PNs), often detected on chest CT scan imaging, represent a frequent diagnostic problem in clinical practice.¹ PNs are most often identified incidentally on CT scan imaging performed for other indications, such as the evaluation of respiratory symptoms.^{2,3} Within the context of lung cancer screening, PNs are identified in 20% to 40% of scans, of which roughly half require further evaluation to

distinguish the small percentage of malignant nodules (approximately 4%) from most of the detected nodules deemed benign.^{4,5} Management of PNs is based on the probability that they are malignant. However, uncertainty in establishing the probability of cancer for any given PN impacts the goal of expediting care for patients with a malignant nodule while minimizing testing for an individual with a benign nodule.

Risk assessment tools to estimate the probability of cancer in patients with PNs based on imaging parameters and other risk factors, such as the Mayo Clinic model, are widely available and have undergone extensive validation to assess their performance.^{6,7} Although guidelines recommend that subsequent management decisions after nodule identification should be based on an assessment of cancer risk, invasive sampling of low-risk nodules and surgical resection of benign nodules remain common, suggesting that additional risk assessment tools are needed.^{1,8}

The development of robust molecular biomarkers to risk-stratify patients with PNs better could be particularly useful in distinguishing patients who warrant further surveillance from those who would benefit from an invasive diagnostic procedure.⁹ In this study, we developed and validated a combined clinical and blood-based risk assessment tool with the goal of improving risk assessment by focusing on patients with PNs in whom the probability of malignancy is between 10% and 84%.

Study Design and Methods

Study Design and Participants

We conducted a multicenter, retrospective diagnostic study using a training validation design with banked plasma samples from patients with benign or malignant PNs collected from five geographically diverse centers across the United States and Canada: British Columbia Cancer Research Institute, Cleveland Clinic, San Francisco Veterans Affairs Medical Center, the University of Pennsylvania, and Vanderbilt University. The

study population included a mix of patients with screening-identified and incidentally identified nodules: the University of Pennsylvania (training and validation set) included patients with incidental nodules, San Francisco Veterans Affairs Medical Center and Cleveland Clinic included high-risk patients undergoing screening, and Vanderbilt University and the British Columbia Cancer Research Institute included a mix of patients with screening-detected and incidentally detected nodules.

The plasma samples were obtained from whole blood collected within 60 days of imaging that identified the nodule at four of five centers; these data were not recorded at one center. Inclusion criteria included age older than 18 years, an indeterminate (noncalcified, not known to be stable on imaging) pulmonary nodule 4 to 30 mm in diameter, pathology-confirmed malignant diagnosis, or confirmation of benign diagnosis by pathologic examination or surveillance imaging demonstrating stability or resolution over 2 years. Exclusion criteria included evidence of metastatic disease, previous nodule biopsy, diagnosis of any cancer within 5 years of nodule detection (except for nonmelanoma skin cancer), or receipt of any blood products within 30 days of enrollment. Study protocols were approved at the individual institutions; all patients provided informed consent.

Allergy and Sleep Medicine (J. K. B.), Department of Medicine, University of California, San Francisco, the VA Medical Center San Francisco (J. K. B.), Department of Medicine, San Francisco, Mag-Array, Inc. (M. B., A. L. F., L. C., and S. X. W.), Milpitas, CA, the Respiratory Institute (P. J. M.), Cleveland Clinic, Cleveland, OH, and the Department of Integrative Oncology (S. L.), British Columbia Cancer Research Institute, University of British Columbia, Vancouver, BC, Canada.

CORRESPONDENCE TO: Anil Vachani, MD, MSCE; email: avachani@pennmedicine.upenn.edu

Copyright © 2022 American College of Chest Physicians. Published by Elsevier Inc. All rights reserved.

DOI: <https://doi.org/10.1016/j.chest.2022.10.038>

Assay Technology

The magnetic nanosensor technology and instruments, developed by MagArray, Inc., are based on magnetoresistive biosensor chips capable of detecting low-level biomolecules in a multiplexed format.¹⁰ Multiple quantitative protein assays were combined into one multiplexed immunoassay. The immunoassays used commercial antibodies and were calibrated with multiplexed recombinant human protein standards. All samples were run in duplicate and included two quality control human plasma samples for assay validity. Acceptable analytical performance was defined as total assay imprecision of $\leq 15\%$.

Identification of Protein Biomarker Candidates

A review of the literature identified 59 protein biomarker candidates, of which 41 were predicted to be detectable in plasma. Antibody screening with the nanosensor technology identified 26 proteins accurately measurable in human plasma. The list was refined further to biomarkers with differentiating levels between the plasma from individuals with benign vs malignant PNs in a subset of the training cohort. Seven proteomic biomarkers (carcinoembryonic antigen, epidermal growth factor receptor, neutrophil activating protein 2, prosurfactant protein B, C-X-C motif chemokine ligand 10, receptor for advanced glycation end products, and tissue inhibitor of metalloproteinases 1) were selected with preliminary evidence of discrimination and low cross-reactivity.

Model Development Using Training Cohort

Natural log-transformed plasma concentrations of the seven biomarkers and 10 clinical risk factors available for all training cohort patients (age, sex, smoking history [current, previous, never], pack-years, years smoked, number of years since quitting, cancer history > 5 years, nodule diameter, lobar location, and spiculation) were evaluated individually and in combination using regression modeling.¹¹ Four machine learning feature selection methods were used to identify an optimal model: binomial error distribution generalized linear models,¹² boosted generalized linear model and boosted tree random forest algorithms,¹³ and radial or linear kernel support vector machines (SVMs).¹⁴ The performance of approximately 1,000 models, fit using these four methods in a random split of the training cohort into a development set (80%) and test set (20%), were ranked to identify the fit method and features that provided improvements in discrimination (assessed using contingency tables, area under the receiver operating characteristic curves [AUCs]), estimated negative predictive value (NPV) and positive predictive value (PPV) at 25% cancer prevalence, and robustness of the risk score to biomarker assay variability. The initial target performance criteria were: (1) an AUC of ≥ 0.80 and a positive AUC difference of ≥ 0.05 vs the Mayo Clinic model, (2) sensitivity of $\geq 80\%$ and specificity of $\geq 40\%$ as recommended by a recent biomarker development policy statement,¹⁵ and (3) estimated NPV and PPV of $\geq 90\%$ and $\geq 40\%$, respectively, at 25% cancer prevalence.¹⁶ Additionally, target model volatility to the control sample biomarker score variations was $\leq \pm 5$ score units.

For model optimization, the top scoring algorithm was refit within the training cohort using 5-fold cross validation. Calibration by a generalized linear model fit of the cohort posterior probability to the malignancy prevalence was used to provide malignancy probabilities.

Risk Reclassification Model

A final algorithm (risk reclassification model [RRM]) was developed to estimate the posttest probability in the validation cohort. First, the method applies four cut points to separate the cohort into five risk groups based on the Mayo Clinic model: low

risk (LR; < 10%), moderate risk (MR; 10%-34%), intermediate risk (35%-69%), high risk (70%-84%), and very high risk (VHR; $\geq 85\%$). Second, the RRM uses the SVM score to provide risk estimates only for individuals with a Mayo Clinic model pretest risk of between 10% and 84% (Fig 1); individuals in the two extreme categories (LR and VHR) do not undergo recategorization. Patients in the intermediate risk category are reclassified liberally based on the SVM score (Fig 1). Based on preliminary analysis in the training cohort, full reclassification (up or down) of patients in either the MR or high-risk categories did not achieve sufficient accuracy, and so the final approach selected allowed only limited reclassification based on the SVM score; patients either remain in their original category or can be reclassified to a one-step decrease or increase in risk category, respectively.

Assessment of Model Performance in the Validation Cohort

We conducted a risk stratum-specific analysis by calculating the observed prevalence of lung cancer within each of the five risk categories using patient-level estimates derived from the Mayo Clinic model and the RRM. To estimate the benefit to harm ratio of the RRM model, we calculated the ratios of the true-positive to false-positive samples within the LR and MR strata and true-negative to false-negative samples within the high-risk and VHR strata. To estimate the predicted probability of cancer for each risk stratum at a population prevalence of 25%, the stratum-specific likelihood ratios were used to convert pretest odds of malignancy to posttest probabilities. Model discrimination was assessed further by calculation of the AUC within the validation cohort and by comparing the sensitivity and specificity and calculating at the two extreme thresholds (< 10% and > 85%) for the RRM and the Mayo Clinic model. NPV and PPV were calculated at the observed prevalence and at a cancer prevalence of 25%. A subgroup analysis determined the sensitivity, specificity, NPV, and PPV for the validation cohort limited to patients with stage I or II disease. Calibration was assessed by comparing the expected prevalence by quintiles with the observed cancer prevalence based on categorization using either the RRM or the Mayo Clinic model.

Statistical Analyses

For assessment of model accuracy within the validation cohort, sample size requirements were estimated based on a cancer prevalence of 0.50 and a desired width of the 95% CI of sensitivity and specificity of no more than 0.10. Given that sensitivity and specificity in the analysis would be assessed at two different thresholds (0.10 and 0.85), sample size was calculated at a target of 0.50 for both sensitivity and specificity, resulting in a minimum sample of 193 cases and 193 control subjects. Although no formal power calculation for model development in the training cohort was performed, it was estimated that the training cohort sample size should include at least 10 events per prediction parameter. Because 17 variables were considered, we estimated needing a minimum of 170 cancers (and at least 170 benign nodules).

Statistical and modeling analyses were carried out using Microsoft Excel version 16 (Microsoft) and R version 4.0.3 or higher software (R Foundation for Statistical Computing) with the following packages from the Comprehensive R Archive Network¹⁷: caret (version 6.0-86), e1071 (version 1.7-4), ggplot2 (version 3.3.5), glmnet (version 4.0-2), OptimalCutpoints (version 1.1-4), ppsr (version 0.0.2), pROC (version 1.10.0), randomForest (version 4.6-14), rms (version 6.2-0), rpart (version 4.1-15), and stats (version 4.0.3).

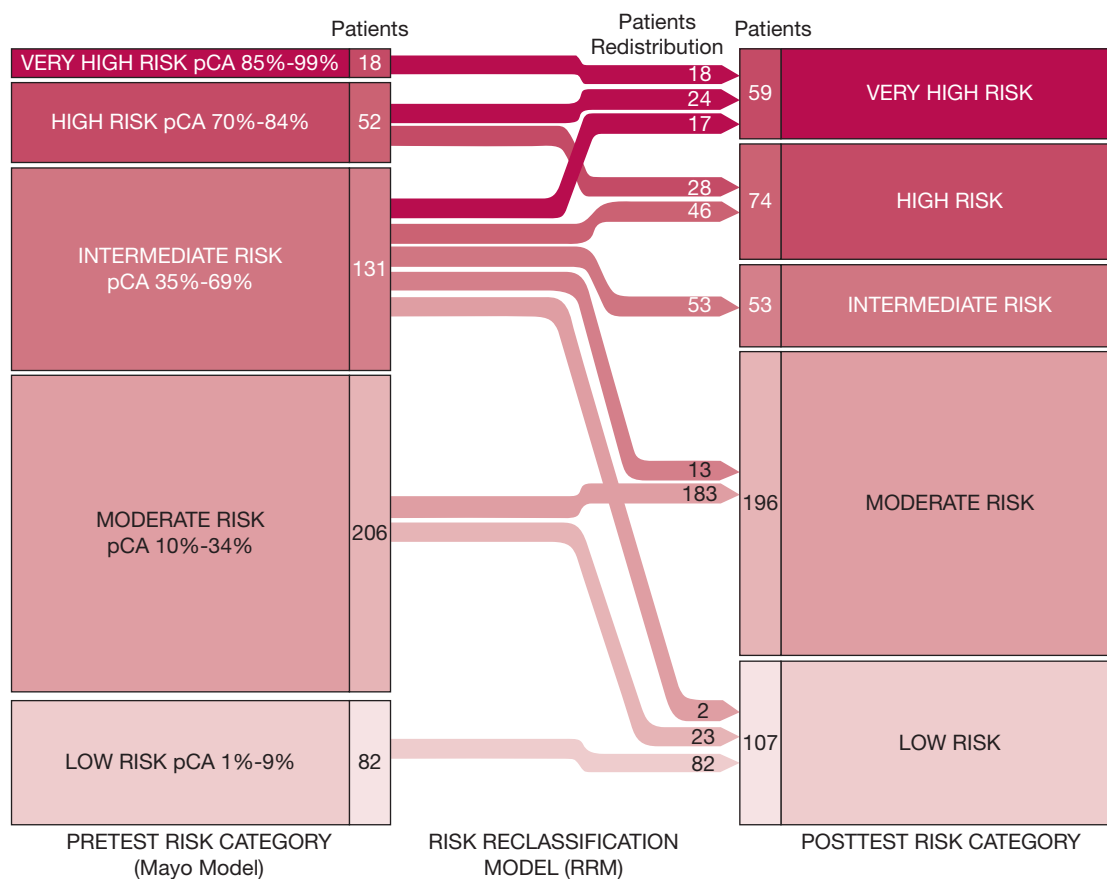


Figure 1 – Diagram showing the redistribution of validation cohort patients using the RRM. Process of reclassification of patients in the validation cohort based on RRM results based on pretest risk assessed by the Mayo Clinic model. Only patients with pretest risk between 10% and 84% were reclassified by the RRM. The numbers of patients by risk category before and after RRM categories (shown in columns) and the number of patients who were redistributed by the RRM (embedded within arrows) are provided. pCA = probability of cancer; RRM = risk reclassification model.

Differences in frequency and distribution of validation cohort patients among the five risk categories by the Mayo Clinic model and the RRM were assessed using the McNemar’s test for marginal homogeneity.

Estimates of diagnostic accuracy for the Mayo Clinic model and RRM at the 10% and 85% risk thresholds were compared using McNemar’s test.

Results

Study Population

The training cohort included 429 patients with a 4- to 30-mm PN found either incidentally or through screening from San Francisco Veterans Affairs Medical Center, University of Pennsylvania, and Vanderbilt University (Table 1, e-Table 1). The three sites showed a similar proportion of patients with cancer with adenocarcinoma histologic findings. Most cancers were stage I or II (79%), and the overall prevalence of lung cancer was 43% (186 of 429). In total, 20% of patients never smoked, 28% were currently smoking, and 52% previously smoked. Malignant PNs were larger than benign PNs (mean diameter, 17.5 mm vs 12.2 mm; $P < .001$). Across both training and validation cohorts, 376

of 520 patients (72%) with nonmalignant findings received a diagnosis of a benign lesion based on nodule stability or resolution over 2 years; the remaining lesions were diagnosed based on pathologic confirmation.

The validation cohort included 489 patients with a 4- to 30-mm PN found incidentally or via lung cancer screening from the British Columbia Cancer Research Institute, Cleveland Clinic, and the University of Pennsylvania (Table 1, e-Table 2). Samples from the University of Pennsylvania were distinct from those included in the training cohort. Similar to the training cohort, the validation cohort showed a lung cancer prevalence of 43% (212 of 489), of which 65% showed adenocarcinoma histologic findings and 70% were stage I or II disease.

TABLE 1] Patient Characteristics

Variable	Training		Validation	
	Benign (n = 243)	Malignant (n = 186)	Benign (n = 277)	Malignant (n = 212)
Age, y	62.7 ± 11.4	65.7 ± 9.15	65.6 ± 7.23	69.3 ± 8.5
Sex, female	83 (34.2)	82 (44.1)	129 (46.6)	120 (56.6)
Race or ethnicity				
White	216 (88.9)	159 (85.5)	249 (89.9)	172 (81.1)
Black	22 (9.1)	22 (11.8)	11 (4.0)	12 (5.7)
Asian	2 (0.8)	2 (1.1)	12 (4.3)	23 (10.8)
Other	2 (0.8)	2 (1.1)	2 (0.7)	3 (1.4)
Missing	1 (0.2)	1 (0.5)	3 (1.1)	2 (0.9)
Cancer history	57 (23.5)	77 (41.4)	17 (6.1)	16 (7.5)
Smoking History				
Current	57 (23.5)	52 (28.0)	105 (37.9)	46 (21.7)
Former	129 (53.1)	97 (52.2)	166 (59.9)	138 (65.1)
Never	57 (23.5)	37 (19.9)	6 (2.2)	28 (13.2)
Pack-years	38.5 ± 39.5	41.1 ± 36.3	42.2 ± 21.3	34.4 ± 28.8
Nodule size, mm	12.2 ± 6.5	17.5 ± 6.3	11.0 ± 6.0	19.8 ± 6.2
Nodule located in upper lobe	107 (44.0)	110 (59.1)	148 (53.4)	134 (63.2)
Nodule spiculation	42 (17.3)	92 (49.5)	19 (6.9)	100 (47.2)
Cancer histology				
Adenocarcinoma	...	139 (74.7)	...	137 (64.7)
Squamous cell carcinoma	...	35 (18.8)	...	39 (18.4)
NSCLC	...	9 (4.8)	...	10 (4.7)
Other	...	2 (0.0)	...	13 (6.1)
Unknown	...	1 (0.0)	...	13 (6.1)
Cancer stage				
I or II	...	147 (79.0)	...	147 (69.3)
III or IV	...	18 (9.7)	...	58 (27.4)
Unknown	...	21 (11.3)	...	7 (3.3)

Data are presented as No. (%) or mean ± SD. NSCLC = non-small cell lung cancer.

Model Derivation

We identified 14 machine learning models that provided the highest relative AUC in the training cohort (AUC, 0.87) (e-Table 3). An SVM algorithm fit using a radial kernel of seven biomarkers (carcinoembryonic antigen, epidermal growth factor receptor, neutrophil activating protein 2, prosurfactant protein B, C-X-C motif chemokine ligand 10, receptor for advanced glycation end products, and tissue inhibitor of metalloproteinases 1) and six clinical factors (age, sex, pack-years, nodule size, nodule spiculation, and nodule location) produced the lowest volatility and was the single SVM model selected for incorporation into the RRM and for further validation (model 1 in e-Table 3).

Model Performance

The AUCs of the RRM and the Mayo Clinic model in the validation cohort were 0.85 (95% CI, 0.80-0.90) and 0.87 (95% CI, 0.82-0.92), respectively (Fig 2A). Sensitivity and specificity plots in the validation cohort show that the RRM, in comparison with the Mayo Clinic model, was more sensitive and less specific at the higher model thresholds and more specific and less sensitive at lower thresholds (Fig 2B).

The RRM was used to reclassify validation cohort patients with a Mayo Clinic model risk of 10% to 84% (e-Table 4), resulting in alterations in the frequency and distribution of the patients in the validation cohort within the five risk categories when compared with a

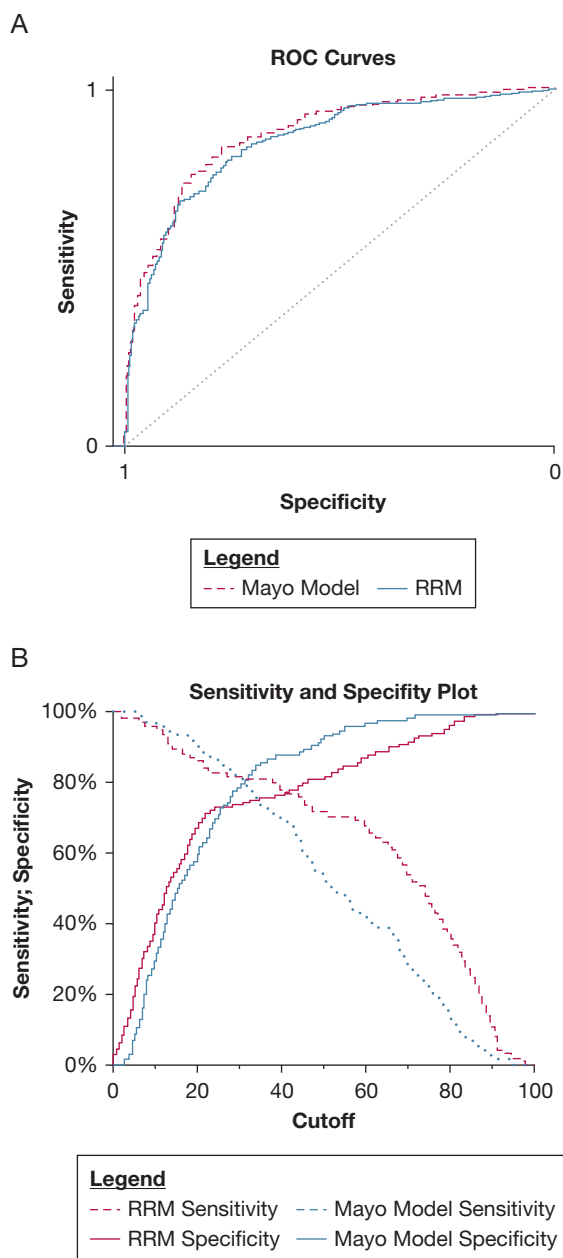


Figure 2 – Line graphs showing the discrimination of pulmonary nodules with the RRM and Mayo Clinic model. A, ROC curves for the Mayo Clinic model (dashed line) and RRM (solid line). Mayo Clinic model area under the ROC curve, 0.87 (95% CI, 0.82-0.92); RRM area under the ROC curve, 0.85 (95% CI, 0.80-0.90). B, Sensitivity and specificity plot across the full range of model cutoffs. Cutoff (x-axis) is the threshold from the model used to define the classification of benign (\leq the cutoff) and malignant ($>$ cutoff) nodules. ROC = receiver operating characteristic.

pretest assessment using the Mayo Clinic model (Fig 1). Among patients in the pretest intermediate risk group, the RRM recategorized 78 patients (60 with malignant findings, 18 with benign findings), resulting in a decrease in the overall proportion of this category from 26.7% to 10.8% ($P < .001$) (Fig 3). The LR category

increased by 25 patients (two with malignant findings, 23 with benign findings; true-negative to false-negative ratio, 11.5), increasing the overall proportion of this category from 16.8% to 21.9% ($P < .001$) (Fig 3). The VHR category increased by 41 patients (40 with malignant findings, one with benign findings; true-positive to false-positive ratio, 40.0), increasing the overall proportion of this category from 3.7% to 12.1% ($P < .001$).

Table 2 summarizes the cancer probabilities observed in the five risk strata based on classification by the Mayo Clinic model and RRM. At both the current cancer prevalence and for estimates at a prevalence of 25%, this approach yields nearly equivalent cancer probabilities for the Mayo Clinic model and RRM for the LR stratum, whereas the RRM provides a slightly higher estimated probability in VHR stratum. Within the LR stratum, the true-negative to false-negative ratios were 16.8 and 19.5 for the RRM and Mayo Clinic model, respectively. The corresponding true-positive to false-positive ratios for the VHR stratum were 28.5 and 17.0 for the RRM and Mayo Clinic model, respectively.

Among the 89 validation cohort patients with 4- to $<$ 8-mm nodules, five patients had malignant findings (5.6% prevalence); 52 patients (58%) had a Mayo Clinic risk score of $<$ 10% and 37 patients (42%) had a Mayo Clinic risk score of 10% to 36%. Of these 37 patients, the RRM reclassified 21 patients downwards to LR, all of whom were proven to have benign nodules. The remaining 16 patients were classified as MR by RRM, of whom 14 were proven to have benign nodules and two to have lung cancer.

Using a threshold-based approach to examine the extreme strata (LR and VHR), the RRM provided higher specificity (36% vs 28%; $P < .001$) while achieving a similar level of sensitivity at the 10% risk threshold compared with the Mayo Clinic model (Table 3). At the 85% risk threshold, the RRM results in higher sensitivity (27% vs 8%, at a similar level of specificity; $P < .001$) (Table 3). In the subgroup analysis limited to patients with stage I or II disease (Table 3), estimates of accuracy (sensitivity, specificity, NPV, and PPV) were similar to the main analysis.

To examine calibration, we compared expected vs observed frequencies of cancer across quintiles of predicted probability. The theoretical expected probability of malignancy is represented by the midpoint of each quintile (10%, 30%, 50%, 70%, and 90%). The

		Benign				Malignant						
RRM pCA	LR	22	1	0	23	1	1	0	2	LR		
	MR	134	5	0	139	49	8	0	57	MR		
	IR	0	19	0	19	0	34	0	34	IR		
	HR	0	12	4	16	0	34	24	58	HR		
	VHR	0	0	1	1	0	17	23	40	VHR		
		156	37	5	Total	50	94	47	Total			
		MR	IR	HR		MR	IR	HR	0			
		10-34%	35-69%	70-84%		10-34%	35-69%	70-84%				
		Mayo Model pCA				Mayo Model pCA						

Figure 3 – Reclassification by RRM of patients with Mayo Clinic Model Risk of 10% to 84%. RRM recategorized 41 patients (40 with malignant findings and one with benign findings) to VHR (dark red) and 25 patients (23 with benign findings and two with malignant findings) to LR (light red). HR = high risk; IR = intermediate risk; LR = low risk; MR = moderate risk; pCA = probability of cancer; RRM = risk reclassification; VHR = very high risk.

observed malignancy proportion was calculated for all patients who fell within each quintile range. The RRM classification aligned more closely with the expected malignancy risk compared with the Mayo Clinic model, which underestimated the risks of malignancy in the middle and high quintiles (Fig 4).

Discussion

In this study, we derived and validated a combined clinical and plasma protein-based lung cancer risk prediction model that improved the estimation of lung cancer risk for PNs. The RRM functions by reclassifying patients with a pretest risk of 10% to 84% as assessed by the Mayo Clinic model. This approach results in a favorable redistribution by decreasing the intermediate risk category by 60% of patients while increasing the proportion identified as LR and VHR by 30% and 228%, respectively. The reclassification into these two categories was highly accurate; of 25 patients reclassified to LR, 23 had benign disease, and of 41 patients reclassified to VHR, 40 had malignant disease. This level of performance in reclassification has the potential to provide tangible results aligning accurate risk assessment with appropriate management decisions. Patients with an intermediate risk of malignancy traditionally have been the most challenging group to manage clinically given the inherent challenges in balancing the desire to avoid invasive procedures in patients with benign disease and to expedite a timely diagnosis in patients with lung cancer. Current guidelines suggest further evaluation with PET and CT scan imaging or biopsy (transthoracic needle aspiration or bronchoscopy) in

patients in this risk range.^{1,8} The performance of PET and CT scan imaging is reasonable for solid nodules, but false-positive results (mostly commonly because of active infection) and false-negative results (eg, subcentimeter tumors, adenocarcinoma of the lepidic subtype) can occur, and the overall accuracy of PET and CT scan imaging is limited in certain geographic regions, such as in areas with high rates of endemic fungal infections.¹⁸ Intermediate risk PNs account for a large proportion of invasive biopsy procedures among patients with benign diagnoses. Biopsy procedures most frequently used in patients with intermediate risk nodules, transthoracic needle aspiration and bronchoscopy, can have a high burden of complications, exposing these patients to unnecessary risk.¹⁹ Strategies to assess risk that more closely align the probability of malignancy with the appropriate management strategy are likely to improve cancer outcomes, to minimize harms, and to yield higher levels of patient satisfaction.

Although the RRM provides favorable reclassification of patients with a Mayo Clinic model risk probability of 10% to 84%, we noted that the overall AUCs of the RRM and Mayo Clinic model were similar in the validation cohort. Notably, the AUC of the Mayo Clinic model was considerably higher in the validation cohort than in the training cohort (AUC, 0.87 vs 0.78), a level of discrimination considerably higher than what has been observed in prior validation assessments of the Mayo Clinic model.^{6,20,21} This suggests that the clinical features used in the Mayo Clinic model exerted greater influence on discrimination in the current validation cohort, which may have limited the contribution of the

TABLE 2] Estimates of Benefit to Harm and Cancer Probability Across Risk Strata in Full Validation Cohort

Variable	Risk Strata									
	LR (1%-9%)		MR (10%-34%)		IR (35%-69%)		HR (70%-84%)		VHR (85%-100%)	
	Benign	Malignant	Benign	Malignant	Benign	Malignant	Benign	Malignant	Benign	Malignant
Mayo Clinic model, <i>n</i>	78	4	156	50	37	94	5	47	1	17
RRM, <i>n</i>	101	6	139	57	19	34	16	58	2	57
Benefit to harm analyses ^a	TN to FN Ratio				...		TP to FP Ratio			
Mayo Clinic model	19.5		3.1				9.4		17.0	
RRM	16.8		2.4				3.6		28.5	
Cancer probability ^b										
Mayo Clinic model	4.9 (4 of 82)		24.3 (50 of 206)		71.8 (94 of 131)		90.4 (47 of 52)		94.4 (17 of 18)	
RRM	5.9 (6 of 101)		29.1 (57 of 196)		64.2 (34 of 53)		78.4 (58 of 74)		96.6 (57 of 59)	
Likelihood ratio (stratum specific)										
Mayo Clinic model	0.07		0.42		3.32		12.28		22.21	
RRM	0.08		0.54		2.34		4.74		37.24	
Estimated cancer probability, % ^b										
Mayo Clinic model	2.2		12.2		52.5		80.4		88.1	
RRM	2.5		15.2		43.8		61.2		92.5	

Data are presented as percentage (No./total No.), unless otherwise indicated. FN = false-negative; FP = false-positive; HR = high-risk; IR = intermediate risk; LR = low risk; MR = moderate risk; RRM = risk reclassification model; TN = true-negative; TP = true-positive; VHR = very high risk.

^aBenefit to harm analysis performed by calculation of within strata of TN to FN ratio for LR and MR categories and TP to FP ratio for HR and VHR categories.

^bCancer probability within strata estimated at 25% prevalence.

plasma proteins in the RRM. That the RRM did not fluctuate together with the Mayo Clinic model between the training and validation sets suggests that the RRM provided independent information from the Mayo Clinic model.

Although 95% of patients with malignant nodules either remained in the original risk category or were reclassified into a higher level of risk, 10 discordant results occurred in which the RRM classified patients into a lower risk group compared with the Mayo Clinic model. Two cancers were classified as LR and eight cancers were categorized as MR by the RRM. These findings substantiate the need to consider a patient's comprehensive clinical presentation when assessing risk and the importance of close imaging surveillance among patients with intermediate risk PNs that do not go on to an invasive strategy. The RRM also will require prospective studies to validate the test performance

further and to assess the impact of the model on the rate of invasive procedures in patients with benign nodules and its impact on time to diagnosis and treatment for patients with malignant nodules.

Our analysis used five risk categories, although two existing guidelines from the American College of Chest Physicians and the British Thoracic Society for incidental nodules have suggested thresholds of 5% or 10% to distinguish between LR and intermediate risk and 65% or 70% to differentiate between intermediate-risk and HR nodules. Although these guidelines are well established, we used a more granular approach for two reasons. First the HR threshold of 65% or 70% was conceived as a level above which surgical biopsy and resection or empiric radiation treatment would be appropriate. However, given the concerns that have emerged regarding the high rate of surgery for benign disease, we incorporated a VHR category that could

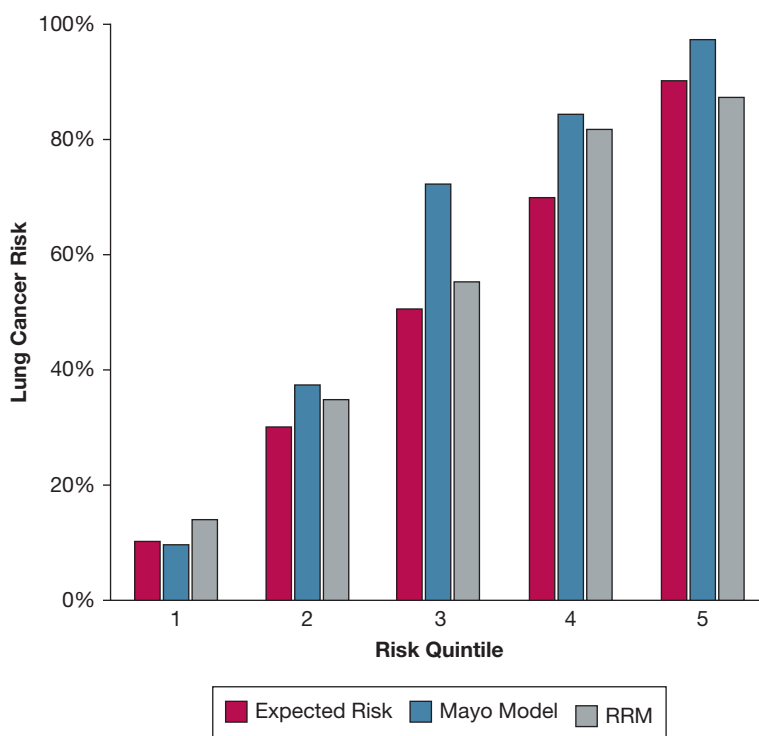
TABLE 3] Diagnostic Performance of RRM and Mayo Clinic Model Using Risk Thresholds

Variable	Risk Threshold							
	10%				85%			
	All Patients (n = 489)		Patients With Stage I or II Disease (n = 431)		All Patients (n = 489)		Patients With Stage I or II Disease (n = 431)	
	RRM	Mayo Clinic	RRM	Mayo Clinic	RRM	Mayo Clinic	RRM	Mayo Clinic
True positive	206	208	150	152	57	17	44	13
False positive	176	199	176	199	2	1	2	1
True negative	101	78	101	78	275	276	275	276
False negative	6	4	4	2	155	195	110	141
Sensitivity (95% CI)	97 (95-100)	98 (96-100)	97 (95-100)	99 (97-100)	27 (17-36)	8 (0-19)	29 (18-40)	8 (0-21)
Specificity (95% CI)	36 (29-44)	28 (20-36)	36 (29-44)	28 (20-36)	99 (97-100)	100 (97-100)	99 (97-100)	100 (97-100)
PPV (95% CI) ^a	54 (46-62)	51 (42-60)	46 (38-54)	43 (34-52)	97 (93-100)	94 (86-100)	96 (91-100)	93 (82-100)
NPV (95% CI) ^a	94 (91-98)	95 (91-99)	96 (93-99)	98 (95-100)	64 (54-74)	59 (40-78)	71 (60-82)	66 (45-87)
PPV (95% CI) ^b	35 (27-42)	32 (24-41)	34 (26-41)	31 (23-40)	93 (88-99)	88 (76-100)	93 (87-99)	89 (75-100)
NPV (95% CI) ^b	97 (95-100)	97 (93-100)	98 (95-100)	98 (96-100)	81 (76-90)	78 (62-94)	81 (75-90)	77 (64-95)

Data are presented as No. or value (95% CI), unless otherwise indicated. RRM = risk reclassification model.

^aCalculated at observed cancer prevalence of 43%.

^bCalculated at cancer prevalence of 25%.



Risk Score Range	Risk Quintile	Cancer Risk	Mayo Model				RRM			
			Total	Benign	Malignant	Prevalence	Total	Benign	Malignant	Prevalence
1-19	1	10%	176	160	16	9%	196	169	27	14%
20-39	2	30%	131	83	48	37%	110	72	38	35%
40-59	3	50%	82	23	59	72%	29	13	16	55%
60-79	4	70%	64	10	54	84%	61	11	50	82%
80-99	5	90%	36	1	35	97%	93	12	81	87%
TOTAL			489	277	212		489	277	212	

Figure 4 – Bar graph showing the model calibration in the validation cohort by risk quintiles ($n = 489$). Diagonal pattern indicates expected risk, light gray indicates Mayo Clinic model, and dark gray indicates the RRM. RRM = risk reclassification model.

minimize potential harms from surgery or radiotherapy for benign disease. Of note, the PPV of 97% at the HR threshold suggests that inappropriate surgery could be limited. Second, since the publication of the American College of Chest Physicians and British Thoracic Society guidelines, advances in our understanding of nodule management have led many to believe that defining management into three buckets requires refinement. For examples, the American College of Radiology Lung-RADS has six categories (1, 2, 3, 4A, 4B, and 4X) to guide shared decision-making. Thus, strategies with > 3 categories, such as those proposed in this study, may facilitate subsequent decision-making.

This study has limitations. First, this was a retrospective validation study that relied on banked plasma samples collected across five institutions. This design can result in certain biases in which patients were identified for inclusion in the study. For example, we observed a cancer prevalence of 43% in both the training and validation sets, which is higher than the prevalence observed in broader clinical settings. We partly account for this by estimating performance of the models at both the observed prevalence and at a prevalence of 25%. Second, full reclassification did not achieve sufficient accuracy in the MR or HR categories to allow for this strategy to be evaluated in the validation cohort,

ultimately limiting the overall impact of the current model. Additional biomarker discovery that can be incorporated into the model may allow for future improvements in performance. Finally, we were unable to perform any comparisons with functional imaging because PET and CT scans for nodule evaluation were limited in our study population. When performed, PET and CT scan timing varied considerably relative to time point of the index CT scan, blood collection, or both.

Interpretation

The RRM improves the assessment of PNs with a pretest risk of 10% to 84% by reclassifying patients within this range into LR and VHR groups, suggesting the potential to improve PN risk prediction and to provide better alignment to subsequent management decisions. These results warrant further investigation in prospective studies to confirm the RRM's performance and assessment of appropriate utility end points.

Acknowledgments

Author contributions: A. V. had full access to all data in the study and takes responsibility for the content of manuscript, including the data and analysis. S. L., P. P. M., J. K. B., M. B., A. L. F., L. C., S. X. W., and P. J. M. contributed substantially to the study design, data analysis and interpretation, and writing of the manuscript.

Role of sponsors: MagArray, Inc., provided financial support for the work reported.

Additional information: Coauthor Pierre P. Massion, MD, passed away on April 1, 2021. The e-Tables are available online under "Supplementary Data."

References

- 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(5 suppl):e93S-e120S.
- Gould MK, Tang T, Liu I-LA, et al. Recent trends in the identification of incidental pulmonary nodules. *Am J Respir Crit Care Med*. 2015;192(10):1208-1214.
- Vachani A, Zheng C, Liu I-LA, Huang BZ, Osuji TA, Gould MK. The probability of lung cancer in patients with incidentally detected pulmonary nodules: clinical characteristics and accuracy of prediction models. *Chest*. 2022;161(2):562-571.
- Aberle DR, DeMello S, Berg CD, et al. Results of the two incidence screenings in

the National Lung Screening Trial. *N Engl J Med*. 2013;369(10):920-931.

- Mazzone PJ, Lam L. Evaluating the patient with a pulmonary nodule: a review. *JAMA*. 2022;327(3):264-273.
- Choi HK, Ghobrial M, Mazzone PJ. Models to estimate the probability of malignancy in patients with pulmonary nodules. *Ann Am Thorac Soc*. 2018;15(10):1117-1126.
- 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(8):849-855.
- Baldwin DR, Callister MEJ; Guideline Development Group. The British Thoracic Society guidelines on the investigation and management of pulmonary nodules. *Thorax*. 2015;70(8):794-798.
- Seijo LM, Peled N, Ajona D, et al. Biomarkers in lung cancer screening: achievements, promises, and challenges. *J Thorac Oncol*. 2019;14(3):343-357.
- Gaster RS, Xu L, Han SJ, et al. Quantification of protein interactions and solution transport using high-density GMR sensor arrays. *Nat Nanotechnol*. 2011;6:314-320.
- Harrell FE Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. New York, NY: Springer; 2015.
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1-22.

13. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32.

- Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273-297.
- Mazzone PJ, Sears CR, Arenberg DA, et al. Evaluating molecular biomarkers for the early detection of lung cancer: when is a biomarker ready for clinical use? An official American Thoracic Society policy statement. *Am J Respir Crit Care Med*. 2017;196(7):e15-e29.
- Tanner NT, Aggarwal J, Gould MK, et al. Management of pulmonary nodules by community pulmonologists: a multicenter observational study. *Chest*. 2015;148(6):1405-1414.
- R Development Core Team. The R Project for Statistical Computing. 2013. Accessed February 20, 2023. <https://cran.r-project.org>
- Deppen SA, Blume JD, Kensinger CD, et al. Accuracy of FDG-PET to diagnose lung cancer in areas with infectious lung disease: a meta-analysis. *JAMA*. 2014;312(12):1227-1236.
- Wiener RS, Schwartz LM, Woloshin S, Welch HG. Population-based risk of complications following transthoracic needle lung biopsy of a pulmonary nodule. *Ann Intern Med*. 2011;155(3):137-144.
- Massion PP, Antic S, Ather S, et al. Assessing the accuracy of a deep learning method to risk stratify indeterminate pulmonary nodules. *Am J Respir Crit Care Med*. 2020;202(2):241-249.
- Kammer MN, Lakhani DA, Balar AB, et al. Integrated biomarkers for the management of indeterminate pulmonary nodules. *Am J Respir Crit Care Med*. 2021;204(11):1306-1316.

Funding/Support

This study was funded by MagArray, Inc. and by National Institutes of Health [Grant U01CA152662]. A. V. is supported in part by the National Institute of Environmental Health Sciences [Grant P30-ES013508].

Financial/Nonfinancial Disclosures

The authors have reported to *CHEST* the following: A. V., J. K. B., P. P. M., and P. J. M. received funding to their institution from MagArray for the work performed in this study. Outside of the submitted work, A. V. reports personal fees as a scientific advisor to the Lung Cancer Initiative at Johnson & Johnson and reports grants to his institution from Optellum, Precyte, the Gordon and Betty Moore Foundation, and the Lungevity Foundation. P. J. M. reports grants to his institution from Adela, Exact Sciences, DELFI, Nucleix, Biodesix, and Veracyte. M. B., A. L. F., and L. C. are employees of MagArray, Inc. S. X. W. is a consultant to MagArray, Inc. None declared (J. K. B.).