

Validation of two models to estimate the probability of malignancy in patients with solitary pulmonary nodules

E M Schultz,¹ G D Sanders,² P R Trotter,² E F Patz Jr,² G A Silvestri,³ D K Owens,^{1,4} M K Gould^{1,4}

¹Stanford School of Medicine, Stanford, California, USA; ²Duke University, Durham, North Carolina, USA; ³Medical University of South Carolina, Charleston, South Carolina, USA; ⁴VA Palo Alto Health Care System, Palo Alto, California, USA

Correspondence to: Ms E M Schultz, Center for Primary Care and Outcomes Research (PCOR), 117 Encina Commons, Stanford, CA 94305, USA; emschultz@stanford.edu

Received 25 May 2007
Accepted 17 October 2007
Published Online First
26 October 2007

ABSTRACT

Background: Effective strategies for managing patients with solitary pulmonary nodules (SPN) depend critically on the pre-test probability of malignancy.

Objective: To validate two previously developed models that estimate the probability that an indeterminate SPN is malignant, based on clinical characteristics and radiographic findings.

Methods: Data on age, smoking and cancer history, nodule size, location and spiculation were collected retrospectively from the medical records of 151 veterans (145 men, 6 women; age range 39–87 years) with an SPN measuring 7–30 mm (inclusive) and a final diagnosis established by histopathology or 2-year follow-up. Each patient's final diagnosis was compared with the probability of malignancy predicted by two models: one developed by investigators at the Mayo Clinic and the other developed from patients enrolled in a VA Cooperative Study. The accuracy of each model was assessed by calculating areas under the receiver operating characteristic (ROC) curve and the models were calibrated by comparing predicted and observed rates of malignancy.

Results: The area under the ROC curve for the Mayo Clinic model (0.80; 95% CI 0.72 to 0.88) was higher than that of the VA model (0.73; 95% CI 0.64 to 0.82), but this difference was not statistically significant ($\Delta = 0.07$; 95% CI -0.03 to 0.16). Calibration curves showed that the probability of malignancy was underestimated by the Mayo Clinic model and overestimated by the VA model.

Conclusions: Two existing prediction models are sufficiently accurate to guide decisions about the selection and interpretation of subsequent diagnostic tests in patients with SPNs, although clinicians should also consider the prevalence of malignancy in their practice setting when choosing a model.

The solitary pulmonary nodule (SPN) is a common and challenging clinical problem.^{1,2} In patients who are surgical candidates, malignancy should be identified promptly (when present) to permit timely resection. Ideally, surgery should be avoided in patients with nodules that prove to be benign. Previous studies have shown that the effectiveness and cost-effectiveness of SPN management strategies depend critically on the “pre-test” probability of malignancy, that is, the probability of malignancy based on clinical characteristics and radiographic findings before performing other tests.^{3–6} While most clinicians use their intuition and clinical judgement to make this assessment, quantitative prediction models^{7,8} and neural

networks^{9–11} have been developed to facilitate this task.

Swensen and colleagues⁷ at the Mayo Clinic retrospectively reviewed the medical records and imaging tests of 629 patients (51% male) with lung nodules measuring 4–30 mm in diameter that were newly discovered between 1984 and 1986; 65% of the nodules were benign, 23% were malignant and 12% were without a final diagnosis. The authors divided the sample into development ($n = 419$) and validation sets ($n = 210$). Using logistic regression analysis, they identified six independent predictors of malignancy: older age, a history of smoking, a history of an extrathoracic cancer >5 years before nodule detection, larger nodule diameter, upper lobe location and spiculated margins (box 1). Model accuracy was good, with an area under the receiver operating characteristic (ROC) curve of 0.83 and a standard error (SE) of 0.02 in the development set. The area under the ROC curve was 0.80 (SE = 0.04) in the validation set, and the model was well calibrated in both sets, that is, the observed frequency of malignancy was similar to the predicted probability of malignancy in patients grouped by deciles of predicted probability. Of note, this study excluded patients with a history of lung cancer and patients with a history of an extrathoracic cancer that was diagnosed within 5 years of nodule identification.

Members of our group⁸ derived another model using data from 375 patients (98% male) with solitary nodules measuring 7–30 mm (inclusive) in diameter who were enrolled in a VA Cooperative Study that examined the diagnostic accuracy of positron emission tomography (PET) imaging for SPN diagnosis. The prevalence of malignancy was 54%. We identified four independent predictors of malignant SPNs by using logistic regression analysis: older age, a history of smoking, larger nodule diameter and shorter time since quitting smoking (box 1). This model had accuracy similar to the Mayo Clinic model with an area under the ROC curve of 0.79 (SE = 0.02) in the development set and an area under the ROC curve of 0.78 (SE = 0.02) by internal cross-validation. The model was well calibrated in patients with predicted probabilities that were <29% or >66%, but it slightly overestimated the probability of malignancy in patients with low to moderate predicted probability (29–48%) and slightly underestimated the probability of malignancy in patients with moderate to high predicted probability (49–66%).

In this study we validated two existing clinical prediction models that estimate the probability that an indeterminate SPN will be malignant.

METHODS

We retrospectively reviewed the electronic medical records of an independent sample of 151 asymptomatic veterans with SPNs discovered incidentally and collected information about patient history, nodule characteristics and final diagnosis. Using these data, we evaluated the accuracy and calibration of the Mayo Clinic and VA models for estimating the probability of malignancy in patients with indeterminate lung nodules.

Patients

We examined the records of 991 consecutive patients with known or suspected lung cancer who underwent PET imaging at a veterans hospital between January 2000 and February 2006 (fig 1). During that time almost all patients with an SPN underwent both PET and CT scanning at this institution. A trained medical abstractor reviewed the medical records to identify patients who had an SPN that measured 7–30 mm in largest diameter on the chest radiograph or CT scan. We excluded patients who did not have an SPN based on the presence of any of the following radiographic findings: lung mass >3 cm in diameter, atelectasis (except linear or subsegmental), pneumonia, consolidation, infiltrate, pleural effusion, hilar or mediastinal lymphadenopathy, or bony abnormalities suggestive of malignancy. In addition, we excluded patients with more than one nodule on the chest radiograph or more than six nodules on the chest CT scan. We also excluded patients who did not have a CT scan or whose initial CT scan was performed more than 30 days after the PET scan ($n = 31$), because we used data from this same sample to validate a decision analysis model that estimates post-test probabilities of malignant nodules following CT and PET.¹² Finally, we

excluded patients who lacked a final diagnosis ($n = 94$), patients with nodules that were observed to grow before the initial PET scan was performed ($n = 64$) and patients for whom the diagnosis was made before the PET scan ($n = 11$).

We considered the final diagnosis to be malignant when histopathology revealed a specific malignant diagnosis. We considered the final diagnosis to be benign when histopathology revealed a specific benign diagnosis, when the nodule was stable over 2 years of clinical and radiographic follow-up, or when the nodule decreased in size at any time. We also considered the final diagnosis to be benign when patients died with a nodule that was stable after at least 1 year of follow-up.

Variables

For eligible patients the abstractor reviewed the initial chest radiograph and chest CT reports to collect information about nodule diameter, location and edge characteristics. In addition, the abstractor reviewed progress notes, admission reports and demographic files to collect information about age, gender, race, ethnicity, pulmonary function, history of cancer and smoking behaviour. Information about smoking behaviour included smoking history (ever or never), smoking status (current or former), average packs/day, years of smoking, pack-years and number of years since quitting.

Data analysis

We tested for differences in clinical and radiological characteristics between patients with malignant and benign nodules using t tests for continuous variables with a normal distribution and the Mann-Whitney U test for continuous variables with a distribution that differed significantly from normal, as determined by the one-sample Kolmogorov-Smirnov test. For categorical variables we performed χ^2 tests or Fisher exact tests as indicated. Mean values and standard deviation (SD) are reported for continuous variables.

Identical validation techniques were used for both models. We described accuracy by comparing the predicted probability of malignancy with the final diagnosis and constructing ROC curves.¹³ We report the area under the ROC curve and 95% confidence intervals (CI). To compare the accuracy of the two models we sampled with replacement and used the bootstrap to generate a 95% CI for the difference in the areas under the ROC curve.¹⁴

Calibration refers to the relationship between the observed and predicted probabilities of malignant SPNs. We calibrated each model by dividing the sample into five equal groups based on predicted probability and plotting the median probability of each quintile against the observed frequency of malignancy for that group. Analyses were performed using SPSS for Windows Version 15.0 (SPSS, Chicago, Illinois, USA) and SAS for Windows Version 9.1 (Cary, North Carolina, USA). Two-sided p values <0.05 were considered to be statistically significant.

RESULTS

The sample of 151 patients was predominantly male (96%) with a mean (SD) age of 66.9 (10.1) years (table 1). Information about race was not available for 59 patients (39%) and information about ethnicity was not available for 66 patients (44%), but of those for whom it was available, 74 (80%) were white and 81 (95%) were not Hispanic or Latino. The prevalence of malignancy was 44%.

Patients with a malignant diagnosis had larger nodules than those with a benign diagnosis ($p < 0.001$) and were more likely

Box 1 Equations for models that estimate the pre-test probability of malignant SPN

The Mayo Clinic model is defined by the equations:

- ▶ Pre-test probability of a malignant SPN = $e^x / (1 + e^x)$
- ▶ $x = -6.8272 + (0.0391 * \text{age}) + (0.7917 * \text{smoke}) + (1.3388 * \text{cancer}) + (0.1274 * \text{diameter}) + (1.0407 * \text{spiculation}) + (0.7838 * \text{upper})$

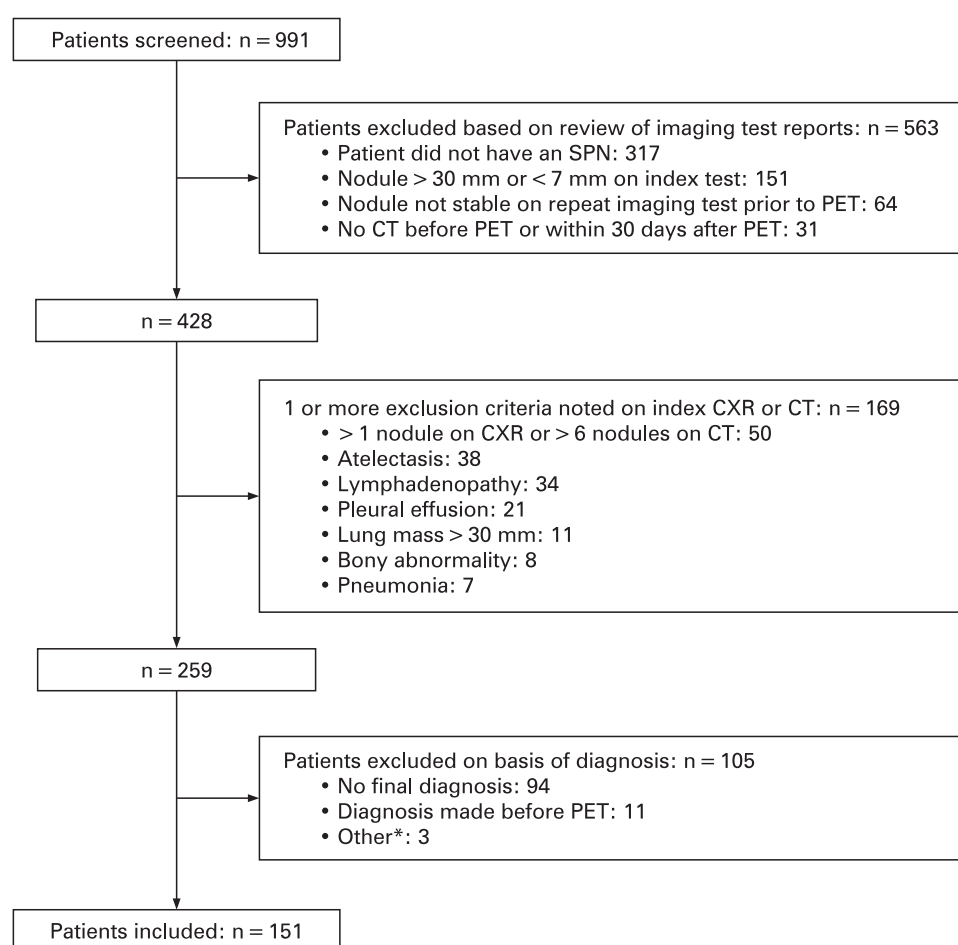
where e is the base of the natural logarithm; age indicates the patient's age in years; smoke indicates smoking history (1 = current or former smoker, 0 = never smoker); cancer indicates history of an extrathoracic cancer ≥ 5 years before nodule identification (1 = yes, 0 = no or not specified); diameter indicates the largest nodule measurement (in mm) reported on initial chest radiograph or CT scan; spiculation indicates mention of nodule spiculation on any imaging test report (1 = yes, 0 = no or not specified); and upper is location of the nodule within the upper lobe of either lung (1 = yes, 0 = no).

The VA model is defined by the equations:

- ▶ Pre-test probability of a malignant SPN = $e^x / (1 + e^x)$
- ▶ $x = -8.404 + (2.061 * \text{smoke}) + (0.779 * \text{age}10) + (0.112 * \text{diameter}) - (0.567 * \text{yearsquit}10)$

where e is the base of the natural logarithm, smoke indicates smoking history (1 = current or former smoker, 0 = never smoker); age10 indicates age in years at the time of nodule identification, divided by 10; diameter indicates the largest nodule measurement (in mm) reported on initial chest radiograph or CT scan; and yearsquit10 indicates the number of years since quitting smoking, divided by 10 (0 indicates not applicable).

Figure 1 Study eligibility and reasons for patient exclusions. Of the 991 patients screened for inclusion in the study, 840 were determined to be ineligible because they met one or more of the exclusion criteria. For those patients who met multiple exclusion criteria, only the first reason for exclusion is presented in this figure. SPN, solitary pulmonary nodule; PET, positron emission tomography; CT, computed tomography; CXR, chest radiograph.



* 3 patients were excluded because they were not eligible for Mayo Clinic model validation due to a history of lung or recent cancer, and were not eligible for VA model validation, because they were used to develop that model.

to have nodules with spiculated borders ($p < 0.001$, table 1). Participants with a malignant diagnosis were somewhat more likely to be older and to be current or former smokers, although these differences were not statistically significant. Former smokers with malignant nodules tended to have quit smoking more recently than those with benign nodules, but these differences were also not statistically significant.

Validation of the Mayo Clinic model

We excluded 33 patients from the validation of this model because they had a history of lung cancer or had been diagnosed with an extrathoracic cancer within 5 years of nodule identification. The prevalence of malignancy in the validation sample was 45%. Of the remaining 118 patients, three were missing information about nodule location and no information on smoking history was available for one.

When these four patients were excluded from the validation, the area under the ROC curve (AUC) was 0.80 (95% CI 0.72 to 0.88), as shown in fig 2A. The AUC changed only slightly when we performed a sensitivity analysis in which we assumed that all four patients with missing data were smokers with nodules in the upper lobe (AUC = 0.81, 95% CI 0.73 to 0.89). The AUC and 95% CI were identical when we assumed that patients with missing data were smokers with nodules in the middle or lower

lobes. In these analyses, we always assumed that the one individual with missing information about smoking history was a smoker because the prevalence of smoking was so high in this sample (89%).

A calibration curve showed that, for patients in all but the lowest quintile of predicted probability (fig 2B), the median predicted probability was lower than the observed frequency of malignant nodules by approximately 2% (fourth quintile) to 12% (second quintile). However, the observed frequency of malignant nodules fell within the range of predicted probabilities for patients in each quintile, except for patients in the third quintile in whom the observed frequency of malignant nodules was 48% and the range of predicted probabilities was 31–47%.

The accuracy (AUC 0.78, 95% CI 0.71 to 0.86) and calibration of the Mayo Clinic model were similar when we included the 33 patients who had a prior diagnosis of lung cancer or an extrathoracic cancer that was diagnosed within 5 years of nodule identification.

Validation of the VA model

We excluded 12 patients from validation of this model because they were included in the sample used to develop the model.⁸ Excluded patients tended to be younger, were more frequently current smokers, were heavier smokers, had a higher proportion

Table 1 Sample characteristics and comparison of patients with malignant and benign solitary pulmonary nodules

Characteristic	Full sample (n = 151)	Malignant (n = 66)	Benign (n = 85)	p Value
Clinical				
Mean (SD) age at index test (years)	66.9 (10.1)	68.6 (8.0)	65.6 (11.4)	0.06
Male	145 (96)	62 (94)	83 (98)	0.41§
Smoking history				
Current or former smoker	135 (89)	63 (95)	72 (85)	0.15
Never smoker	12 (8)	3 (5)	9 (11)	
Current smoker	59 (39)	29 (44)	30 (35)	0.63
Former smoker	69 (46)	31 (47)	38 (45)	
Mean (SD) pack-years*	49.6 (31.2)	52.5 (30.5)	46.8 (31.9)	0.34
Mean (SD) years quit†	15.0 (12.8)	12.2 (10.7)	17.7 (14.2)	0.09
History of other cancer ≤5 years previously	25 (17)	10 (15)	15 (18)	0.68
History of other cancer >5 years previously	19 (13)	11 (17)	8 (9)	0.18
Any history of lung cancer	9 (6)	3 (5)	6 (7)	0.73§
Radiological				
Nodule in right lung	98 (65)	44 (67)	54 (64)	0.69
Location of nodule				
Upper lobe	80 (53)	40 (61)	44 (47)	0.11
Middle or lower lobe	68 (45)	25 (38)	43 (51)	
Median (range) nodule size (mm)‡	15 (7–30)	20 (8–30)	13 (7–30)	<0.001¶
Nodule spiculated	57 (38)	40 (61)	17 (20)	<0.001
PET results positive (any uptake)	65 (43)	55 (83)	10 (12)	<0.001

SD, standard deviation; PET, positron emission tomography. Data are presented as number (%) unless otherwise specified. Subcategories do not sum to 100% because the variable value was unknown or unspecified for the remainder of patients. Unless otherwise indicated, χ^2 tests were performed for categorical variables and *t* tests were performed for continuous variables. Missing data were excluded from all statistical tests. *N = 108. Pack-years were unknown for 27 smokers. †Applies only to former smokers (N = 64). Years quit were unknown for 5 former smokers. ‡Only patients with SPN(s) measuring 7–30 mm (inclusive) in diameter were eligible for this study. §Fisher exact test. ¶Mann-Whitney U test.

of positive PET scans and had larger nodules, although none of these differences was statistically significant. The prevalence of malignancy among patients used to validate the VA model was 42% compared with 67% for those excluded. These differences probably reflect the very high prevalence of malignancy in the development set (54%). Patients excluded from the analysis did not differ significantly or meaningfully with respect to sex, race, ethnicity, nodule location or spiculation.

Of the remaining 139 patients used for validation, 15 (11%) were missing information necessary for estimating pre-test probability with the VA model. Of these, four patients were known to be former smokers but were missing information about time since quitting; seven patients were known to have smoked but were missing information about their current smoking status and years since quitting smoking; and four patients were missing all information about smoking history.

Patients missing values for smoking-related variables did not differ significantly from those with complete information with respect to age, sex, race, ethnicity, cancer history, nodule location, nodule edge characteristics, method of first nodule detection or method of final diagnosis. However, patients with missing information tended to have smaller nodules (mean difference −5 mm, *p* = 0.07) and were more likely to have a benign final diagnosis (*p* = 0.07).

Excluding patients with missing data, the area under the ROC curve for the VA model was 0.73 (95% CI 0.64 to 0.82), as shown in fig 3A. The results for the VA model did not vary when we used three different sets of assumptions (Appendix) to impute values for missing data (AUC 0.72–0.75; lower limit of 95% CI 0.64 to 0.67; upper limit of 95% CI 0.81 to 0.83). In contrast to the Mayo Clinic model, the median predicted probability for patients in each quintile was higher than the

observed frequency of malignant nodules by approximately 2% (first quintile) to 16% (second quintile) (fig 3B). When constructing the ROC and calibration curves shown in fig 3, we excluded patients with missing data.

The Mayo Clinic model appeared to be slightly more accurate than the VA model, although the difference in the areas under the ROC curves was not statistically significant (Δ = 0.07; 95% CI −0.03 to 0.16).

DISCUSSION

We have validated two models that use clinical history and radiographic findings to estimate the probability of malignancy in patients with SPNs. Both models appeared to be sufficiently accurate to inform clinical decisions about the choice and interpretation of subsequent diagnostic tests. The accuracy of both models was similar to that reported in the papers that described their development.^{7 8} The Mayo Clinic model was slightly more accurate than the VA model, but this difference was not statistically significant.

In our validation, the Mayo Clinic model underestimated the probability of malignancy for patients in all but the lowest quintile of predicted probabilities. This is perhaps not surprising, given that the Mayo Clinic model was developed in a sample of patients with a much lower prevalence of malignancy (23%) than our sample (44%).

Another group recently examined the accuracy of the Mayo Clinic model and obtained similar results. Using data from 106 patients with SPNs in the Netherlands, of whom 61 (57%) had malignant nodules, Herder and colleagues¹⁵ found that the AUC was 0.79 (95% CI 0.70 to 0.87), almost identical to our estimate. As in our study, the authors found that the model underestimated the clinical probability of malignancy.

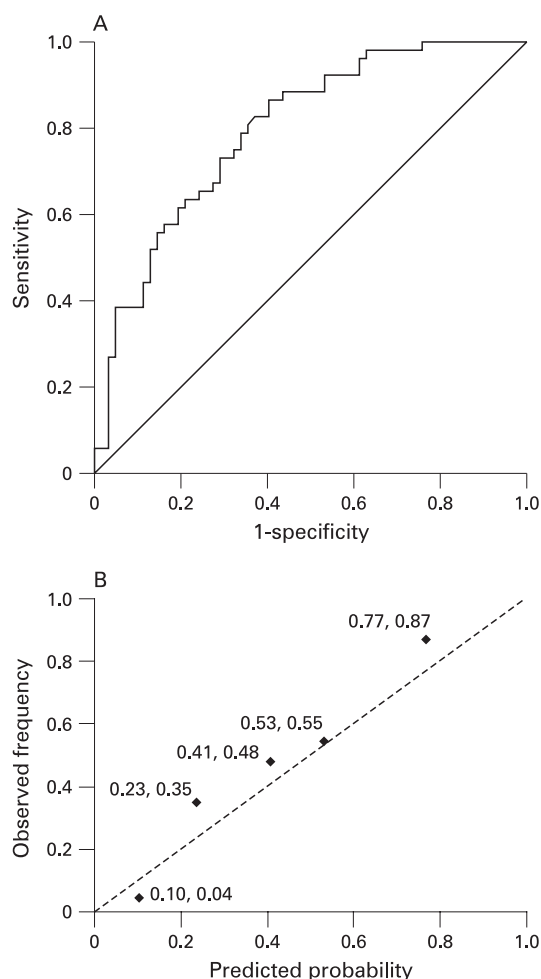


Figure 2 (A) Receiver operating characteristic curve and (B) calibration curve for validation of the Mayo Clinic model. (A) AUC = 0.80 (95% CI 0.72 to 0.88). (B) The calibration curve plots the median predicted probability of a malignant SPN by the observed frequency of malignancy for patients in each quintile of predicted probability. The range of predicted probabilities for patients in the first, second, third, fourth and fifth quintiles were <0.15, 0.15–0.30, 0.31–0.46, 0.47–0.66 and >0.66, respectively. AUC, area under the receiver operating characteristic curve; CI, confidence interval; SPN, solitary pulmonary nodule.

In contrast to the Mayo Clinic model, the VA model was developed from a sample with a relatively high prevalence of malignancy (54%). As might be expected, this model overestimated probabilities of malignancy for patients in our validation sample, among whom prevalence was lower (44%).

Thus, one option for practising clinicians would be to use both models to generate a range of pre-test probabilities: the Mayo Clinic model would provide the lower bound and the VA model would provide the upper bound. Alternatively, it may be reasonable to use the VA model instead of the Mayo Clinic model in some circumstances because it may be preferable to overestimate, rather than underestimate, pre-test probability. While overestimates might result in unnecessary surgery or biopsy in some patients with benign nodules (false positive diagnosis), underestimates might lead to a more dire delayed diagnosis and missed opportunities for surgical cure in patients with malignant nodules (false negative diagnosis).

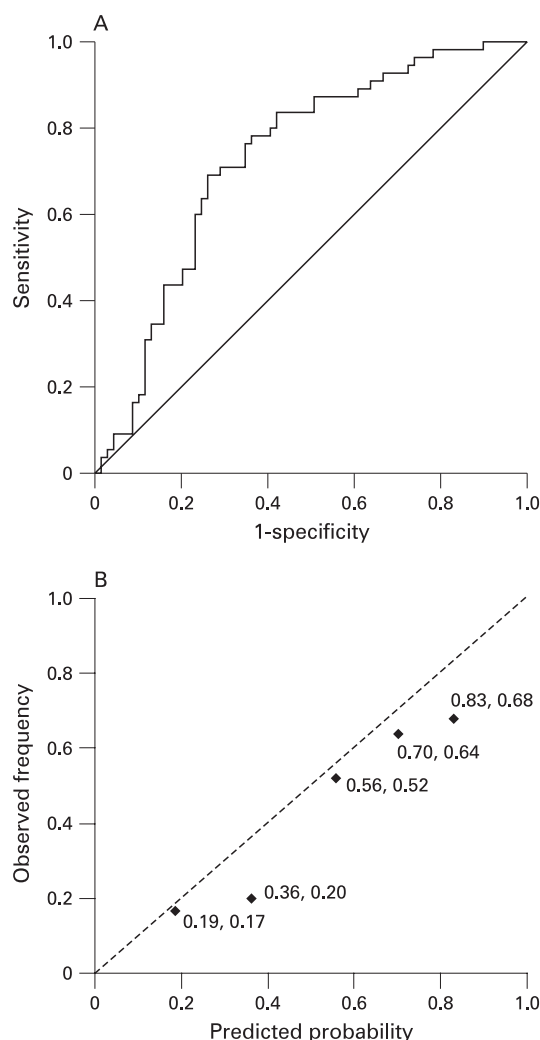


Figure 3 (A) Receiver operating characteristic curve and (B) calibration curve for validation of the VA Cooperative Study model. (A) AUC = 0.73 (95% CI 0.64 to 0.82). (B) The calibration curve plots the median predicted probability of a malignant SPN by the observed frequency of malignancy for patients in each quintile. The range of predicted probabilities for patients in the first, second, third, fourth and fifth quintiles were <0.26, 0.26–0.45, 0.46–0.65, 0.66–0.75 and >0.75, respectively. AUC, area under the receiver operating characteristic curve; CI, confidence interval; SPN, solitary pulmonary nodule.

At present we recommend that, when choosing a model, clinicians should carefully weigh the relative consequences of a false positive diagnosis compared with those for a false negative diagnosis. They should also be mindful of the prevalence of malignant nodules in their practice setting when interpreting model estimates. Although further research is needed to determine whether the pattern of over- and under-calibration that we observed is generalisable to other patient groups or other models, the Mayo Clinic model is probably a better choice in practice settings with a low prevalence of malignant lung nodules. Conversely, the VA model may be a better choice in settings with a high prevalence of malignant SPN.

For an example of how to use these models, we present the hypothetical case of a previously healthy 54-year-old former smoker who quit smoking 10 years ago. The patient has a

smooth-bordered non-calcified nodule in the right upper lobe that measures 16 mm in diameter. Using the equations presented in box 1, the estimated probability of a malignant SPN is 25% according to the Mayo Clinic model and 29% according to the VA model. Previous research has shown that use of PET imaging was most cost-effective in patients with indeterminate nodules on CT scanning and a low to moderate (<40%) probability of malignancy,⁶ so it would be reasonable to order a PET scan in this case. In contrast, the estimated probability of malignancy in a 68-year-old smoker with a 22 mm spiculated nodule in the right upper lobe is 78% by the Mayo Clinic model and 81% by the VA model. If this patient is a good surgical candidate and has no evidence of intrathoracic or distant metastasis on the CT scan, the probability of malignancy is so high that it would be reasonable to proceed directly to surgical resection. The probability estimates generated by the models could also be incorporated into future clinical decision support systems.

Our study has several limitations. We collected information about nodule size, location and spiculation from radiology reports rather than by examining chest radiographs and CT images directly. Similarly, we relied on progress notes in patients' medical records for information about smoking and cancer history. Inaccurate or incomplete information taken from these records, when used to calculate probabilities of malignancy, would result in lower model accuracy. Accordingly, this study provides a conservative test of model accuracy. More accurate data would yield higher, rather than lower, measures of accuracy.

When available, we used the largest nodule diameter from the first abnormal chest radiograph to calculate nodule size. However, 73 patients (48%) did not have a chest radiograph or measurements from the first abnormal chest radiograph were not recorded. When a chest radiograph measurement was unavailable, we calculated nodule size using measurements from the CT scan instead. If chest radiograph and CT measurements differed substantially, this would affect our results. However, when we used only CT measurements, which were available for 96% of patients, we found that the areas under the ROC curve and 95% confidence intervals for both models were nearly identical to results using both chest radiograph and CT measurements. Model calibrations were also very similar. This suggests that both chest radiograph and CT measurements may be used with these models without affecting accuracy, making them more widely applicable in clinical practice.

Our validation was also limited by missing data. In the sample used to validate the VA model, 11% of patients were missing information about smoking history. However, we found that model accuracy and calibration were not sensitive to multiple methods of imputing missing values for these patients. Fewer than 5% of patients in the sample used to validate the Mayo Clinic model were missing information. Again, we found very similar model accuracy and calibration using multiple methods of imputing missing values. In general, characteristics of patients with and without missing data were similar.

Finally, because our sample was composed primarily of older white men with a history of smoking, our results may not be generalisable to other groups of patients. In particular, this study does not directly address questions about the accuracy of these models when predicting pre-test probabilities of malignant SPNs in women. However, there is no evidence that characteristics of benign and malignant pulmonary nodules

differ between men and women. The potential advantages of quantitatively estimating the pre-test probability of malignancy in women with SPNs probably outweigh the risk in generalising to this population. In addition, the validity of the VA model for predicting malignancy in nodules that measure <7 mm in diameter has not been established. Additional research using this or other models to accurately assess the probability of malignancy in small "sub-centimetre" nodules is needed because the detection of small nodules is on the rise as more CT scans of the chest are being performed.

In conclusion, our study demonstrates that both the Mayo Clinic and VA models predict the pre-test probability of malignant SPNs with accuracy that is sufficient to make them clinically useful, especially since neither model is intended to be used as a stand-alone diagnostic test. By using a validated model to estimate pre-test probability, clinicians will be able to make better informed decisions about the selection and interpretation of subsequent diagnostic tests in patients with SPNs, and may thereby deliver more effective and efficient care.

Acknowledgements: We would like to thank Mithat Gönen for his assistance in applying the bootstrap method.

Funding: This study was supported by the National Cancer Institute (R01 CA117840). The National Cancer Institute had no role in study design, data collection, data analysis or manuscript preparation. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or the National Cancer Institute.

Competing interests: None.

Ethics approval: The Institutional Review Board at Stanford University approved this study and waived the requirements for informed consent.

REFERENCES

1. Ost D, Fein AM, Feinsilver SH. Clinical practice. The solitary pulmonary nodule. *N Engl J Med* 2003;**348**:2535–42.
2. Tan BB, Flaherty KR, Kazerooni EA, et al. The solitary pulmonary nodule. *Chest* 2003;**123**:89–96S.
3. Cummings SR, Lillington GA, Richard RJ. Managing solitary pulmonary nodules. The choice of strategy is a "close call". *Am Rev Respir Dis* 1986;**134**:453–60.
4. Dietlein M, Weber K, Gandjour A, et al. Cost-effectiveness of FDG-PET for the management of solitary pulmonary nodules: a decision analysis based on cost reimbursement in Germany. *Eur J Nucl Med* 2000;**27**:1441–56.
5. Gambhir SS, Shepherd JE, Shah BD, et al. Analytical decision model for the cost-effective management of solitary pulmonary nodules. *J Clin Oncol* 1998;**16**:2113–25.
6. Gould MK, Sanders GD, Barnett PG, et al. Cost-effectiveness of alternative management strategies for patients with solitary pulmonary nodules. *Ann Intern Med* 2003;**138**:724–35.
7. Swensen SJ, Silverstein MD, Ilstrup DM, et al. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med* 1997;**157**:849–55.
8. Gould MK, Ananth L, Barnett PG. A clinical model to estimate the pre-test probability of lung cancer in patients with solitary pulmonary nodules. *Chest* 2007;**131**:383–8.
9. Henschke CI, Yankelevitz DF, Mateescu I, et al. Neural networks for the analysis of small pulmonary nodules. *Clin Imaging* 1997;**21**:390–9.
10. Matsuki Y, Nakamura K, Watanabe H, et al. Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution CT: evaluation with receiver operating characteristic analysis. *AJR Am J Roentgenol* 2002;**178**:657–63.
11. Nakamura K, Yoshida H, Engelmann R, et al. Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules with use of artificial neural networks. *Radiology* 2000;**214**:823–30.
12. Thompson EM, Sanders GD, Trotter PR, et al. Validation of a decision model for managing solitary pulmonary nodules [abstract]. *Am J Respir Crit Care Med* 2007;**175**:A869.
13. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36.
14. Davison AC, Hinkley DV. *Bootstrap methods and their application*. New York: Cambridge University Press, 1997.
15. Herder GJ, van Tinteren H, Golding RP, et al. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18-F-fluorodeoxyglucose positron emission tomography. *Chest* 2005;**128**:2490–6.

APPENDIX

(A) Assumptions for imputing missing data

We used three different sets of assumptions to impute values for missing data about smoking behaviour when validating the VA model. In all three cases, if a patient was known or assumed to be a former smoker, we assumed that time since quitting smoking was 15.1 years (the mean value in all former smokers within the VA validation sample). In the first set of assumptions we assumed that participants with missing data about smoking history had smoked and that patients with missing data about smoking status were current smokers. In the second set of assumptions we assumed that patients with missing data about smoking history had smoked and that patients with missing data about current smoking status were former smokers. In the

third set of assumptions we assumed that patients with missing data about smoking history had never smoked and that patients with missing data about current smoking status were former smokers. All three sets of assumptions yielded similar results.

(B) Characteristics of samples used to validate each model

Of the 151 patients included in the study, 33 were excluded from validation of the Mayo Clinic model because they had a history of lung cancer or a history of an extrathoracic cancer within 5 years of nodule identification. The characteristics of the 118 patients used to validate the Mayo Clinic model are shown in table B1. Of the 151 patients included in this study, 12 were excluded from validation of the VA model because those patients were used to develop the model in previous work.⁸ The characteristics of the 139 patients used to validate the VA model are shown in table B2.

Table B1 Characteristics of patients used to validate the Mayo Clinic model

Characteristic	Full sample (n = 118)	Malignant (n = 53)	Benign (n = 65)	p Value
<i>Clinical</i>				
Mean (SD) age at index test (years)	65.9 (10.1)	68.3 (7.7)	63.9 (11.3)	0.01
Male	114 (97)	51 (96)	63 (97)	1.0*
Smoking history				
Current or former smoker	107 (91)	50 (94)	57 (88)	0.35*
Never smoker	10 (9)	3 (6)	7 (11)	
Current smoker	48 (41)	24 (45)	24 (37)	0.56
Former smoker	52 (44)	23 (43)	29 (45)	
Mean (SD) pack-years‡	49.0 (31.1)	53.3 (31.7)	44.9 (30.3)	0.21
Mean (SD) years quit§	15.3 (13.2)	12.3 (10.1)	18.0 (15.2)	0.13
History of other cancer >5 years previously	7 (6)	4 (8)	3 (5)	0.70¶
<i>Radiological</i>				
Nodule in right lung	74 (63)	34 (64)	40 (62)	0.77
Location of nodule				0.28
Upper lobe	60 (51)	30 (57)	30 (46)	
Middle or lower lobe	55 (47)	22 (42)	33 (51)	
Median (range) nodule size (mm)**	16.0 (7–30)	20.0 (8–30)	13.0 (7–30)	<0.001†
Nodule spiculated	42 (36)	33 (62)	9 (14)	<0.001
PET results positive (any uptake)	55 (47)	47 (89)	8 (12)	<0.001

For definitions, see footnote to table B2.

Table B2 Characteristics of patients used to validate the VA model

Characteristic	Full sample (n = 139)	Malignant (n = 58)	Benign (n = 81)	p Value
<i>Clinical</i>				
Mean (SD) age at index test (years)	67.1 (10.2)	68.7 (8.2)	66.0 (11.3)	0.11
Male	133 (96)	54 (93)	79 (98)	0.24*
Smoking history				
Current or former smoker	123 (88)	55 (95)	68 (84)	0.19
Never smoker	12 (9)	3 (5)	9 (11)	
Current smoker	51 (37)	24 (41)	27 (33)	0.67
Former smoker	65 (47)	28 (48)	37 (46)	
Mean (SD) pack-years¶	49.2 (31.4)	51.8 (30.7)	46.7 (32.1)	0.42
Mean (SD) years quit††	15.1 (13.1)	12.0 (11.1)	17.7 (14.2)	0.09
History of other cancer ≤5 years previously	25 (18)	10 (17)	15 (19)	0.85
History of other cancer >5 years previously	19 (14)	11 (19)	8 (10)	0.12
Any history of lung cancer	9 (7)	3 (5)	6 (7)	0.74*
<i>Radiological</i>				
Nodule in right lung	89 (64)	39 (67)	50 (62)	0.50
Location of nodule				0.16
Upper lobe	74 (53)	35 (60)	39 (48)	
Middle or lower lobe	62 (45)	22 (38)	40 (49)	
Median (range) nodule size (mm)**	15.0 (7–30)	20.0 (8–30)	13.0 (7–30)	<0.001†
Nodule spiculated	52 (37)	35 (60)	17 (21)	<0.001
PET results positive (any uptake)	57 (41)	47 (81)	10 (12)	<0.001

SD, standard deviation; PET, positron emission tomography. Data are presented as number (%) unless otherwise specified. Subcategories do not sum to 100% because the variable value was unknown or unspecified for the remainder of patients. Unless otherwise indicated, χ^2 tests were performed for categorical variables and t tests were performed for continuous variables. Missing data were excluded from all statistical tests. *Fisher exact test. †Mann-Whitney U test. ‡N = 86 (pack-years unknown for 21 smokers). §Applies only to former smokers (N = 49). Years quit were unknown for 3 former smokers. ¶N = 100. Pack-years unknown for 23 smokers. **Only patients with SPN(s) measuring between 7 and 30 mm (inclusive) in diameter were eligible for this study. ††Applies only to former smokers (N = 61). Years quit unknown for 4 former smokers.