Validation of Two Models to Estimate the Probability of Malignancy in Patients with Solitary Pulmonary Nodules

Ellen M Schultz¹, Gillian D Sanders², Priscilla R Trotter², Edward F Patz, Jr², Gerard A Silvestri³, Douglas K Owens¹,⁴, Michael K Gould¹,⁴

¹Stanford School of Medicine, Stanford, CA, United States; ²Duke University, Durham, NC, United States; ³Medical University of South Carolina, Charleston, SC, United States and ⁴VA Palo Alto Health Care System, Palo Alto, CA, United States

Correspondence to: E Schultz, emschultz@stanford.edu, Center for Primary Care and Outcomes Research (PCOR), 117 Encina Commons, Stanford, CA 94305 USA, Phone (650) 736-0397, Fax (650) 723-1919

Keywords: Lung Neoplasms; Coin Lesion, pulmonary; Diagnosis; Models, statistical; Receiver Operating Characteristic (ROC) curve

Word count: 3676
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 solitary pulmonary nodule (SPN) is malignant, based on clinical characteristics and radiographic findings.

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

Results: The area under the ROC curve for the Mayo Clinic model (0.80; 95% CI 0.72-0.88) was higher than that of the VA model (0.73; 95% CI 0.64-0.82), but this difference was not statistically significant (Δ=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.
INTRODUCTION
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, i.e. the probability of malignancy based on clinical characteristics and radiographic findings prior to performing other tests.[3-6] While most clinicians use their intuition and clinical judgment to make this assessment, quantitative prediction models [7, 8] and neural networks [9-11] have been developed to facilitate this task.

Swensen and colleagues at the Mayo Clinic retrospectively reviewed the medical records and imaging tests of 629 patients (52% male) with lung nodules measuring 4 to 30 mm in diameter that were newly discovered between 1984 and 1986.[7] Sixty-five percent 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 more than 5 years before nodule detection, larger nodule diameter, upper lobe location, and spiculated margins (Box).

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, i.e. 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 group[8] derived another model using data from 375 patients (98% male) with solitary nodules measuring 7 to 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. 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).

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% to 48%) and slightly underestimated the probability of malignancy in patients with moderate to high predicted probability (49% to 66%).

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

METHODS
The Institutional Review Board at Stanford University approved this study and waived the requirements for informed consent.
We retrospectively reviewed the electronic medical records of an independent sample of 151 asymptomatic veterans with incidentally-discovered SPNs, 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 indeterminant 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 (Figure 1). During that time, almost all patients with an SPN underwent both PET and computed tomography (CT) at this institution. A trained medical abstractor reviewed medical records to identify patients who had an SPN that measured 7 to 30 mm in largest diameter on chest x-ray or chest CT. 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 chest x-ray or more than six nodules on chest CT. 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 x-ray 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 behavior. Information about smoking behavior included smoking history (ever or never), smoking status (current or former), average packs-per-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. We performed 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 1-sample Kolmogorov-Smirnov test. For categorical variables, we performed \( \chi^2 \) tests or Fisher’s exact test, as indicated. We report the mean and standard deviation (SD) for continuous variables.

We used identical validation techniques 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% confidence interval 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, IL) and SAS for Windows version 9.1 (Cary, NC). We considered two-sided \( P \)-values < 0.05 to be statistically significant.

**RESULTS**

The sample of 151 patients was predominantly male (96%) with a mean age of 66.9 years (SD = 10.1) (Table). 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. Prevalence of malignancy was 44%.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Full Sample (n = 151)</th>
<th>Malignant (n = 66)</th>
<th>Benign (n = 85)</th>
<th>( P )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD Age at Index Test (years)</td>
<td>66.9 ± 10.1</td>
<td>68.6 ± 8.0</td>
<td>65.6 ± 11.4</td>
<td>0.06</td>
</tr>
<tr>
<td>Male</td>
<td>145 (96)</td>
<td>62 (94)</td>
<td>83 (98)</td>
<td>0.41*</td>
</tr>
<tr>
<td>Smoking History</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Current or Former Smoker</em></td>
<td>135 (89)</td>
<td>63 (95)</td>
<td>72 (85)</td>
<td>0.15</td>
</tr>
<tr>
<td><em>Never Smoker</em></td>
<td>12 (8)</td>
<td>3 (5)</td>
<td>9 (11)</td>
<td></td>
</tr>
<tr>
<td><em>Current Smoker</em></td>
<td>59 (39)</td>
<td>29 (44)</td>
<td>30 (35)</td>
<td>0.63</td>
</tr>
<tr>
<td><em>Former Smoker</em></td>
<td>69 (46)</td>
<td>31 (47)</td>
<td>38 (45)</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD Pack-years*</td>
<td>49.6 ± 31.2</td>
<td>52.5 ± 30.5</td>
<td>46.8 ± 31.9</td>
<td>0.34</td>
</tr>
<tr>
<td>Mean ± SD Years Quit**</td>
<td>15.0 ± 12.8</td>
<td>12.2 ± 10.7</td>
<td>17.7 ± 14.2</td>
<td>0.09</td>
</tr>
<tr>
<td>History Other Cancer ≤ 5 Years Prior</td>
<td>25 (17)</td>
<td>10 (15)</td>
<td>15 (18)</td>
<td>0.68</td>
</tr>
<tr>
<td>History Other Cancer &gt; 5 Years Prior</td>
<td>19 (13)</td>
<td>11 (17)</td>
<td>8 (9)</td>
<td>0.18</td>
</tr>
<tr>
<td>Any History Lung Cancer</td>
<td>9 (6)</td>
<td>3 (5)</td>
<td>6 (7)</td>
<td>0.73*</td>
</tr>
<tr>
<td><strong>Radiological</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodule in Right Lung</td>
<td>98 (65)</td>
<td>44 (67)</td>
<td>54 (64)</td>
<td>0.69</td>
</tr>
<tr>
<td>Location of Nodule</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Upper Lobe</em></td>
<td>80 (53)</td>
<td>40 (61)</td>
<td>44 (47)</td>
<td>0.11</td>
</tr>
<tr>
<td>*Middle or Lower Lobe</td>
<td>68 (45)</td>
<td>25 (38)</td>
<td>43 (51)</td>
<td></td>
</tr>
<tr>
<td>Median (range) Nodule Size (mm)†</td>
<td>15 (7-30)</td>
<td>20 (8-30)</td>
<td>13 (7-30)</td>
<td>&lt;0.0001\††</td>
</tr>
<tr>
<td>Nodule Spiculated</td>
<td>57 (38)</td>
<td>40 (61)</td>
<td>17 (20)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>PET Results Positive (any uptake)</td>
<td>65 (43)</td>
<td>55 (83)</td>
<td>10 (12)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

SD, Standard deviation; PET, Positron Emission Tomography.
Data are presented as number (%) unless otherwise specified. Subcategories in italics do not sum to 100% because the variable value was unknown or unspecified for the remainder of patients. Unless otherwise indicated, \( \chi^2 \) 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 was unknown for 5 former smokers.
† Only patients with SPN(s) measuring between 7 and 30 mm (inclusive) in diameter were eligible for this study.
‖ Fisher's Exact test.
†† Mann-Whitney U test.

Patients with a malignant diagnosis had larger nodules than those with a benign diagnosis ($P < 0.0001$) and were more likely to have nodules with spiculated borders ($P < 0.0001$) (Table). 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. Prevalence of malignancy in the validation sample was 45%. Of the remaining 118 patients, three were missing information about nodule location, and one patient had no information on smoking history available.

When these four patients were excluded from the validation, the area under the ROC curve was 0.80 (95% CI 0.72 to 0.88), as shown in Figure 2, Panel A. 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 0.73-0.89). The AUC and 95% confidence interval 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 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 (Figure 2, Panel B), 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% to 47%.

The accuracy (AUC 0.78, 95% CI 0.71-0.86) 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 development the model.[8] Excluded patients tended to be younger, were more frequently current smokers, were heavier smokers, had a higher proportion of positive PET scans, and had larger nodules, although none of these differences was statistically significant. Prevalence of malignancy among patients used to validate the VA model was 42%, compared to 67% for those excluded. These differences likely 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 0.64-0.82), as shown in Figure 3, Panel A. 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 0.64-0.67; upper limit of 95% CI 0.81-0.83). 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) (Figure 3, Panel B). When constructing the ROC and calibration curves shown in Figure 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 (\( \Delta = 0.07; 95\% \text{ CI} = -0.03 \text{ to } 0.16 \)).

**DISCUSSION**

We validated two models that use clinical history and radiographic findings to estimate the probability of malignancy in patients with SPNs, and determined that both models appear 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, among whom 61 (57%) had malignant nodules, Herder and colleagues [15] found that the AUC was 0.79 (95% CI 0.70-0.87), almost identical to our estimate. As in our study, the authors found that the model underestimated the clinical probability of malignancy.

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 (a false positive diagnosis), underestimates might lead to more dire delayed diagnosis and missed opportunities for surgical cure in patients with malignant nodules (a false negative diagnosis).

At present, we recommend that when choosing a model, clinicians carefully weigh the relative consequences of a false positive diagnosis, compared to those for a false negative diagnosis. Clinicians 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 generalizable to other patient groups or other models, the Mayo Clinic model is likely 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 the Box, the estimated probability of a malignant SPN is 25% according to the Mayo Clinic model and 29% according to the VA model. Previous research showed that use of PET imaging was most cost-effective in patients with indeterminate nodules on CT 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 CT, 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 x-ray 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 x-ray to calculate nodule size. However, 73 patients (48%) did not have a chest x-ray, or measurements from the first abnormal chest x-ray were not recorded. When a chest x-ray measurement was unavailable, we calculated nodule size using measurements from CT instead. If chest x-ray 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 x-ray and CT measurements. Model calibrations were also very similar.
This suggests that both chest x-ray 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. Eleven percent of patients in the sample used to validate the VA model 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 five percent 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 generalizable 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 generalizing 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-centimeter” 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.

COMPETING INTERESTS
None of the authors has any competing interest to report.

FUNDING
This study was supported by National Cancer Institute (RO1 CA117340-2). 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.
FIGURE LEGENDS

Figure 1: Study Eligibility and Reasons for Patient Exclusions.
Of the 991 patients screened for inclusion in this 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 x-ray.

Figure 2: Receiver Operating Characteristic Curve (Panel A) and Calibration Curve (Panel B) for Validation of the Mayo Clinic Model. (A) AUC = 0.80 (95% CI 0.72-0.88). (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 to 0.30, 0.31 to 0.46, 0.47 to 0.66 and >0.66, respectively.
AUC = area under the receiver operating characteristic curve; CI = confidence interval; SPN = solitary pulmonary nodule.

Figure 3: Receiver Operating Characteristic Curve (Panel A) and Calibration Curve (Panel B) for Validation of the VA Cooperative Study Model. (A) AUC = 0.73 (95% CI 0.64 -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 to 0.45, 0.46 to 0.65, 0.66 to 0.75 and >0.75, respectively.
AUC = area under the receiver operating characteristic curve; CI = confidence interval; SPN = solitary pulmonary nodule.
REFERENCES


The Mayo Clinic model is defined by the equations:
1. Pre-test probability of a malignant SPN = $e^x/(1 + e^x)$
2. $x = -6.8272 + (0.0391 \cdot \text{age}) + (0.7917 \cdot \text{smoke}) + (1.3388 \cdot \text{cancer}) + (0.1274 \cdot \text{diameter}) + (1.0407 \cdot \text{spiculation}) + (0.7838 \cdot \text{upper});$

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 or more years before nodule identification (1 = yes, 0 = no or not specified); diameter indicates the largest nodule measurement, in mm, reported on initial chest x-ray or CT; 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:
1. Pre-test probability of a malignant SPN = $e^x/(1 + e^x)$
2. $x = -8.404 + (2.061 \cdot \text{smoke}) + (0.779 \cdot \text{age}_{10}) + (0.112 \cdot \text{diameter}) - (0.567 \cdot \text{yearsquit}_{10});$

where $e$ is the base of the natural logarithm, smoke indicates smoking history (1 = current or former smoker, 0 = never smoker); age$_{10}$ 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 x-ray or CT; and yearsquit$_{10}$ indicates the number of years since quitting smoking, divided by 10 (0 indicates not applicable).
APPENDIX

A. Assumptions for Imputing Missing Data

We used three different sets of assumptions to impute values for missing data about smoking behavior 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 five years of nodule identification. Characteristics of the 118 patients used to validate the Mayo Clinic model are 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] Characteristics of the 139 patients used to validate the VA model are in Table B2.

<table>
<thead>
<tr>
<th>Table B1: Mayo Clinic Model Validation Sample Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
</tr>
<tr>
<td>Clinical</td>
</tr>
<tr>
<td>Mean ± SD Age at Index Test (years)</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Smoking History</td>
</tr>
<tr>
<td>Current or Former Smoker</td>
</tr>
<tr>
<td>Never Smoker</td>
</tr>
<tr>
<td>Current Smoker</td>
</tr>
<tr>
<td>Former Smoker</td>
</tr>
<tr>
<td>Mean ± SD Pack-years‡</td>
</tr>
<tr>
<td>Mean ± SD Years Quit§</td>
</tr>
<tr>
<td>History Other Cancer &gt; 5 Years Prior</td>
</tr>
<tr>
<td>Radiological</td>
</tr>
<tr>
<td>Nodule in Right Lung</td>
</tr>
<tr>
<td>Location of Nodule</td>
</tr>
<tr>
<td>Upper Lobe</td>
</tr>
<tr>
<td>Middle or Lower Lobe</td>
</tr>
<tr>
<td>Median (range) Nodule Size (mm)&quot;</td>
</tr>
<tr>
<td>Nodule Spiculated</td>
</tr>
<tr>
<td>PET Results Positive (any uptake)</td>
</tr>
</tbody>
</table>

group.bmj.com on April 8, 2017 - Published by group.bmj.com
Table B2: VA Model Validation Sample Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Full Sample</th>
<th>Malignant</th>
<th>Benign</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 139)</td>
<td>(n = 58)</td>
<td>(n = 81)</td>
<td></td>
</tr>
<tr>
<td><strong>Clinical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD Age at Index Test (years)</td>
<td>67.1 ± 10.2</td>
<td>68.7 ± 8.2</td>
<td>66.0 ± 11.3</td>
<td>0.11</td>
</tr>
<tr>
<td>Male</td>
<td>133 (96)</td>
<td>54 (93)</td>
<td>79 (98)</td>
<td>0.24*</td>
</tr>
<tr>
<td>Smoking History</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current or Former Smoker</td>
<td>123 (88)</td>
<td>55 (95)</td>
<td>68 (84)</td>
<td>0.19</td>
</tr>
<tr>
<td>Never Smoker</td>
<td>12 (9)</td>
<td>3 (5)</td>
<td>9 (12)</td>
<td></td>
</tr>
<tr>
<td>Current Smoker</td>
<td>51 (37)</td>
<td>24 (41)</td>
<td>27 (33)</td>
<td>0.67</td>
</tr>
<tr>
<td>Former Smoker</td>
<td>65 (47)</td>
<td>28 (48)</td>
<td>37 (46)</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD Pack-years</td>
<td></td>
<td>49.2 ± 31.4</td>
<td>51.8 ± 30.7</td>
<td>46.7 ± 32.1</td>
</tr>
<tr>
<td>Mean ± SD Years Quit°</td>
<td>15.1 ± 13.1</td>
<td>12.0 ± 11.1</td>
<td>17.7 ± 14.2</td>
<td>0.09</td>
</tr>
<tr>
<td>History Other Cancer ≤ 5 Years Prior</td>
<td>25 (18)</td>
<td>10 (17)</td>
<td>15 (19)</td>
<td>0.85</td>
</tr>
<tr>
<td>History Other Cancer &gt; 5 Years Prior</td>
<td>19 (14)</td>
<td>11 (19)</td>
<td>8 (10)</td>
<td>0.12</td>
</tr>
<tr>
<td>Any History Lung Cancer</td>
<td>9 (7)</td>
<td>3 (5)</td>
<td>6 (7)</td>
<td>0.74*</td>
</tr>
<tr>
<td><strong>Radiological</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodule in Right Lung</td>
<td>89 (64)</td>
<td>39 (67)</td>
<td>50 (62)</td>
<td>0.50</td>
</tr>
<tr>
<td>Location of Nodule</td>
<td></td>
<td></td>
<td></td>
<td>0.16</td>
</tr>
<tr>
<td>Upper Lobe</td>
<td>74 (53)</td>
<td>35 (60)</td>
<td>39 (48)</td>
<td></td>
</tr>
<tr>
<td>Middle or Lower Lobe</td>
<td>62 (45)</td>
<td>22 (38)</td>
<td>40 (49)</td>
<td></td>
</tr>
<tr>
<td>Median (range) Nodule Size (mm)**</td>
<td>15.0 (7-30)</td>
<td>20.0 (8-30)</td>
<td>13.0 (7-30)</td>
<td>&lt;0.0001†</td>
</tr>
<tr>
<td>Nodule Spiculated</td>
<td>52 (37)</td>
<td>35 (60)</td>
<td>17 (21)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>PET Results Positive (any uptake)</td>
<td>57 (41)</td>
<td>47 (81)</td>
<td>10 (12)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

SD, Standard deviation; PET, Positron Emission Tomography.
Data are presented as number (%) unless otherwise specified. Subcategories in italics do not sum to 100% because the variable value was unknown or unspecified for the remainder of patients. Unless otherwise indicated, χ² tests were performed for categorical variables, and t tests were performed for continuous variables. Missing data were excluded from all statistical tests.
* Fisher’s Exact test
† Mann-Whitney U test
** Only patients with SPN(s) measuring between 7 and 30 mm (inclusive) in diameter were eligible for this study.
‡ N = 86. Pack-years was unknown for 21 smokers.
§ Applies only to former smokers. N = 49. Years Quit was unknown for 3 former smokers.
‖ N = 100. Pack-years was unknown for 23 smokers.
○ Applies only to former smokers. N = 61. Years Quit was unknown for 4 former smokers.
LICENCE
The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in THORAX editions and any other BMJPG Ltd products to exploit all subsidiary rights, as set out in our licence (http://thorax.bmj.com/fora/licence.pdf).
Figure 1: Study Eligibility and Reasons for Patient Exclusions

Patients screened: n = 991

Patients excluded based on review of imaging test reports: n = 563
  • Patient did not have an SPN: 317
  • Nodule >30 mm or <7 mm on index test: 151
  • Nodule not stable on repeat imaging test prior to PET: 64
  • No CT before PET or within 30 days after PET: 31

n = 428

1 or more exclusion criteria noted on index CXR or CT: n = 169
  • >1 nodule on CXR or >6 nodules on CT: 50
  • Atelectasis: 38
  • Lymphadenopathy: 34
  • Pleural effusion: 21
  • Lung mass >30 mm: 11
  • Bony abnormality: 8
  • Pneumonia: 7

n = 259

Patients excluded on basis of diagnosis: n = 105
  • No final diagnosis: 94
  • Diagnosis made before PET: 11
  • Other*: 3

Patients included: n = 151

*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.
Figure 2, Panel A: Receiver Operating Characteristic Curve for Validation of the Mayo Clinic Model
Figure 2, Panel B: Calibration Curve for Validation of the Mayo Clinic Model
Figure 3, Panel A: Receiver Operating Characteristic Curve for Validation of the VA Cooperative Study Model
Figure 3, Panel B: Calibration Curve for Validation of the VA Cooperative Study Model
Validation of Two Models to Estimate the Probability of Malignancy in Patients with Solitary Pulmonary Nodules

Ellen M Schultz, Gillian D Sanders, Priscilla R Trotter, Edward F Patz, Jr., Gerard A Silvestri, Douglas K Owens and Michael K Gould

Thorax published online October 26, 2007

Updated information and services can be found at:
http://thorax.bmj.com/content/early/2007/10/26/thx.2007.084731

These include:

- Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Articles on similar topics can be found in the following collections

- Lung neoplasms (608)
- Health effects of tobacco use (211)
- Health education (1223)
- Histopathology (28)
- Smoking (1037)
- Tobacco use (1039)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/