A Prediction Model to Help with the Assessment of Adenopathy in Lung Cancer: HAL

Oisin J. O'Connell^{1*}, Francisco A. Almeida², Michael J. Simoff³, Lonny Yarmus⁴, Ray Lazarus^{1†}, Benjamin Young², Yu Chen^{3§}, Roy Semaan³, Timothy M. Saettele^{1||}, Joseph Ciceria², Harmeet Bedi³, Corrine Kliment⁴, Liang Li⁵, Sonali Sethi², Javier Diaz-Mendoza³, David Feller-Kopman⁴, Juhee Song⁵, Thomas Gildea², Hans Lee⁴, Horiana B. Grosu¹, Michael Machuzak², Macarena Rodriguez-Vial¹, George A. Eapen¹, Carlos A. Jimenez¹, Roberto F. Casal¹, and David E. Ost¹

¹Department of Pulmonary Medicine and ⁵Department of Biostatistics, MD Anderson Cancer Center, Houston, Texas; ²Department of Pulmonary Medicine, Cleveland Clinic, Cleveland, Ohio; ³Department of Pulmonary and Critical Care Medicine, Henry Ford Hospital, Detroit, Michigan; and ⁴Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, Maryland

ORCID ID: 0000-0001-6351-5552 (D.E.O.).

Abstract

Rationale: Estimating the probability of finding N2 or N3 (prN2/3) malignant nodal disease on endobronchial ultrasound–guided transbronchial needle aspiration (EBUS-TBNA) in patients with non–small cell lung cancer (NSCLC) can facilitate the selection of subsequent management strategies.

Objectives: To develop a clinical prediction model for estimating the prN2/3.

Methods: We used the AQUIRE (American College of Chest Physicians Quality Improvement Registry, Evaluation, and Education) registry to identify patients with NSCLC with clinical radiographic stage T1–3, N0–3, M0 disease that had EBUS-TBNA for staging. The dependent variable was the presence of N2 or N3 disease (vs. N0 or N1) as assessed by EBUS-TBNA. Univariate followed by multivariable logistic regression analysis was used to develop a parsimonious clinical prediction model to estimate prN2/3. External validation was performed using data from three other hospitals.

Measurements and Main Results: The model derivation cohort (n = 633) had a 25% prevalence of malignant N2 or N3 disease. Younger age, central location, adenocarcinoma histology, and higher positron emission tomography–computed tomography N stage were associated with a higher prN2/3. Area under the receiver operating characteristic curve was 0.85 (95% confidence interval, 0.82–0.89), model fit was acceptable (Hosmer–Lemeshow, $P = 0.62$; Brier score, 0.125). We externally validated the model in 722 patients. Area under the receiver operating characteristic curve was 0.88 (95% confidence interval, 0.85–0.90). Calibration using the general calibration model method resulted in acceptable goodness of fit (Hosmer–Lemeshow test, $P = 0.54$; Brier score, 0.132).

Conclusions: Our prediction rule can be used to estimate prN2/3 in patients with NSCLC. The model has the potential to facilitate clinical decision making in the staging of NSCLC.

Keywords: lung cancer; lung cancer staging; endobronchial ultrasound; mediastinal adenopathy

(Received in original form July 7, 2016; accepted in final form December 19, 2016)

*Present address: Mater Misericordiae University Hospital, Dublin, Ireland.

†Present address: Baylor College of Medicine and Michael E. DeBakey VA Medical Center, Houston, Texas.

§Present address: The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China.

||Present address: Department of Pulmonary Medicine, Saint Luke's Hospital, Kansas City, Missouri.

The American College of Chest Physicians funded the database construction for the AQUIRE program. Some of the data used for this publication were provided through AQUIRE. The researchers are solely responsible for the analysis and any conclusions drawn from the data presented in this report. This research was supported in part by the National Institutes of Health through a Cancer Center Support Grant (P30CA016672), biostatistics core, at MD Anderson Cancer Center. Statistical analysis work was supported in part by the Cancer Center Support Grant (NCI grant P30 CA016672).

Author Contributions: D.E.O. was the principal investigator (PI) for this study and was responsible for project conception, oversight, organization, data collection and auditing, statistical analysis, and manuscript writing. O.J.O'C. was involved in data collection, auditing, and manuscript writing. L.L. and J.S. were the primary biostatisticians for the project, constructed the models and analyses, and contributed to writing. F.A.A. (PI, Cleveland Clinic), M.J.S. (PI, Henry Ford Hospital), L.Y. (PI, Johns Hopkins), R.L., B.Y., Y.C., R.S., T.M.S., J.C., H.B., C.K., S.S., J.D.-M., D.F.-K., T.G., H.L., H.B.G., M.M., M.R.-V., G.A.E., C.A.J., and R.F.C. contributed to the data collection and writing.

Correspondence and requests for reprints should be addressed to David E. Ost, M.D., Department of Pulmonary Medicine, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 1462, Houston, TX 77030. E-mail: dost@mdanderson.org

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 195, Iss 12, pp 1651–1660, Jun 15, 2017

Copyright © 2017 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201607-1397OC on December 21, 2016

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: Estimating the probability that endobronchial ultrasound–guided transbronchial needle aspiration will detect N2 or N3 malignant nodal disease in patients with non–small cell lung cancer can facilitate the selection of proper management strategies.

What This Study Adds to the

Field: A model that uses positron emission tomography/computed tomography N stage, patient age, location of the tumor (central vs. peripheral), and histology was able to accurately predict the probability of N2 or N3 disease being identified by endobronchial ultrasound–guided transbronchial needle aspiration in patients with non–small cell lung cancer. The model was tested at three other hospitals and had excellent discrimination and was able to predict absolute risk after calibration.

Treatment decisions for patients with non–small cell lung cancer (NSCLC) without distant metastases are highly dependent on the extent of lymph node involvement (1–3). Given the limited accuracy of computed tomography (CT) and positron emission tomography (PET) for nodal staging and the superior safety profile of endobronchial ultrasound–guided transbronchial needle aspiration (EBUS-TBNA) over mediastinoscopy, the American College of Chest Physicians (ACCP) evidence-based lung cancer guidelines recommend a needle-based technique as the initial investigation of choice for staging the mediastinum in patients with NSCLC (3–5).

However, mediastinal sampling is not required for all patients with NSCLC (4). If the probability of N2 or N3 disease is sufficiently low, for instance in patients with cT1aN0Mx disease by PET-CT, then proceeding directly to surgery is a reasonable choice (4, 6). What constitutes a sufficiently low probability of disease to justify proceeding directly to surgery is not otherwise well defined. The European Society of Thoracic Surgery working group considers a rate of unforeseen N2

disease at the time of surgery of 10% as acceptable (6).

Providing more precise estimates of the probability of detecting N2/3 disease by EBUS-TBNA for individual patients before deciding on a staging or treatment strategy would therefore be valuable, because this would better inform the decision-making process in terms of the risks and benefits and would also improve the informed consent process. Indeed, such predictive models for estimating the probability of malignancy in solitary pulmonary nodules have been developed and used successfully for some time and their use is now recommended in the ACCP lung cancer guidelines (7–11). However, there are no corresponding models for mediastinal disease.

The primary aim of this study was to develop models to predict whether or not EBUS-TBNA will identify N2 or N3 disease in patients with NSCLC without distant metastases. Secondary goals were to explore the impact of the conditional sensitivity of PET and CT in modeling and to quantify the information value of PET in terms of discriminatory function for predicting the probability of finding N2 or N3 disease on EBUS-TBNA (prN2/3).

Methods

We performed a multicenter retrospective cohort study of consecutive patients with NSCLC that underwent EBUS-TBNA for staging between September 1, 2009, and January 31, 2013. The study was approved by the Institutional Review Board Committee 4, Protocol DR09–0101, at the University of Texas MD Anderson Cancer Center. Data were prospectively collected as part of the AQuIRE (ACCP Quality Improvement Registry, Evaluation, and Education) registry as previously reported (12–14). Data were collected using standardized definitions, quality control checks, and protocols and was entered via a Web-based interface.

All consecutive patients with treatment-naïve NSCLC undergoing EBUS-TBNA for staging were included. Patients with distant metastasis, suspected or confirmed synchronous primaries, recurrent lung cancer, small cell lung cancer, or other active malignancies were excluded. Patients with T4 disease caused by direct mediastinal invasion by CT were also

excluded, because there would be little uncertainty as to nodal involvement.

Patient clinical characteristics, procedural information, and pathology results from the AQuIRE database were used (14, 15). In addition a review of chest CT and PET scans was performed. These were used to define the radiographic N stage and location of the tumor (central vs. peripheral). CT and PET N stage was defined as the highest abnormal nodal station using the International Association for the Study of Lung Cancer staging system. Additional details on variable definitions are provided in the online supplement.

All EBUS-TBNA procedures were performed in a standard manner, with sampling of N3 followed by N2 and then N1 nodes. All lymph nodes measuring 0.5 cm or larger by EBUS were sampled, irrespective of CT and PET results.

Statistical Analysis

Prediction model development. We used data from The University of Texas MD Anderson Cancer Center to develop a model to help with the assessment of adenopathy in lung cancer (HAL). The primary outcome was the presence of either N2 or N3 disease (vs. N0 or N1) as determined by EBUS-TBNA. Additional details on outcome selection and rationale are provided in the online supplement. N2 or N3 disease and N0 or N1 disease groups were compared for each variable using chi-square or Fisher exact test for categorical variables and two sample Student's *t* test or Wilcoxon-Mann-Whitney test.

We evaluated the interaction between CT N stage and PET N stage, based on previous work that suggested the sensitivity of PET for mediastinal lymph node involvement is conditional on the size of the node on CT (16). We specified *a priori* that we would model N stage using interaction terms for PET-CT and tested this interaction using the Cochran-Mantel-Haenszel test. Because PET-CT images do not use contrast we chose to lump N0/N1 disease together for CT but kept N0 and N1 separate for PET when creating interaction terms.

We used univariate logistic regression to identify variables associated with N2 or N3 disease. We specified *a priori* that variables that had a *P* value less than 0.2 on univariate analysis would be candidate

variables for multivariable logistic regression models. We used backward selection with a P value less than 0.05 to stay in the model.

External Validation

Data from three other centers (Johns Hopkins, Henry Ford, and Cleveland Clinic) were subsequently used to externally validate the model. Consecutive cases were entered using identical definitions, forms, and quality control checks as in the development cohort. The model was calibrated to each center using a general calibration model presented in Steyerberg and coworkers (17, 18) (*see online supplement for details*).

Alternative Clinical Scenarios

The baseline model predicts prN2/3 using all information including CT, PET, and histology of the primary tumor. Because staging before diagnosis may be required (19, 20), and because PET imaging may not have been done before EBUS-TBNA, secondary analyses evaluated model prediction without histology and/or PET.

Model Performance Assessment

For each model, area under the receiver operating characteristic (ROC) curve (AUC) was calculated. The Hosmer-Lemeshow goodness-of-fit test, Brier score, and observed versus predicted plots were used to assess model performance. A P value less than 0.05 was considered statistically significant and all tests were two-sided. All statistical analyses were performed in SAS version 9.4 (SAS Institute, Cary, NC) or STATA/SE 14.1 (StataCorp LP, College Station, TX).

Results

A total of 694 patients with clinical radiographic stage T1–3, N0–3, M0 disease underwent staging EBUS-TBNA for NSCLC in the model development cohort. Clinical characteristics according to N stage as determined by EBUS-TBNA are provided in Table E1 in the online supplement. Of these 61 did not have PET imaging performed, so the primary analysis was conducted on 633 patients (Table 1). Of these 633 patients, 160 (25%) had N2 or N3 disease as determined by EBUS-TBNA. The

distribution of positive N2 and N3 nodes by location is shown in Table E2.

Model Development

We found that the sensitivity of PET for nodal disease was higher when there were enlarged lymph nodes on CT (Cochrane-Mantel-Haenszel test, $P < 0.001$), indicating that there was significant effect measure modification between CT N stage and PET N stage for the outcome of nodal involvement. Therefore, for the multivariable analysis we used interaction terms to capture all combinations of PET (N0, N1, or N2/3) and CT (N0/1 or N2/3).

In the multivariable model (Table 2), younger age, location within the central one-third of the hemithorax, adenocarcinoma histology, and higher nodal stage by CT and PET were associated with an increased prN2/3. ROC AUC was 0.85 (95% confidence interval, 0.82–0.89) (Figure 1A), model fit was acceptable as assessed by observed versus predicted plots (Figure 2), and as indicated by the Hosmer-Lemeshow test ($P = 0.62$). The Brier score was 0.125.

The predicted prN2/3 for a variety of scenarios are provided in Table 3, based on the model development cohort. Predicted probabilities are for patients with ages anywhere from 40 to 80 (*see Table E10 for a more detailed prN2/3 and a risk calculator for each specific decade of age*).

External Validation

External validation was conducted at three other centers. A total of 839 patients with clinical radiographic stage T1–3, N0–3, M0 disease underwent staging EBUS-TBNA for NSCLC. Of these, 117 did not have PET imaging so the external validation analysis used 722 patients. Clinical characteristics according to ultimate N stage as determined by EBUS-TBNA are provided in Table E3. ROC AUC was 0.88 (95% confidence interval, 0.85–0.91) (Figure 1B). The AUC for different institutions ranged from 0.82 to 0.92 (Table 4).

Although discrimination as assessed by ROC curves was good, assessment of model fit by observed versus predicted plots (Figure 3A) and by Hosmer-Lemeshow test indicated that calibration of the model was slightly off in two of the three external validation sites (Table 4) ($P < 0.001$), possibly because of the different patient characteristics at different institutions. Of

note, the overall prevalence of N2 disease in the validation sites (28%, 45%, and 56%) was different than the development cohort (25%). However, after calibration using the general calibration model method, multiplying the linear predictor of the logistic regression by $b = 1.1$ and adding $a = 0.72$ (17), goodness of fit was acceptable (Hosmer-Lemeshow test, $P = 0.54$) and observed versus predicted plots (Figure 3B) confirmed good model calibration. The Brier score with calibration was 0.132.

Alternative Clinical Scenarios

Secondary analyses developed models for alternative scenarios, including when PET imaging was not available, histology was not available, and when both PET and histology were not available. The details of model development and validation for each scenario are provided in the online supplement (*see Tables E4–E9, Figures E1–E6*). Model performance when used on the external validation cohorts was fair to good, with ROC AUC ranging from 0.76 to 0.87. Using the general calibration method, Hosmer-Lemeshow test P values ranged from 0.38 to 0.68, and Brier scores ranged from 0.135 to 0.193 (Table 5).

Discussion

In this study we developed a parsimonious clinical prediction model for prN2/3 in patients with NSCLC. Younger age, adenocarcinoma histology, central location, and higher N stage by PET-CT were associated with a higher prN2/3. The model performed well, with good discrimination (AUC ROC, 0.85) and calibration (Hosmer-Lemeshow, $P = 0.62$) in the development cohort. External validation confirmed good discrimination (AUC ROC, 0.88) but calibration was required for higher prN2/3. Calibration of the model using the general calibration method resulted in good model performance as assessed by the Hosmer-Lemeshow test, Brier scores, and plots of observed versus predicted prN2/3. In terms of clinical value, the model adds to the existing body of knowledge by providing more precise estimates of prN2/3 for individual patients. This in turn can help inform and improve the decision-making process.

Other investigators have tried to predict mediastinal lymph node metastases (21–30). Most of these studies focused on

Table 1. Clinical Characteristics by N Stage (0–1 vs. 2–3) as Determined by EBUS-TBNA for 633 Patients with Both CT and PET

	N0 or N1 Disease (n = 473)	N2 or N3 Disease (n = 160)	P Value*	Univariate OR (95% CI)
Age, yr, mean ± SD	68.7 ± 9.4	65.2 ± 10.5	<0.001 [†]	0.96 (0.95–0.98)
Sex, n (%)			0.28	
Female	222 (46.9)	83 (51.9)		1.00 (ref)
Male	251 (53.1)	77 (48.1)		0.80 (0.57–1.13)
Race, n (%)			0.61	
White	397 (84.5)	137 (86.2)		1.00 (ref)
Black	35 (7.4)	13 (8.2)		1.06 (0.57–1.97)
Hispanic	23 (4.9)	7 (4.4)		0.98 (0.45–2.13)
Asian	15 (3.2)	2 (1.3)		0.56 (0.16–1.99)
ASA score, n (%)			0.43 [‡]	
1	3 (0.6)	3 (1.9)		1.00 (ref)
2	35 (7.4)	14 (8.8)		0.45 (0.08–2.45)
3	429 (90.7)	142 (88.8)		0.34 (0.07–1.72)
4	6 (1.3)	1 (0.6)		0.17 (0.01–2.37)
Tobacco history, n (%)			0.65	
Never used	47 (9.9)	12 (7.5)		1.05 (0.55–2.02)
Prior use	326 (68.9)	114 (71.3)		1.04 (0.68–1.58)
Current	100 (21.1)	34 (21.3)		1.00 (ref)
ECOG, n (%)			0.14	
0	135 (28.5)	39 (24.4)		1.00 (ref)
1	238 (50.3)	96 (60)		1.37 (0.91–2.06)
2	85 (18)	23 (14.4)		1.06 (0.62–1.81)
3	15 (3.2)	2 (1.3)		0.45 (0.01–2.03)
Size of the tumor, n (%)			0.39	
≤3 cm	209 (44.2)	63 (39.4)		1.00 (ref)
>3 cm but ≤5 cm	150 (31.7)	60 (37.5)		1.38 (0.93–2.04)
>5 cm	114 (24.1)	37 (23.1)		1.27 (0.83–1.96)
Lobar location of tumor, n (%)			0.23	
Left upper or lingula	141 (29.8)	40 (25)		1.00 (ref)
Left lower lobe	74 (15.6)	20 (12.5)		0.96 (0.54–1.72)
Right upper lobe	158 (33.4)	55 (34.4)		1.29 (0.83–2.01)
Right lower or middle lobe	100 (21.1)	45 (28.1)		1.74 (1.09–2.78)
Location, n (%)			0.05	
Outer two-thirds of lung	365 (77.2)	111 (69.4)		1.00 (ref)
Central one-third of lung	108 (22.8)	49 (30.6)		1.42 (0.98–2.07)
Histology, n (%)			0.02	
Adenocarcinoma	234 (49.5)	97 (60.6)		1.00 (ref)
Squamous cell carcinoma	177 (37.4)	40 (25)		0.51 (0.34–0.75)
Non-small cell carcinoma	40 (8.5)	18 (11.3)		0.95 (0.54–1.67)
Other primary lung cancer	22 (4.7)	5 (3.1)		0.51 (0.21–1.28)
CT characteristics, n (%)			0.09	
Cavitary	18 (3.8)	4 (2.5)		1.00 (ref)
Ground glass, semi-solid or infiltrate	32 (6.8)	4 (2.5)		0.66 (0.18–2.45)
Solid	423 (89.4)	152 (95)		1.33 (0.48–3.62)
Satellite lesion in same lobe, n (%)			0.14	
No	452 (95.6)	157 (98.1)		1.00 (ref)
Yes	21 (4.4)	3 (1.9)		0.38 (0.11–1.27)
N stage by CT, n (%)			<0.001	
0	217 (45.9)	23 (14.4)		1.00 (ref)
1	68 (14.4)	15 (9.4)		1.91 (0.95–3.82)
2	119 (25.2)	79 (49.4)		6.74 (4.09–11.11)
3	69 (14.6)	43 (26.9)		6.61 (3.82–11.45)
N stage by PET (n = 633), n (%)			<0.001	
0	256 (54.1)	15 (9.4)		1.00 (ref)
1	101 (21.4)	17 (10.6)		2.87 (1.38–5.97)
2	72 (15.2)	93 (58.1)		22.04 (12.04–40.36)
3	44 (9.3)	35 (21.9)		13.58 (6.85–26.91)

Definition of abbreviations: ASA = American Society of Anesthesiologists; CI = confidence interval; CT = computed tomography; ECOG = Eastern Cooperative Oncology Group; EBUS-TBNA = endobronchial ultrasound-guided transbronchial needle aspiration; OR = odds ratio; PET = positron emission tomography; ref = reference standard.

P values reported are for *chi-square test except where otherwise noted, [†]two sample Student's *t* test, or [‡]Fisher exact test, and are not the *P* values for the odds ratios reported from univariate analysis.

Table 2. Multivariable Logistic Regression to Predict N Stage (N0–1 vs. N2–3) as Determined by EBUS-TBNA with Interaction Terms for PET-CT

	Parameter Estimate	OR	95% CI		P Value
Intercept	−1.0663				0.21
Age, yr	−0.0311	0.97	0.95	0.99	0.006
Location					
Outer two-thirds of lung	0	1.00			
Central one-third of lung	0.5881	1.80	1.07	3.03	0.03
Histology					
Adenocarcinoma	0	1.00			
Squamous cell carcinoma	−0.8235	0.44	0.26	0.731	0.002
Non-small cell carcinoma	0.1423	1.15	0.55	2.43	0.71
Other primary lung cancer	−0.5079	0.60	0.19	1.91	0.39
N stage by CT and PET					
CT = N0 or N1, PET = N0	0	1.00			
CT = N2 or N3, PET = N0	0.9694	2.64	0.91	7.61	0.07
CT = N0 or N1, PET = N1	1.5766	4.84	1.83	12.8	0.002
CT = N2 or N3, PET = N1	0.9009	2.46	0.67	9.04	0.17
CT = N0 or N1, PET = N2 or N3	2.3726	10.73	4.14	27.76	<0.001
CT = N2 or N3, PET = N2 or N3	3.7531	42.65	18.59	97.87	<0.001

Definition of abbreviations: CI = confidence interval; CT = computed tomography; EBUS-TBNA = endobronchial ultrasound-guided transbronchial needle aspiration; OR = odds ratio; PET = positron emission tomography.

stage I NSCLC, and in some instances the investigators focused on one particular histologic type of NSCLC (23–26). Others restricted the population to patients with CT N0 disease or PET N0/1 disease (22, 27–30). Although these approaches

provide useful research insights, for clinical prediction physicians need a broader rule, because in many cases the histology may be unknown and patients may have larger tumors or more advanced stage disease based on PET-CT.

A broader, more generalizable approach was used by investigators that derived a prediction rule from the Canadian Lung Oncology Group's randomized study of routine versus selective mediastinoscopy (21). The investigators used chest radiographs and clinical risk factors, and their AUC was 0.70. Although their study was useful for estimating pretest probability before PET-CT, it did not use CT or PET data, making the study less useful in current practice.

Our model is consistent with and adds to the existing body of literature on prediction of mediastinal metastasis. This study confirms the findings of other investigators that young age and central location are risk factors for N2 and N3 disease (21, 22, 27, 28, 31). In addition, this study integrates CT and PET data into the model, so the decision context is more applicable to current practice than prior studies (21). Also, previous studies focused on surgical populations. But most patients with N2/N3 disease will never be treated surgically, which limits the generalizability of previous prediction rules developed from surgical cohorts.

Our model also provides insight into the incremental information value of PET when added to CT. Often a patient with a suspected lung cancer presents to their

Figure 1. Receiver operating characteristic curves for the prediction model in (A) test development cohort and (B) external validation cohort. Prediction models use positron emission tomography/computed tomography N stage, age, location of tumor (central one-third vs. other), and histology.

Figure 2. Model development cohort observed versus predicted frequencies. The model uses positron emission tomography/computed tomography N stage, age, location (central one-third vs. outer two-thirds), and histology. The figure plots the probability of N2/N3 disease by decile of expected risk as a function of the actual observed risk in that group. This includes an interaction term for positron emission tomography/computed tomography. The observed probability for each decile is on the vertical axis, the predicted probability according to decile is on the horizontal axis. A perfect model, where observed = predicted is shown by the line.

pulmonologist without a PET scan. A frequent question is should we proceed to EBUS-TBNA or do PET? Our data shows that if the patient has N2 disease by CT, then the incremental information value of PET is really based on its ability to detect extrapulmonary metastases. Finding extrapulmonary metastases on PET will change management. But in the absence of metastatic disease, information about mediastinal lymph node activity on PET will probably not impact the EBUS-TBNA decision because the CT combined with the clinical data suggest that the prN2/3 will be high enough to warrant EBUS-TBNA whether or not the mediastinal lymph nodes are PET positive (*see* Table E10 and the online supplement on applying the model to clinical situations). PET will, however, improve the accuracy of the prN2/3 estimate significantly. Conversely, in patients who have N0 disease by CT, the incremental information value of PET is larger. PET may detect extrapulmonary metastasis and it may also detect abnormal mediastinal activity. Either of these findings would directly impact the decision on whether or not to do EBUS-TBNA.

This serves to highlight the importance of discordant CT and PET results. This study confirms the interaction between size of lymph nodes on CT imaging and PET sensitivity and specificity. This has been described previously in a metaanalysis (16) but it has not to our knowledge been integrated into prior prediction models. This finding is important, because discordant PET-CT images are not uncommon (*see* Table E11).

The model also adds to the existing body of evidence on whether size of the tumor impacts prN2/3. Although some prior studies have reported an association between larger size and prN2/3 (21, 27, 28, 32–36), others have failed to demonstrate an association (29, 37). But many of these studies were limited to T1 patients, most did not or could not adjust for CT and PET N status, and none were predicting EBUS-TBNA yield. Our study failed to demonstrate a relationship between tumor size and prN2/3 after adjustment for age, location, and N status by PET-CT. This is a relevant clinical question, because although current guidelines address the issue of cT1aN0Mx by PET-CT, they do not address the rarer but relevant question of what to do with patients who have large tumors that are N0 by PET-CT. Although prior studies might suggest prN2/3 is high in such cases, our data suggest that prN2/3 is still low. For instance, in this study there were 56 patients who were PET-CT N0 with tumors greater than 5 cm in size. Only two (4%) had N2/3 disease by EBUS-TBNA. Proceeding directly to surgery in such patients, particularly if they are older and have peripheral squamous cell tumors, would be a consideration, depending on the comorbidities present.

When all of these elements are integrated into a model, it can help inform clinical decision making. A good prediction model must have both good discrimination and good calibration (*see* online supplement for details). But model performance in the development cohort usually overestimates actual performance, so validation requires reassessment, ideally with external cohorts that are distinct from the development cohort (38–40).

To our knowledge this is the first study of EBUS-TBNA to externally validate a general clinical prediction model for prN2/3 (21–28). One prior surgical study externally validated a more limited model, focusing only on patients with a PET-negative

mediastinum (29, 30). Their AUC was 0.65 and the model overestimated risk in both the original and external validation cohorts, but age was not included and PET-CT interactions were not accounted for (29, 30).

Although our model demonstrated good discrimination in the external validation cohort, inspection of the observed versus predicted plots (Figure 3A) suggests that at prN2/3 greater than 20% our model underestimates the observed prN2/3. However, in terms of deciding whether or not to do EBUS-TBNA initially, once the prN2/3 is that high, it arguably makes little difference, because mediastinal sampling is warranted whether the true prN2/3 is 20% or 40% (4). So even in high prevalence validation sites, where calibration is limited when prN2/3 is greater than 20%, the model works well in deciding on whether or not to do EBUS-TBNA.

However, there are other benefits to having more precise estimates of prN2/3 that make calibration desirable. ACCP lung cancer guidelines suggest that if the pretest probability of N2 or N3 disease is sufficiently high, then if EBUS-TBNA is negative, mediastinoscopy is warranted (4). However, this is predicated on having accurate estimates of pretest probability before EBUS-TBNA to calculate posterior probabilities (i.e., negative predictive value). Although our model does not predict pretest probability of N2 or N3 disease based on the true gold standard of thoracotomy, the high sensitivity and close to 100% specificity of EBUS-TBNA make it reasonable to use the calibrated model's predicted prN2/3 as an approximation for the true probability of N2/3 disease. Because EBUS-TBNA is not 100% sensitive, this approximation represents a low estimate of the true probability of N2/3 disease, but at least we know the direction of the bias. This in turn can be used to estimate the posterior probability of disease following a negative EBUS-TBNA, provided we know the sensitivity and specificity of EBUS-TBNA. If this posterior probability is greater than the decision threshold (10% according to European Society of Thoracic Surgery), then it would be reasonable to consider a confirmatory mediastinoscopy.

Note that although the original uncalibrated model can be informative as to which patients may benefit from EBUS-TBNA, because it has good calibration up to

Table 3. Predicted Probabilities of N2/N3 Disease and Estimated Posterior Probabilities of N2/N3 Disease following a Negative EBUS-TBNA in the Model Development Cohort

CT PET N stage	Histology	Location	Predicted prN2/3	Probability of N2/3 Post Negative EBUS*
CT = N0 or N1, PET = N0	Adeno	Central	0.049–0.152	0.006–0.019
CT = N0 or N1, PET = N0	Adeno	Outer two-thirds	0.028–0.090	0.003–0.011
CT = N0 or N1, PET = N0	Squamous	Central	0.022–0.073	0.002–0.009
CT = N0 or N1, PET = N0	Squamous	Outer two-thirds	0.012–0.042	0.001–0.005
CT = N0 or N1, PET = N1	Adeno	Central one-third	0.199–0.464	0.027–0.087
CT = N0 or N1, PET = N1	Adeno	Outer two-thirds	0.122–0.324	0.015–0.050
CT = N0 or N1, PET = N1	Squamous	Central one-third	0.099–0.275	0.012–0.040
CT = N0 or N1, PET = N1	Squamous	Outer two-thirds	0.057–0.174	0.007–0.023
CT = N0 or N1, PET = N2 or N3	Adeno	Central one-third	0.356–0.657	0.057–0.174
CT = N0 or N1, PET = N2 or N3	Adeno	Outer two-thirds	0.235–0.516	0.033–0.105
CT = N0 or N1, PET = N2 or N3	Squamous	Central one-third	0.195–0.457	0.026–0.085
CT = N0 or N1, PET = N2 or N3	Squamous	Outer two-thirds	0.119–0.318	0.015–0.049
CT = N2 or N3, PET = N0	Adeno	Central one-third	0.120–0.320	0.015–0.049
CT = N2 or N3, PET = N0	Adeno	Outer two-thirds	0.070–0.207	0.008–0.028
CT = N2 or N3, PET = N0	Squamous	Central one-third	0.056–0.171	0.007–0.022
CT = N2 or N3, PET = N0	Squamous	Outer two-thirds	0.032–0.103	0.004–0.012
CT = N2 or N3, PET = N1	Adeno	Central one-third	0.113–0.305	0.014–0.046
CT = N2 or N3, PET = N1	Adeno	Outer two-thirds	0.066–0.196	0.008–0.026
CT = N2 or N3, PET = N1	Squamous	Central one-third	0.053–0.162	0.006–0.021
CT = N2 or N3, PET = N1	Squamous	Outer two-thirds	0.030–0.097	0.003–0.012
CT = N2 or N3, PET = N2 or N3	Adeno	Central one-third	0.687–0.884	0.195–0.456
CT = N2 or N3, PET = N2 or N3	Adeno	Outer two-thirds	0.550–0.809	0.118–0.318
CT = N2 or N3, PET = N2 or N3	Squamous	Central one-third	0.491–0.770	0.096–0.269
CT = N2 or N3, PET = N2 or N3	Squamous	Outer two-thirds	0.349–0.650	0.056–0.170

Definition of abbreviations: CT = computed tomography; EBUS-TBNA = endobronchial ultrasound-guided transbronchial needle aspiration; PET = positron emission tomography; prN2/3 = probability of EBUS-TBNA demonstrating N2 or N3 disease.

*Probability of N2/3 post negative EBUS: post-test probability of N2/3 disease following a negative EBUS-TBNA in a patient that had a pretest probability of disease equal to the prN2/3 shown in the column to the left. This is predicated on EBUS sensitivity and specificity being 89% and 100%, respectively, as per American College of Chest Physicians guidelines. Probability ranges are for patients from 40 to 80 years old. Younger patients have lower probabilities of N2/3 disease. For age-specific probabilities see Table E10.

a prN2/3 of 20%, it would not be suitable for informing which patients at an outside center should have mediastinoscopy after a negative EBUS-TBNA. That is because this clinical application requires accurate calibration throughout the whole range of prN2/3, and we have established that the uncalibrated model underestimates prN2/3 at the high end in some high-prevalence

centers. However, with calibration (Figure 3B), the model is sufficiently accurate to help inform the decision as to which patients with a negative EBUS-TBNA might benefit from mediastinoscopy.

The online supplement provides further information on using the model to inform the decision on whether a

mediastinoscopy is warranted following a negative EBUS-TBNA. Although some of the implications of the model are fairly straightforward (e.g., mediastinoscopy is usually warranted in patients with CT N2 PET N2 peripheral adenocarcinoma with a negative EBUS-TBNA) (Table 3), other implications of the model are not so obvious. For instance in a 70 year old with

Table 4. Model Performance at Different Institutions: Predictions Using Full Information without Calibration*

Institution	Prevalence of N2 Disease in that Cohort (%)	AUC ROC (95% CI)	Hosmer-Lemeshow P Value [†]	Brier Score [‡]
MD Anderson Cancer Center (development cohort)	160/633 (25)	0.86 (0.82–0.89)	0.62	0.125
Cleveland Clinic (validation)	87/310 (28)	0.87 (0.83–0.91)	0.03	0.129
Johns Hopkins (validation)	107/186 (58)	0.82 (0.76–0.89)	<0.001	0.181
Henry Ford Hospital (validation)	102/226 (45)	0.92 (0.88–0.95)	<0.001	0.139

Definition of abbreviations: AUC ROC = area under the receiver operating characteristic curve; CI = confidence interval.

*Full model as specified in Table 2 (age, location, histology, and computed tomography/positron emission tomography interaction). Note that the Hosmer-Lemeshow and Brier scores use predictions from the uncalibrated model in this table.

[†]Hosmer-Lemeshow test, $P < 0.05$ indicates poor calibration.

[‡]Brier scores reflect the mean squared difference between predicted outcomes and the actual outcomes. Brier scores range from 0 to 1, with lower scores being better.

Figure 3. External validation cohort, observed versus predicted frequencies. The model uses positron emission tomography/computed tomography N stage, age, location (central one-third vs. outer two-thirds), and histology. The figure plots the probability of N2/N3 disease by decile of expected risk as a function of the actual observed risk in that group. This includes an interaction term for positron emission tomography/computed tomography. The observed probability for each decile is on the *vertical axis*; the predicted probability according to decile is on the *horizontal axis*. A perfect model, where observed = predicted, is shown by the *line*. (A) The uncalibrated model using data from the development cohort to predict outcomes in the external validation cohort. Note that although accurate at low probabilities (<20%), it is less accurate at high probabilities. (B) The calibrated model using the general calibration method (17). The *line* shows that the observed frequency of N2/N3 disease was similar to the predicted probability.

CT N2 PET N2 peripheral squamous cell carcinoma with a negative EBUS-TBNA the prN2/3 is only 7% (see Table E10). So the model provides additional insight to further inform the decision.

Our model is analogous to clinical prediction models used to calculate the probability of cancer in pulmonary nodules (7, 39–43). Pulmonary nodule prediction rules are widely available and have been integrated into guidelines and can even be

downloaded as apps for mobile phones. Although some nodule studies have used cross-validation, few studies have externally validated the original models (39, 40). The most rigorous studies evaluated the VA Cooperative and Mayo models and found a fair to good AUC (0.73 and 0.80, respectively) when applied to external cohorts (39). However, calibration curves showed that the Mayo model underestimated probability of

malignancy, whereas the VA model overestimated.

Compared with pulmonary nodule models, our model demonstrated better discrimination, with an AUC of 0.88 in the external validation cohort, but we had similar problems with underestimation of observed probabilities, which required model calibration. Possible reasons that would account for local variations in model calibration include differences in patient populations and variability in technology performance and interpretation. Examples of differences in patient populations would include the prevalence of background diseases (e.g., histoplasmosis) that might cause false-positive PET scans. Similarly, variability in PET performance or interpretation between centers may impact model calibration (16, 30, 44, 45). Although the specific reasons underlying variability remain unknown, it is known that such factors as time to FDG injection, type of reconstruction, filter, attenuation, and scan length can impact standardized uptake value measurements (44, 45). Indeed, prior metaanalysis has demonstrate that PET sensitivity and specificity vary significantly between studies (interquartile range, 67–91% and 82–96%, respectively); the same is true for CT (interquartile range, 50–71% and 66–89%, respectively) (16). Thus, because regional population differences will always be present, and because there will probably always be subtle but important differences in PET performance between centers, local calibration will probably always be necessary for any model to be applicable in the clinical arena.

This serves to emphasize the importance of external validation to assess absolute risk prediction as compared with looking solely at relative risk prediction. In this study we used the general calibration model method to calibrate our model (17). Note this method does not change the model's performance in terms of discrimination because it is just a linear transformation of the log odds. But the absolute predicted probabilities do change. To our knowledge, no prior prediction model of either pulmonary nodules (7, 39–42) or mediastinal nodal metastasis (21–28) has applied this method.

We believe this method of calibration is important because it provides a practical solution that could be widely implemented. This applies not only to our model, but to

Table 5. Model Performance When Used on External Validation Cohort for Different Clinical Scenarios

Model	AUC ROC Validation Cohort (95% CI)	Hosmer-Lemeshow Test <i>P</i> Value*	Brier Score†
Full model, with interaction terms‡	0.88 (0.85–0.91)	0.54	0.132
No PET	0.77 (0.74–0.80)	0.39	0.190
No histology	0.87 (0.85–0.90)	0.38	0.135
No PET, no histology	0.76 (0.73–0.79)	0.68	0.193

Definition of abbreviations: AUC ROC = area under the receiver operating characteristic curve; CI = confidence interval; PET = positron emission tomography

Hosmer-Lemeshow and Brier scores are for calibrated model using the general calibration method (17). For model performance without calibration, see Tables E5 (No PET), E7 (No histology), and E9 (No PET, no histology).

*Hosmer-Lemeshow test, *P* < 0.05 indicates poor calibration.

†Brier scores reflect the mean squared difference between predicted outcomes and the actual outcomes. Brier scores range from 0 to 1, with lower scores being better.

‡Full model as specified in Table 2.

other models as well, such as those for pulmonary nodules. The factors listed in our model (age, PET-CT N stage, central location, and histology) are readily captured through the electronic health record. A relatively modest amount of data collected through the electronic health record would allow the prediction rule to be calibrated to local patterns, allowing the model to adapt and learn (18). This would provide patients and their physicians with improved decision support tools that would be meaningful at the local level.

Finally, this study should prompt reconsideration of current guidelines regarding which tests are most suitable for initial mediastinal evaluation. Current ACCP lung cancer guidelines recommend EBUS-TBNA, endoscopic ultrasound-needle aspiration (EUS-NA), or a combined procedure as the first test in evaluating patients with a sufficiently high probability of mediastinal metastases (4). Looking at the N2 and N3 nodes identified in this study (see Table E2), it is clear that in many cases EUS-NA would not have been able to reach the nodes involved. Stations 4R, 2R, 11RS, and 11L accounted for 44% of all positive N2 or N3 nodes and these cannot be routinely accessed with EUS-NA. We therefore suggest that the guidelines be

refined, to indicate that EUS-NA by itself should not routinely be considered a suitable first test for mediastinal staging. EBUS-TBNA or combined EBUS-TBNA/EUS-NA would be the best options.

Although the data provided are practical and extend the work of prior investigators, it is important to recognize the limitations of this study. Although this was a multicenter study and the model was externally validated, it is a retrospective study and it therefore only includes patients who were referred for evaluation. Because of this it is difficult to judge how well the model would perform in patients who were not referred. Model performance in such groups is partially addressed by our retrospective validation of the model in three other centers. The finding that although discrimination was good, calibration was off in these other centers suggests that a combination of selection bias and residual confounding within strata may be impacting model performance (see online supplement for a detailed description of selection bias and residual confounding). Although this improved with use of the general calibration model, it still requires future prospective validation. In addition, the accuracy of mediastinal staging with EBUS-TBNA

depends on technical factors of how the procedure is performed (46). Our prediction rule is only externally valid in centers that perform EBUS-TBNA in a similar systematic manner, sampling all lymph nodes that are 5 mm or more in size, moving from N3 to N2 to N1. In addition, all of the centers involved in this study are high-volume centers, a characteristic that has been associated with higher diagnostic yield (15). Results might be different in other centers.

Conclusions

This multicenter study developed a parsimonious clinical prediction model (HAL) for calculating prN2/3 in patients with NSCLC. Younger age, central location, adenocarcinoma histology, and higher PET-CT N stage were associated with increased prN2/3 disease. The HAL model was externally validated and proved to have good discrimination. A general calibration model was used to successfully improve model calibration (17). The HAL model has the potential to facilitate clinical decision making in the staging of NSCLC when calibrated to local patterns. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

- Ramnath N, Dilling TJ, Harris LJ, Kim AW, Michaud GC, Balekian AA, Diekemper R, Detterbeck FC, Arenberg DA. Treatment of stage III non-small cell lung cancer: diagnosis and management of lung cancer, 3rd ed.: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143:e314S–340S.
- Ost DE, Yeung SC, Tanoue LT, Gould MK. Clinical and organizational factors in the initial evaluation of patients with lung cancer: diagnosis and management of lung cancer, 3rd ed.: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143:e121S–141S.
- Almeida FA, Uzbeck M, Ost D. Initial evaluation of the nonsmall cell lung cancer patient: diagnosis and staging. *Curr Opin Pulm Med* 2010;16:307–314.
- Silvestri GA, Gonzalez AV, Jantz MA, Margolis ML, Gould MK, Tanoue LT, Harris LJ, Detterbeck FC. Methods for staging non-small cell lung cancer: diagnosis and management of lung cancer, 3rd ed.: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143:e211S–250S.
- Silvestri GA, Gould MK, Margolis ML, Tanoue LT, McCrory D, Toloza E, Detterbeck F. Noninvasive staging of non-small cell lung cancer: ACCP evidenced-based clinical practice guidelines (2nd edition). *Chest* 2007;132:178S–201S.
- De Leyn P, Dooms C, Kuzdzal J, Lardinois D, Passlick B, Rami-Porta R, Turna A, Van Schil P, Venuta F, Waller D, et al. Revised ESTS guidelines for preoperative mediastinal lymph node staging for non-small-cell lung cancer. *Eur J Cardiothorac Surg* 2014;45:787–798.
- Gould MK, Ananth L, Barnett PG; Veterans Affairs SNAP Cooperative Study Group. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest* 2007;131:383–388.
- Gould MK, Donington J, Lynch WR, Mazzone PJ, Midhun DE, Naidich DP, Wiener RS. 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:e93S–120S.
- Ost D, Fein A. Evaluation and management of the solitary pulmonary nodule. *Am J Respir Crit Care Med* 2000;162:782–787.
- Ost D, Fein AM, Feinsilver SH. Clinical practice: the solitary pulmonary nodule. *N Engl J Med* 2003;348:2535–2542.
- Ost DE, Gould MK. Decision making in patients with pulmonary nodules. *Am J Respir Crit Care Med* 2012;185:363–372.
- Ost D, Simoff M, Feller-Kopman D, Kovitz K, Eapen G, Greenhill S, Michaud G, Majid A, HeyM J, Jimenez C, et al. Risk adjusted diagnostic yield and outcomes for endobronchial ultrasound guided TBNA: results of the ACCP AQuIRE Registry. *Chest* 2009;136:69S.
- Ost DE, Ernst A, Lei X, Kovitz KL, Benzaquen S, Diaz-Mendoza J, Greenhill S, Toth J, Feller-Kopman D, Puchalski J, et al.; AQuIRE Bronchoscopy Registry. Diagnostic yield and complications of bronchoscopy for peripheral lung lesions: results of the AQuIRE Registry. *Am J Respir Crit Care Med* 2016;193:68–77.
- Eapen GA, Shah AM, Lei X, Jimenez CA, Morice RC, Yarmus L, Filner J, Ray C, Michaud G, Greenhill SR, et al.; American College of Chest Physicians Quality Improvement Registry, Education. Complications,

- consequences, and practice patterns of endobronchial ultrasound-guided transbronchial needle aspiration: results of the AQUIRE registry. *Chest* 2013;143:1044–1053.
15. Ost DE, Ernst A, Lei X, Feller-Kopman D, Eapen GA, Kovitz KL, Herth FJ, Simoff M, Registry AQB; AQUIRE Bronchoscopy Registry. Diagnostic yield of endobronchial ultrasound-guided transbronchial needle aspiration: results of the AQUIRE Bronchoscopy Registry. *Chest* 2011;140:1557–1566.
 16. Gould MK, Kuschner WG, Rydzak CE, Maclean CC, Demas AN, Shigemitsu H, Chan JK, Owens DK. Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: a meta-analysis. *Ann Intern Med* 2003;139:879–892.
 17. Steyerberg E. Evaluation of performance. Clinical prediction models: a practical approach to development, validation, and updating. 1st ed. New York: Springer-Verlag; 2009. p. 273.
 18. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med* 2004;23:2567–2586.
 19. Ost DE, Niu J, S Elting L, Buchholz TA, Giordano SH. Quality gaps and comparative effectiveness in lung cancer staging and diagnosis. *Chest* 2014;145:331–345.
 20. Almeida FA, Casal RF, Jimenez CA, Eapen GA, Uzbeck M, Sarkiss M, Rice D, Morice RC, Ost DE. Quality gaps and comparative effectiveness in lung cancer staging: the impact of test sequencing on outcomes. *Chest* 2013;144:1776–1782.
 21. Shafazand S, Gould MK. A clinical prediction rule to estimate the probability of mediastinal metastasis in patients with non-small cell lung cancer. *J Thorac Oncol* 2006;1:953–959.
 22. Chen K, Yang F, Jiang G, Li J, Wang J. Development and validation of a clinical prediction model for N2 lymph node metastasis in non-small cell lung cancer. *Ann Thorac Surg* 2013;96:1761–1768.
 23. Tsutani Y, Murakami S, Miyata Y, Nakayama H, Yoshimura M, Okada M. Prediction of lymph node status in clinical stage IA squamous cell carcinoma of the lung. *Eur J Cardiothorac Surg* 2015;47:1022–1026.
 24. Tsutani Y, Miyata Y, Nakayama H, Okumura S, Adachi S, Yoshimura M, Okada M. Prediction of pathologic node-negative clinical stage IA lung adenocarcinoma for optimal candidates undergoing sublobar resection. *J Thorac Cardiovasc Surg* 2012;144:1365–1371.
 25. Park SY, Yoon JK, Park KJ, Lee SJ. Prediction of occult lymph node metastasis using volume-based PET parameters in small-sized peripheral non-small cell lung cancer. *Cancer Imaging* 2015;15:21.
 26. Takenaka T, Yano T, Morodomi Y, Ito K, Miura N, Kawano D, Shoji F, Baba S, Abe K, Honda H, et al. Prediction of true-negative lymph node metastasis in clinical IA non-small cell lung cancer by measuring standardized uptake values on positron emission tomography. *Surg Today* 2012;42:934–939.
 27. Zhang Y, Sun Y, Xiang J, Zhang Y, Hu H, Chen H. A prediction model for N2 disease in T1 non-small cell lung cancer. *J Thorac Cardiovasc Surg* 2012;144:1360–1364.
 28. Koike T, Koike T, Yamato Y, Yoshiya K, Toyabe S. Predictive risk factors for mediastinal lymph node metastasis in clinical stage IA non-small-cell lung cancer patients. *J Thorac Oncol* 2012;7:1246–1251.
 29. Farjah F, Lou F, Sima C, Rusch VW, Rizk NP. A prediction model for pathologic N2 disease in lung cancer patients with a negative mediastinum by positron emission tomography. *J Thorac Oncol* 2013;8:1170–1180.
 30. Farjah F, Backhus LM, Varghese TK, Manning JP, Cheng AM, Mulligan MS, Wood DE. External validation of a prediction model for pathologic N2 among patients with a negative mediastinum by positron emission tomography. *J Thorac Dis* 2015;7:576–584.
 31. Yang F, Chen H, Xiang J, Zhang Y, Zhou J, Hu H, Zhang J, Luo X. Relationship between tumor size and disease stage in non-small cell lung cancer. *BMC Cancer* 2010;10:474.
 32. Ketchedjian A, Daly BD, Fernando HC, Florin L, Hunter CJ, Morelli DM, Shemin RJ. Location as an important predictor of lymph node involvement for pulmonary adenocarcinoma. *J Thorac Cardiovasc Surg* 2006;132:544–548.
 33. Lee PC, Port JL, Korst RJ, Liss Y, Meherally DN, Altorki NK. Risk factors for occult mediastinal metastases in clinical stage I non-small cell lung cancer. *Ann Thorac Surg* 2007;84:177–181.
 34. Kanzaki R, Higashiyama M, Fujiwara A, Tokunaga T, Maeda J, Okami J, Kozuka T, Hosoki T, Hasegawa Y, Takami M, et al. Occult mediastinal lymph node metastasis in NSCLC patients diagnosed as clinical N0-1 by preoperative integrated FDG-PET/CT and CT: risk factors, pattern, and histopathological study. *Lung Cancer* 2011;71:333–337.
 35. Cerfolio RJ, Bryant AS, Eloubeidi MA. Routine mediastinoscopy and esophageal ultrasound fine-needle aspiration in patients with non-small cell lung cancer who are clinically N2 negative: a prospective study. *Chest* 2006;130:1791–1795.
 36. Takamochi K, Nagai K, Suzuki K, Yoshida J, Ohde Y, Nishiwaki Y. Clinical predictors of N2 disease in non-small cell lung cancer. *Chest* 2000;117:1577–1582.
 37. Al-Sarraf N, Aziz R, Gately K, Lucey J, Wilson L, McGovern E, Young V. Pattern and predictors of occult mediastinal lymph node involvement in non-small cell lung cancer patients with negative mediastinal uptake on positron emission tomography. *Eur J Cardiothorac Surg* 2008;33:104–109.
 38. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–138.
 39. Schultz EM, Sanders GD, Trotter PR, Patz EF Jr, Silvestri GA, Owens DK, Gould MK. Validation of two models to estimate the probability of malignancy in patients with solitary pulmonary nodules. *Thorax* 2008;63:335–341.
 40. Herder GJ, van Tinteren H, Golding RP, Kostense PJ, Comans EF, Smit EF, Hoekstra OS. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. *Chest* 2005;128:2490–2496.
 41. Swensen SJ, Silverstein MD, Edell ES, Trastek VF, Aughenbaugh GL, Ilstrup DM, Schleck CD. Solitary pulmonary nodules: clinical prediction model versus physicians. *Mayo Clin Proc* 1999;74:319–329.
 42. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med* 1997;157:849–855.
 43. Tammemagi MC, Freedman MT, Pinsky PF, Oken MM, Hu P, Riley TL, Ragard LR, Berg CD, Prorok PC. Prediction of true positive lung cancers in individuals with abnormal suspicious chest radiographs: a prostate, lung, colorectal, and ovarian cancer screening trial study. *J Thorac Oncol* 2009;4:710–721.
 44. Toney LK, Vesselle HJ. Neural networks for nodal staging of non-small cell lung cancer with FDG PET and CT: importance of combining uptake values and sizes of nodes and primary tumor. *Radiology* 2014;270:91–98.
 45. Hasenclever D, Kurch L, Kluge R. Sources of variability in FDG PET imaging and the qPET value: reply to Laffon and Marthan. *Eur J Nucl Med Mol Imaging* 2014;41:2155–2157.
 46. Dettmerbeck F, Puchalski J, Rubinowitz A, Cheng D. Classification of the thoroughness of mediastinal staging of lung cancer. *Chest* 2010;137:436–442.