# External validation of a prediction model for pathologic N2 among patients with a negative mediastinum by positron emission tomography

# Farhood Farjah<sup>1,2</sup>, Leah M. Backhus<sup>1</sup>, Thomas K. Varghese<sup>1</sup>, James P. Manning<sup>2</sup>, Aaron M. Cheng<sup>1</sup>, Michael S. Mulligan<sup>1</sup>, Douglas E. Wood<sup>1</sup>

<sup>1</sup>Division of Cardiothoracic Surgery, <sup>2</sup>Surgical Outcomes Research Center, University of Washington, Seattle, USA *Correspondence to:* Farhood Farjah, MD, MPH. Division of Cardiothoracic Surgery, University of Washington, 1959 NE Pacific St, Box 356310, Seattle, Washington 98195, USA. Email: ffarjah@uw.edu.

**Background:** A prediction model for pathologic N2 (pN2) among lung cancer patients with a negative mediastinum by positron emission tomography (PET) was recently internally validated. Our study sought to determine the external validity of that model.

**Methods:** A cohort study [2005-2013] was performed of lung cancer patients with a negative mediastinum by PET. Previously published model coefficients were used to estimate the probability of pN2 based on tumor location and size, nodal enlargement by computed tomography (CT), maximum standardized uptake value (SUV<sub>max</sub>) of the primary tumor, N1 disease by PET, and pretreatment histology.

**Results:** Among 239 patients, 18 had pN2 [7.5%, 95% confidence interval (CI): 4.5-12%]. Model discrimination was excellent (c-statistic 0.80, 95% CI: 0.75-0.85) and the model fit the data well (P=0.191). The accuracy of the model was as follows: sensitivity 100%, 95% CI: 81-100%; specificity 49%, 95% CI: 42-56%; positive predictive value (PPV) 14%, 95% CI: 8-21%, and negative predictive value (NPV) 100%, 95% CI: 97-100%. CI inspection revealed a significantly higher c-statistic in this external validation cohort compared to the internal validation cohort. The model's apparently poor specificity for patient selection is in fact significantly better than usual care (i.e., aggressive but allowable guideline concordant staging) and minimum guideline mandated selection criteria for invasive staging.

**Conclusions:** A prediction model for pN2 is externally valid. The high NPV of this model may allow pulmonologists and thoracic surgeons to more comfortably minimize the number of invasive procedures performed among patients with a negative mediastinum by PET.

Keywords: Lung neoplasms; positron emission tomography (PET); mediastinum; neoplasm staging

Submitted Nov 06, 2014. Accepted for publication Jan 26, 2015. doi: 10.3978/j.issn.2072-1439.2015.02.09 View this article at: http://dx.doi.org/10.3978/j.issn.2072-1439.2015.02.09

### Introduction

Mediastinal nodal evaluation is the cornerstone of staging patients with non-metastatic non-small cell lung cancer (NSCLC). Practice guidelines recommend routine invasive mediastinal staging for patients with a positive mediastinum by positron emission tomography (PET) (1,2). When the mediastinum is negative by PET, guidelines recommend selective invasive staging based on the presence or absence of other radiographic risk factors for mediastinal nodal disease—for example, tumor size and/or location, mediastinal lymphadenopathy, or fluorodeoxyglucose (FDG) uptake in N1 nodes. Because there are multiple risk factors for unsuspected N2 (3-9), there is an opportunity to use risk-prediction to facilitate better patient selection for invasive diagnostic procedures.

Several groups have developed prediction models to improve patient selection for invasive mediastinal staging (10-14). One group of investigators recently developed and internally validated a prediction model for pathologic N2 (pN2) using six pre-operative risk-factors among North American patients with a negative mediastinum by PET (14). An important limitation of this work and other contemporary studies is that validation occurred at only one institution—specifically the same institution where the model was developed. The goal of this investigation was to evaluate the performance of the previously internally validated prediction model for pN2 at another North American site. A secondary aim was to explore the risk-prediction model's potential impact on care.

#### **Materials and methods**

A retrospective investigation (November 2005-March 2013) was conducted of NSCLC patients with a negative mediastinum by PET who underwent nodal evaluation by invasive staging and/or at the time of pulmonary resection at the University of Washington. The Institutional Review Board approved this study and waived the need for consent for this minimal risk retrospective review (committee EA, approval number 44939). Study inclusion/exclusion criteria matched those of a prior North American study that originally developed and internally validated the prediction model under investigation (14). Patients were eligible for study if they were initially staged by computed tomography (CT) and PET and had no evidence of metastatic disease. Those ineligible for study included: suspicion of synchronous, metachronous, or recurrent lung cancer; evidence of T3 (except by size criteria) or T4 tumors by CT; and/or receipt of induction therapy without pathologic confirmation of nodal disease. The American Joint Committee on Cancer classification for T-status changed over the study period such that tumors greater than 7 cm were considered T3 rather than T2 after 2010. Because the population in which the prediction model was originally developed included patients regardless of size, we allowed for inclusion of subjects with T3 tumors defined exclusively by size criteria.

The previously published North American prediction model utilized logistic regression to estimate the probability of pN2 based on six risk factors ascertained prior to treatment—tumor size by CT, tumor location by CT (central versus peripheral), extent of nodal disease by CT (lymphadenopathy defined by size greater than 1 cm), maximum standardized uptake value (SUV<sub>max</sub>) of the primary tumor, FDG uptake in ipsilateral hilar nodes, and tumor histology if the patient underwent a pretreatment biopsy of the primary lesion (14). Coefficient estimates from this model were used to estimate the probability of pN2 for each patient. Like the original prediction model study (14), radiology and pathology reports were used to ascertain model inputs and the true status of mediastinal lymph nodes—defined by pathologic confirmation of disease in nodal tissue.

Model performance was evaluated using the same metrics utilized by the internal validation study (14)discrimination, calibration (fit), and accuracy. A c-statistic describes a model's ability to discriminate between two outcomes. The metric ranges from 0.5 (no different than a coin-toss) to 1.0 (perfect discrimination among two outcomes). Values greater than 0.7 indicate reasonable discriminatory ability, and values greater than 0.8 indicate strong discriminatory ability (15). A non-significant goodness-of-fit test indicates good fit. Visual evaluation of calibration (fit) was also performed by plotting actual versus predicted pN2 rates. In order to describe the model's accuracy for patient selection in terms of sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), patients must be dichotomized into high-versus low-risk groups based on a threshold probability. The original prediction model used an empirical method called the Youden Index to identify the cut-off that maximizes the sum of sensitivity and specificity (16). A probability of pN2 disease >8.3% defined the high-risk group in the original prediction model study (14), and this threshold was used for the current investigation.

A secondary hypothesis-generating aim of this study was to explore the potential impact of the prediction model on care by comparing the frequency and accuracy of patient selection for invasive staging across varying assumptions (Tables S1-S3). Usual care at the University of Washington historically consisted of an aggressive but guideline allowable staging strategy (1,2) of invasive mediastinal staging for all patients except for those with a relative contraindication and/or those in whom management would not be expected to change based on mediastinal nodal status. We did not use the prediction model as a part of usual care because it was not available until 2012 and because we considered it investigational without evidence of external validity. The accuracy of patient selection with usual care was calculated using a  $2 \times 2$  table of patients who actually underwent invasive staging. Minimum practice guideline recommendations for selective staging essentially require all patients to undergo invasive staging except those with a radiographic, peripheral stage IA tumor (1,2). The accuracy of patient selection using the minimum recommended standards for invasive staging was estimated using a  $2\times 2$  table after classifying patients as having a peripheral stage IA or not. The accuracy of patient selection using the risk-prediction model was estimated using a  $2\times 2$  table after classifying patients as being high- or low-risk using the probability cut-off described earlier.

STATA (Special Edition 12.1; Statacorp, College Station, Texas, USA) was used to conduct all statistical analyses. Binary variables were summarized using proportions and 95% exact binomial confidence intervals (CI).

### Results

Table 1 provides an overall summary of the cohort. Nearly every patient underwent invasive staging with a median of 3 (range, 0-5) mediastinal lymph node stations sampled. Two patients with pN2 identified by invasive staging did not undergo operative management because one had singlestation bulky disease and the other opted for definitive chemoradiation therapy. Intra-operatively, a median of 1 (range, 0-8) mediastinal lymph node station was sampled. The overall median number of mediastinal lymph node stations sampled by any means (invasive staging or intraoperatively) was 3 (range, 1-9), with 86% of patients having at least three mediastinal nodal stations sampled. A total of 18 patients had pN2 (7.5%, 95% CI: 4.5-12%)-all with single-station disease. Table 2 shows the distribution of risk factors and the prevalence of pN2 across each risk factor. The median primary tumor size by CT and  $SUV_{max}$  were 2.5 cm and 5.6, respectively. Patients with FDG uptake in N1 nodes had the highest prevalence of pN2 disease (25%).

Table 3 summarizes model performance for the external validation cohort as well as the reported performance of the original development and validation cohorts. Model discrimination was excellent in our external validation cohort and significantly higher than that reported in the internal validation cohort (Figure 1). The model fit the data well across all cohorts (P>0.05). Figure 2 reveals that the model tended to overestimate the risk of pN2 in the external validation cohort. Model sensitivity, specificity, PPV, and NPV were not significantly different across cohorts. A post-hoc analysis was conducted among patients with at least three nodal stations sampled by any means (n=205, pN2 =7.3%). Model performance was similar to that observed in our primary analysis [c-statistic 0.80 (95% CI: 0.73-0.85), goodness-of-fit test P=0.50, sensitivity 100% (95% CI: 78-100), specificity 46% (95% CI: 3954%), PPV 13% (95% CI: 7.3-20%), and NPV 100% (95% CI: 96-100%)]. An additional post-hoc analysis was conducted among patients with at least three nodal stations sampled intraoperatively (n=58, pN2 =12%). Again model performance was similar to that observed in our primary analysis except that the PPV was significantly higher [c-statistic 0.78 (95% CI: 0.65-0.87), goodness-of-fit test P=0.62, sensitivity 100% (95% CI: 59-100%), specificity 59% (95% CI: 44-72%), PPV 72% (95% CI: 53-87%), and NPV 100% (95% CI: 88-100%)].

Table 4 summarizes the frequency and accuracy of patient selection for invasive staging across varying assumptions. Usual care—representing an aggressive but guideline allowable invasive staging strategy—was characterized by the highest utilization of invasive procedures and lowest accuracy in terms of patient selection. Had the minimum practice guideline recommended indications for invasive staging been adhered to, use of invasive procedures would have been significantly less, and the specificity and NPV of patient selection would have been significantly higher. Compared to both usual care and the minimum recommended indications for invasive staging, use of the prediction model would have resulted in even less use of invasive diagnostic tests and further improvements in the specificity of patient selection.

# Discussion

The clinical application of a prediction model for nodal disease in operable lung cancer patients is to guide the use of invasive staging procedures prior to first treatment. Patients at high-risk for nodal disease will still require invasive staging in order to characterize the extent of disease (i.e., N1, single-station N2, multi-station N2, N3). However, patients at low-risk may proceed directly to resection with intraoperative nodal evaluation. A few prediction models have been developed for mediastinal staging among NSCLC patients (10-14), but none have been externally validated. We report that a previously developed and internally validated prediction model for pN2 among North American patients with a negative mediastinum by PET is externally valid. The potential impact of such a model is a reduction in the use of invasive mediastinal staging procedures and improvements in patient selection for invasive diagnostic tests.

External validation revealed that model performance was at least as good as that observed in the original development and internal validation cohorts (14). Although

# Journal of Thoracic Disease, Vol 7, No 4 April 2015

| Table 1 Cohort characteristics (n=239)    |             |  |
|---|-------------|--|
| Characteristics                           |             |  |
| Median age, years [range]                 | 67 [24-92]  |  |
| Male (%)                                  | 49.0        |  |
| Race (%)                                  |             |  |
| White                                     | 87.0        |  |
| Black                                     | 2.9         |  |
| Asian/Pl                                  | 6.7         |  |
| Other                                     | 2.9         |  |
| Missing                                   | 0.4         |  |
| Comorbidity* (%)                          |             |  |
| HTN                                       | 55.0        |  |
| CAD                                       | 18.0        |  |
| CHF                                       | 4.2         |  |
| CVD                                       | 6.7         |  |
| PVD                                       | 3.8         |  |
| COPD                                      | 15.5        |  |
| IPF                                       | 1.3         |  |
| Pulmonary HTN                             | 0.8         |  |
| DM  | 15.1        |  |
| Dialysis                                  | 0.8         |  |
| Steroids                                  | 4.6         |  |
| ASA (%)                                   |             |  |
| - I                                       | 1.3         |  |
| П   | 11.0        |  |
| III                                       | 83.0        |  |
| IV  | 4.2         |  |
| Smoking status (%)                        |             |  |
| Current                                   | 63.0        |  |
| Former                                    | 22.0        |  |
| Never                                     | 15.0        |  |
| Median pack-years [range]                 | 30 [0-135]  |  |
| Median predicted FEV <sub>1</sub> [range] | 79 [28-135] |  |
| Median predicted DLCO [range]             | 65 [23-122] |  |
| Prior CTS surgery (%)                     | 13.0        |  |
| Invasive staging* (%)                     |             |  |
| Mediastinoscopy                           | 97.0        |  |
| Mediastinotomy                            | 0.4         |  |
| EBUS                                      | 3.5         |  |
| None                                      | 3.3         |  |
| Table 1 (continued)                       |             |  |

| Table 1 (continued)          |      |  |
|------------------------------|------|--|
| Characteristics              |      |  |
| Resection (%)                |      |  |
| Lobectomy                    | 77.0 |  |
| Wedge                        | 13.0 |  |
| Segmentectomy                | 1.7  |  |
| Sleeve lobectomy             | 0.8  |  |
| Bilobectomy                  | 2.1  |  |
| Pneumonectomy                | 3.4  |  |
| None                         | 0.8  |  |
| Approach (%)                 |      |  |
| Thoracotomy                  | 54.0 |  |
| VATS                         | 44.0 |  |
| Robotic                      | 1.3  |  |
| No resection                 | 0.8  |  |
| Histology (%)                |      |  |
| Adenocarcinoma               | 60.0 |  |
| Squamous                     | 26.0 |  |
| Large cell/NSCLC NOS         | 7.5  |  |
| AIS                          | 5.0  |  |
| Small cell                   | 0.8  |  |
| Other                        | 0.7  |  |
| Grade (%)                    |      |  |
| Well                         | 19.0 |  |
| Moderate                     | 47.0 |  |
| Poor/undifferentiated        | 28.0 |  |
| Missing                      | 5.4  |  |
| Stage (%)                    |      |  |
| I                            | 74.0 |  |
| II                           | 15.0 |  |
| IIIA                         | 8.8  |  |
| IIIB                         | 2.1  |  |
| Prolonged length-of-stay (%) | 6.3  |  |
| 30-day mortality (%)         | 1.3  |  |

\*, columns may not add to 100% because patients may have had more than 1 comorbid condition and/or more than 1 type of invasive mediastinal staging procedure. PI, Pacific Islander; HTN, hypertension; CAD, coronary artery disease; CHF, congestive heart failure; CVD, cerebrovascular disease; PVD, peripheral vascular disease; COPD, chronic obstructive pulmonary disease; IPF, interstitial pulmonary fibrosis; DM, diabetes mellitus; ASA, American Society of Anesthesiologists; FEV<sub>1</sub>, forced expiratory volume in 1 second; DLCO, diffusion lung capacity for carbon monoxide; CTS, cardiothoracic surgery; EBUS, endobronchial ultrasound-guided nodal aspiration; VATS, video-assisted thoracic surgery; NSCLC, non-small cell lung cancer; NOS, not otherwise specified, AIS, adenocarcinoma-in-situ.

# Farjah et al. Predicting N2 among lung cancer patients

| Table 2 Prevalence of risk factors and risk of pN2 (n=239)  |                                   |                            |  |
|---|-----------------------------------|----------------------------|--|
|   | Frequency of risk factor (95% CI) | Prevalence of pN2 (95% CI) |  |
| Primary tumor size, cm  |                                   |                            |  |
| <1.0  | 5.0% (2.6-8.6%)                   | 0.0% (0-26%)               |  |
| 1.0-2.0   | 33% (27-39%)                      | 1.3% (0.0-6.9%)            |  |
| 2.1-3.0   | 26% (20-32%)                      | 6.6% (1.8-16.0%)           |  |
| 3.1-5.0   | 26% (20-31%)                      | 6.7% (1.8-16.0%)           |  |
| 5.1-7.0   | 7.0% (3.9-11.0%)                  | 13% (5.9-25.0%)            |  |
| >7.0  | 5.0% (2.6-8.6%)                   | 6.3% (0.2-30.0%)           |  |
| Primary tumor location  |                                   |                            |  |
| Peripheral  | 30% (24-36%)                      | 2.8% (3.4-9.8%)            |  |
| Central   | 70% (64-76%)                      | 9.5% (5.5-15.0%)           |  |
| SUV <sub>max</sub>  |                                   |                            |  |
| ≤1.9  | 10% (6.9-15.0%)                   | 4.0% (0.1-20.0%)           |  |
| 2.0-3.4   | 18% (13-23%)                      | 0.0% (0.0-8.2%)            |  |
| 3.5-5.9   | 24% (19-30%)                      | 8.6% (2.9-19.0%)           |  |
| 6.0-10.4  | 21% (16-27%)                      | 9.8% (3.2-21.0%)           |  |
| ≥10.5   | 26% (21-32%)                      | 11% (4.7-22.0%)            |  |
| N-status (lymphadenopathy) by CT  |                                   |                            |  |
| NO  | 79% (73-84%)                      | 5.3% (2.6-9.5%)            |  |
| N1  | 9.6% (6.2-14.0%)                  | 22% (7.4-44.0%)            |  |
| N2/N3   | 11% (7.5-16.0%)                   | 11% (2.4-29.0%)            |  |
| N1 (FDG-uptake) by PET  |                                   |                            |  |
| NO  | 88% (84-92%)                      | 5.2% (2.6-9.1%)            |  |
| N1  | 12% (7.9-16.0%)                   | 25% (11-45%)               |  |
| Pretreatment histology  |                                   |                            |  |
| No biopsy/other   | 69% (63-75%)                      | 7.2% (3.8-12.0%)           |  |
| Adenocarcinoma  | 20% (15-26%)                      | 8.3% (2.3-20.0%)           |  |
| Adenocarcinoma w/AIS  | 0.4% (0.0-2.3%)                   | 0.0% (0-98%)               |  |
| Squamous  | 10% (6.9-15.0%)                   | 5.4% (1.0-26.0%)           |  |
| SUV, standardized uptake value; CT, computed tomography; FDG, fluorodeoxyglucose; PET, positron emission tomography; AIS, |                                   |                            |  |

adenocarcinoma-in-situ; CI, confidence interval.

| Table 3 Model performance            |   |   |  |
|--------------------------------------|---|---|--|
|                                      | Current study, external validation cohort (n=239) | Prior study (14),<br>development cohort (n=625) | Prior study (14), internal validation cohort (n=313) |
| c-statistic (95% CI)                 | 0.80 (0.75-0.85)                                  | 0.70 (0.63-0.77)                                | 0.65 (0.56-0.74)                                     |
| Goodness-of-fit test (P value)       | 0.191   | 0.560   | 0.190  |
| Performance characteristics (95% CI) |   |   |  |
| Sensitivity                          | 100% (81-100%)                                    | 77% (64-86%)                                    | 86% (68-96%)   |
| Specificity                          | 49% (42-56%)                                      | 54% (50-58%)                                    | 48% (42-54%)   |
| Positive predictive value            | 14% (8.3-21.0%)                                   | 16% (12-21%)                                    | 15% (10-21%)   |
| Negative predictive value            | 100% (97-100%)                                    | 95% (92-97%)                                    | 97% (93-99%)   |
| CI, confidence interval.             |   |   |  |



Figure 1 Receiver operating curve.



**Figure 2** Calibration plot. The x-axis represents predicted rates of pN2 and the y-axis represents actual rates of pN2. Patients were divided into 3 groups based on terciles of predicted pN2. Each group's average predicted pN2 rate was plotted against their average actual rate of pN2. The dashed 45-degree line represents (as a reference) perfect concordance between actual and predicted rates. Vertical bars represent 95% CI for each group's average actual pN2 rate. pN2, pathologic N2; CI, confidence interval.

model discrimination was significantly better in our external validation cohort, superior performance in this metric did not translate into observable improvements in the accuracy of patient selection for invasive staging. One potential explanation is the relatively small number of patients and events in this study. Importantly, however, we found no evidence of performance decrements in terms of discrimination, calibration (fit), or accuracy. The tendency of the model to overestimate risk was also observed in the internal validation study (14). On balance, the performance of the model appeared to be similar across two different centers among patient populations defined by the same inclusion/exclusion criteria.

External validation of risk-prediction models is important because the ascertainment and/or prevalence of risk-factors can vary across patient populations at different centers despite identical inclusion/exclusion criteria. Compared to the population in which the prediction model was originally developed and validated (14), patients in our study tended to have slightly larger tumors radiographically, higher  $SUV_{max}$ , and a higher frequency of lymphadenopathy by CT and FDG-uptake in N1 nodes. Interestingly, despite a higher prevalence of radiographic risk factors for pN2, the overall rate of pN2 did not vary significantly across studies. These findings support conventional wisdom suggesting variation in the interpretation of advanced imaging across centers. Although the reasons underlying such variation are often unknown, in some cases they are expected. For instance, factors that influence SUV measurement include the time from FDG injection to imaging, type of image reconstruction algorithm, reconstruction filter; scan length, and attenuation correction methods (17). The significance of variation in interpreting advanced imaging is that model performance may also vary from one site to another. While it is desirable to mitigate practice variation,

| Table 4 Exploratory analysis comparing use of the prediction model versus usual care |                  |  |                  |
|--|------------------|--|------------------|
|  | Usual care*      | Minimum recommended practice guideline<br>indications for invasive staging | Prediction model |
| Frequency of invasive staging (95% CI)   | 97% (94-99%)     | 77% (72-83%)   | 55% (48-61%)     |
| Accuracy of patient selection (95% Cl)   |                  |  |                  |
| Sensitivity  | 83% (59-96%)     | 100% (81-100%)   | 100% (81-100%)   |
| Specificity  | 2.3% (0.7-5.2%)  | 24% (19-31%)   | 49% (42-56%)     |
| Positive predictive value  | 6.5% (3.7-10.0%) | 10% (5.8-15.0%)  | 14% (8.3-21.0%)  |
| Negative predictive value  | 63% (24-91%)     | 100% (93-100%)   | 100% (97-100%)   |
| *, aggressive but guideline allowable invasive staging; CI, confidence interval.     |                  |  |                  |

it is not always practical to do so. For this reason, it is imperative to externally validate the performance of riskprediction models across varying settings. The similar performance of the prediction model across two different North American centers provides evidence in support of its generalizability. However, one important limitation of the current analysis is that model performance was evaluated at another high-volume, academic, National Cancer Institute (NCI)-designated comprehensive cancer center. Future studies should evaluate model performance across multiple environments including non-academic centers; non-NCI designated cancer centers, low-volume hospitals, and integrated health systems.

Our exploratory aim suggests that use of a risk-prediction model for mediastinal staging may lead to a decreased need for invasive diagnostic tests through better patient selection. To the extent that diagnostic accuracy, appropriate treatment selection, and patient outcomes are not adversely impacted by the use of a risk-prediction model, a reduction in the number of invasive diagnostic tests will increase the value of thoracic oncologic care. Certainly, fewer procedures will translate into less exposure to procedurerelated risks and lower costs. The efficiency of care may also improve by omitting one step (invasive staging) in a complex diagnostic algorithm for working up suspected or confirmed NSCLC (1,2). Another potential benefit of riskprediction is reducing unnecessary provider-level variation in the use of invasive procedures. Practice variation in the use of invasive staging modalities has not yet been described at the population level, but it is strongly suspected based on clinical experience. Importantly, the use of a riskprediction model does not eliminate patient-level variation in the use of invasive staging procedures arising from differences in individual-level risk. Risk-prediction provides an opportunity to standardize provider- or institutionallevel approaches to lung cancer staging while ensuring the delivery of personalized cancer care.

This study has several important limitations. Our approach to defining true pathologic nodal status is antiquated, even though it was based on a conventional criterion appropriate for the period of time under study. The traditional gold standard definition for mediastinal nodal disease is pathologic confirmation of disease in nodal tissue. However, this definition does not take into account the thoroughness of mediastinal nodal evaluation (18). A secondary analysis of the ACOSOG Z0030 trial revealed that 99% of patients had at least three nodal stations sampled (19). Subsequently, contemporary practice guidelines and quality improvement initiatives regard assessment of at least three nodal stations to be the standard (2,20). One potential consequence of lymph node sampling is that the true prevalence of nodal disease is underestimated, which may in turn falsely elevate sensitivity and falsely decrease specificity. A single institution study evaluated the prevalence of pN2 when usual care consists of routine lymphadenectomy and reported a prevalence of pN2 only slightly higher (10%) than our rate of pN2 (21). Furthermore, a post-hoc analysis restricted to patients in our study that had at least three nodal stations sampled (by any means or intraoperatively) revealed model performance similar to that observed in the internal validation study. We found no evidence to suggest that the pragmatic nature of our retrospective study has substantially biased results pertaining to model performance. Future validation studies should ideally define the gold standard for determining nodal disease by both pathologic confirmation of disease and adequate extent of mediastinal nodal assessment.

One potential criticism of our study is a missed opportunity to evaluate the performance of other previously published prediction models for pN2. All four prior models/studies did not use information from PET to define the population under investigation and/or include as model inputs. One North American study evaluated patients prior to the availability of PET (10). Two studies from China and one from Japan alluded to the selective use of PET in their countries because of high costs and/ or delayed adoption (11-13). Because PET is the standard of care in North American, we only considered validating the prediction model developed among a population that routinely underwent PET. As a consequence, the prediction model under study may only be generalizable to patients undergoing routine PET. Finally, another potential criticism of this study is that the availability of minimallyinvasive endosonography obviates the need for riskprediction. However, practice guidelines recommending the use of endosonography for invasive mediastinal staging emphasize the importance of available equipment and the skill and experience of the operator, which may not yet exist in all practice settings (1,2). Furthermore, the ASTER trial revealed that up to 53% of patients undergoing first-line endosonography will require surgical staging to definitely rule-out mediastinal disease (22). Finally, endosonography is not without risk (23). A risk-prediction model compliments new technology in reducing procedure-related risks for lung cancer patients.

In conclusion, a previously developed and internally

# Journal of Thoracic Disease, Vol 7, No 4 April 2015

validated prediction model for pN2 among patients with a negative mediastinum by PET is externally valid. The potential benefit of using risk-prediction is a reduction in the overall number of invasive diagnostic procedures achieved through better patient selection.

# Acknowledgements

Authors' contributions: F Farjah contributed substantially to the concept and study design, acquisition of data, data analysis and interpretation, and writing and revising the manuscript. JP Manning contributed substantially to the study design, acquisition of data, and revising the manuscript. LM Backhus, TK Varghese, AM Cheng, MS Mulligan, and DE Wood contributed substantially to the study design, interpretation of the data, and revision of the manuscript. All authors (F Farjah, LM Backhus, TK Varghese, JP Manning, AM Cheng, MS Mulligan, and DE Wood) have provided final approval of the version to be published and have agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

*Disclosure:* Dr. Backhus has received payment as a consultant for Disney and Engine Room Production Company. Dr. Mulligan has received payment as a consultant for Covidien. Dr. Wood has received payment as a consultant for Spiration as well as grant support for Spiration-related research. Dr. Farjah is a Cancer Research Network Scholar and the recipient of a Cancer Research Network Pilot Grant (Grant Number 1U24 CA171524 from the NCI).

# References

- Silvestri GA, Gonzalez AV, Jantz MA, et al. 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-50S.
- National Comprehensive Cancer Network. Clinical practice guidelines in oncology–v.2.2013: Non-small cell lung cancer. NCCN Guidelines 2013.
- Al-Sarraf N, Aziz R, Gately K, et al. 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;33:104-9.
- 4. Cerfolio RJ, Bryant AS. Survival of patients with

unsuspected N2 (stage IIIA) nonsmall-cell lung cancer. Ann Thorac Surg 2008;86:362-6; discussion 366-7.

- Defranchi SA, Cassivi SD, Nichols FC, et al. N2 disease in T1 non-small cell lung cancer. Ann Thorac Surg 2009;88:924-8.
- Kanzaki R, Higashiyama M, Fujiwara A, 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;71:333-7.
- Lee PC, Port JL, Korst RJ, et al. Risk factors for occult mediastinal metastases in clinical stage I non-small cell lung cancer. Ann Thorac Surg 2007;84:177-81.
- 8. Park HK, Jeon K, Koh WJ, et al. Occult nodal metastasis in patients with non-small cell lung cancer at clinical stage IA by PET/CT. Respirology 2010;15:1179-84.
- Trister AD, Pryma DA, Xanthopoulos E, et al. Prognostic value of primary tumor FDG uptake for occult mediastinal lymph node involvement in clinically N2/N3 nodenegative non-small cell lung cancer. Am J Clin Oncol 2014;37:135-9.
- 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;1:953-9.
- Koike T, Koike T, Yamato Y, et al. Predictive risk factors for mediastinal lymph node metastasis in clinical stage IA non-small-cell lung cancer patients. J Thorac Oncol 2012;7:1246-51.
- Zhang Y, Sun Y, Xiang J, et al. A prediction model for N2 disease in T1 non-small cell lung cancer. J Thorac Cardiovasc Surg 2012;144:1360-4.
- Chen K, Yang F, Jiang G, et al. Development and validation of a clinical prediction model for N2 lymph node metastasis in non-small cell lung cancer. Ann Thorac Surg 2013;96:1761-8.
- Farjah F, Lou F, Sima C, et al. A prediction model for pathologic N2 disease in lung cancer patients with a negative mediastinum by positron emission tomography. J Thorac Oncol 2013;8:1170-80.
- Hosmer DW, Lemeshow S. Applied logistic regression, second edition. New York: John Wiley and Sons; 2000.
- Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32-5.
- 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;270:91-8.

#### Farjah et al. Predicting N2 among lung cancer patients

- Detterbeck F, Puchalski J, Rubinowitz A, et al. Classification of the thoroughness of mediastinal staging of lung cancer. Chest 2010;137:436-42.
- Darling GE, Allen MS, Decker PA, et al. Randomized trial of mediastinal lymph node sampling versus complete lymphadenectomy during pulmonary resection in the patient with N0 or N1 (less than hilar) non-small cell carcinoma: results of the American College of Surgery Oncology Group Z0030 Trial. J Thorac Cardiovasc Surg 2011;141:662-70.
- Katlic MR, Facktor MA, Berry SA, et al. ProvenCare lung cancer: a multi-institutional improvement collaborative. CA Cancer J Clin 2011;61:382-96.

**Cite this article as:** 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(4):576-584. doi: 10.3978/ j.issn.2072-1439.2015.02.09

- 21. Cerfolio RJ, Bryant AS, Minnich DJ. Complete thoracic mediastinal lymphadenectomy leads to a higher rate of pathologically proven N2 disease in patients with nonsmall cell lung cancer. Ann Thorac Surg 2012;94:902-6.
- 22. Annema JT, van Meerbeeck JP, Rintoul RC, et al. Mediastinoscopy vs endosonography for mediastinal nodal staging of lung cancer: a randomized trial. JAMA 2010;304:2245-52.
- von Bartheld MB, van Breda A, Annema JT. Complication rate of endosonography (endobronchial and endoscopic ultrasound): a systematic review. Respiration 2014;87:343-51.

# 584

| Table S1 | Usual care (aggressive but guideline allowable invasive |
|----------|---|
| staging) |   |

|                     | pN2 | No pN2 |
|---------------------|-----|--------|
| Invasive staging*   | 15  | 216    |
| No invasive staging | 3   | 5      |

\*, reflects the actual number of patients who underwent invasive staging in this study. pN2, pathologic N2.

| Table S2 Minimum guideline recommended indications for |     |        |  |
|--|-----|--------|--|
| invasive staging                                       |     |        |  |
|  | pN2 | No pN2 |  |
| Invasive staging*                                      | 18  | 167    |  |
| No invasive staging                                    | 0   | 54     |  |

\*, reflects the number of patients with a radiographic stage  $\geq 1B$  cancer and/or a central tumor. pN2, pathologic N2.

| Table S3 Prediction model                                  |     |        |  |
|--|-----|--------|--|
|  | pN2 | No pN2 |  |
| Invasive staging*  | 18  | 108    |  |
| No invasive staging  | 0   | 113    |  |
| *, reflects the number of a patients considered high-risk  |     |        |  |
| (probability ≥8.3%) for nodal disease. pN2, pathologic N2. |     |        |  |