A Prediction Model to Help with Oncologic Mediastinal Evaluation for Radiation: HOMER

Gabriela Martinez-Zayas
Francisco A. Almeida
Michael Simoff
Henry Ford Health System, msimoff1@hfhs.org
Lonny Yarmus
Sofia Molina

See next page for additional authors

Follow this and additional works at: https://scholarlycommons.henryford.com/pulmonary_articles

Recommended Citation
Martinez-Zayas G, Almeida FA, Simoff MJ, Yarmus L, Molina S, Young B, Feller-Kopman D, Sagar AS, Gildea T, Debiane LG, Grosu HB, Casal RF, Arain MH, Eapen GA, Jimenez CA, Noor LZ, Baghaie S, Song J, Li L, and Ost DE. A Prediction Model to Help with Oncologic Mediastinal Evaluation for Radiation: HOMER. Am J Respir Crit Care Med 2019.

This Article is brought to you for free and open access by the Pulmonary and Critical Care Medicine at Henry Ford Health System Scholarly Commons. It has been accepted for inclusion in Pulmonary and Critical Care Medicine Articles by an authorized administrator of Henry Ford Health System Scholarly Commons.
A Prediction Model to Help with Oncologic Mediastinal Evaluation for Radiation:

HOMER

Gabriela Martinez-Zayas1,2; Francisco A. Almeida, MD, MS3; Michael J. Simoff, MD4; Lonny Yarmus, DO5; Sofia Molina, MD1,2; Benjamin Young, MD6; David Feller-Kopman, MD5; Ala-Eddin S. Sagar, MD2; Thomas Gildea, MD, MS3; Labib G. Debiane, MD5; Horiana B. Grosu, MD, MPH2; Roberto F. Casal, MD2; Muhammad H. Arain, MBBS2; George A. Eapen, MD3; Carlos A. Jimenez, MD2; Laila Z. Noor, MD2; Shiva Baghaie, MPH2; Juhee Song, PhD7; Liang Li, PhD7; David E. Ost, MD, MPH2

1Tecnologico de Monterrey, School of Medicine and Health Sciences, Monterrey, NL, Mexico; 2Department of Pulmonary Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; 3Department of Pulmonary Medicine, Cleveland Clinic, Cleveland, OH, USA; 4Department of Pulmonary and Critical Care Medicine, Henry Ford Hospital, Detroit, MI, USA; 5Johns Hopkins University, Baltimore, MD, USA; 6Division of Pulmonary and Critical Care Medicine, University Hospitals, Cleveland, OH, USA; 7Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

Corresponding Author Information:

David E. Ost, MD, MPH, The University of Texas MD Anderson Cancer Center, Pulmonary Department, 1515 Holcombe Blvd, Unit 1462, Houston, Tx, 77030; Email: dost@mdanderson.org; phone: 713-745-8775; fax: 713-749-4922.

Detail of each author’s contributions to the study:
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. G.M.Z. was involved in data collection, statistical analysis and manuscript writing. SMG was involved in statistical analysis and contributed to writing. L.L. and J.S were the primary biostatisticians for the project, constructed models and analysis and contributed to writing. F.A.A. (PI, Cleveland Clinic), M.J.S. (PI, Henry Ford Hospital), L.Y. (PI, Johns Hopkins), B.Y., D.F.-K., A.E.S., T.G., L.G.D., H.B.G., R.F.C., G.A.E., C.A.J., contributed to data collection and writing. L.Z.N., M.H.A., and S.B. were involved in data extraction.

Funding: Statistical analysis work supported in part by the Cancer Center Support Grant (NCI Grant P30 CA016672).
Descriptor number: 9.24 Interventional Pulmonology/Bronchoscopy
Total word count for the body manuscript: 4243

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

Rationale: When stereotactic ablative radiotherapy (SABR) is an option for non-small cell lung cancer (NSCLC) patients, distinguishing between N0, N1 and N2 or N3 (N2|3) disease is important.

Objectives: To develop a prediction model for estimating the probability of N0, N1, and N2|3 disease.

Methods: Consecutive patients with clinical-radiographic stage T1-3/N0-3/M0 NSCLC that underwent endobronchial ultrasound-guided staging from a single center were included. Multivariate ordinal logistic regression analysis was used to predict the presence of N0, N1 or N2|3 disease. Temporal validation used consecutive patients from three years later at the same center. External validation used three other hospitals.

Results: In the model development cohort (n=633), younger age, central location, adenocarcinoma and higher PET-CT nodal stage were associated with a higher probability of having advanced nodal disease. Area under the receiver operating characteristic curves (AUC) were 0.84 and 0.86 for predicting N1 or higher (vs. N0) disease and N2|3 (vs. N0|1) disease respectively. Model fit was acceptable (Hosmer-Lemeshow p=0.960; Brier score 0.36). In the temporal validation cohort (n=473) AUCs were 0.86 and 0.88. Model fit was acceptable (Hosmer-Lemeshow p=0.172; Brier score 0.30). In the external validation cohort (n=722), AUCs were 0.86 and 0.88, but required calibration (Hosmer-Lemeshow p<0.001; Brier score 0.38). Calibration using the general calibration method resulted in acceptable model fit (Hosmer-Lemeshow p=0.094; Brier score, 0.34).
Conclusions: This prediction model can estimate the probability of N0, N1 and N2/3 disease in NSCLC patients. The model has the potential to facilitate decision-making in NSCLC patients when SABR is an option.

Key words: lung cancer; lung cancer staging; endobronchial ultrasound, mediastinal adenopathy

Total abstract word count: 250
Introduction

In patients with non-small cell lung cancer (NSCLC), correct staging is necessary to offer appropriate treatment. Treatment options and prognosis vary significantly by stage. After ruling out metastatic disease, knowing the N stage is necessary to determine the best treatment strategy. Patients with stage I or a subset of stage II (T1-2, N1) NSCLC are generally candidates for surgical resection and mediastinal lymph node dissection. For patients with early-stage NSCLC who are medically inoperable, or in patients who refuse surgery, stereotactic ablative radiotherapy (SABR) is recommended. For higher cancer stages, multimodality therapy with chemoradiation, chemotherapy or targeted therapy is preferred.

Previously, O’Connell et al. published a prediction model to help with the assessment of Adenopathy in Lung cancer (HAL model). This model predicts the probability of having N2 or N3 (prN2|3) disease as determined by endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) in NSCLC patients. In this model, younger age, central tumor location, adenocarcinoma, and higher N stage by positron emission tomography-computed tomography (PET-CT) were all associated with increased prN2|3 disease vs. N0 or N1 (N0|1) disease.

However, for patients being considered for SABR, it is important to distinguish between N0 and N1 disease, since ablative radiation is directed only to the primary tumor without covering N1 nodes. If N1 disease is present, SABR may not suffice. Predicting the probability of N0, N1, N2|3 nodal disease in NSCLC patients requires a different model, since HAL cannot distinguish between N0 and N1 disease. Accurate estimates of the probability of N0, N1 and N2|3 disease are central to the decision-making process and help to drive staging and treatment decision in NSCLC patients in which SABR is being considered.
In this study, our objective was to create a prediction model for estimating the probability of N0, N1, and N2|3 lymph node involvement in NSCLC patients. The secondary objective was to temporally and externally validate the model.

Some of the results of this study have been presented in abstract form during the 2019 American Thoracic Society Conference. (9)

Methods

The development, temporal validation and external validation cohorts shared the same inclusion and exclusion criteria. All consecutive untreated NSCLC patients, clinical radiographic stage T1-3, N0-3, M0 who underwent EBUS-TBNA for staging were included. Patients with distant metastasis, mediastinal invasion by CT, suspected or confirmed synchronous primaries, recurrent lung cancer, and small cell were excluded. Patients without PET imaging prior to treatment were excluded.

For the development cohort, we performed a retrospective analysis of consecutive NSCLC patients that underwent staging EBUS-TBNA from September 2009 to January 2013. The study was approved by the Institutional Review Board Committee 5, Protocol PA16-0107, at the University of Texas MD Anderson Cancer Center (MDACC). Data was collected prospectively as part of the American College of Chest Physicians (ACCP) Quality Improvement Registry, Evaluation and Education (AQuIRE) registry as previously reported.(4, 10-13) We used standardized definitions, quality control checks, and entered data into a Web-based interface (REDCap™). This data set was the same one used to develop the HAL model.(4)

PET-CT scan was used to define location of the tumor (central vs. peripheral) and N stage. Definitions of all variables were developed prior to data abstraction and provided to all sites. For
computed tomography (CT) scans, abnormal lymph nodes were defined as being greater or equal to 1 cm in their short axis. Lymph node N stage was determined by review of the radiology report and further review by an interventional pulmonologist or an interventional pulmonary fellow under supervision in order to assign an N stage to the patient. If both contrast and non-contrast CT were available, the contrast-enhanced images were used to determine CT N stage. Positive emission tomography (PET) N stage was based on the radiologist’s interpretation of mediastinal lymph node fluorodeoxyglucose (FDG) avidity. In some cases standardized uptake values (SUV) measurements were recorded. In those cases that SUV measures on lymph nodes were available, a SUV value equal or greater than 2.5 was considered positive. Based on the radiologist’s reading and further review by an interventional pulmonologist or a supervised interventional pulmonary fellow, the PET N stage of the lesion was determined. Radiographic N stage by PET-CT was defined as the highest abnormal nodal station using The Eights Edition Lung Cancer Stage Classification. (14)

Tumors in the inner one-third of the hemithorax were defined as central (see online supplement, Figure E1).(15, 16) All EBUS-TBNA procedures sampled N3 following by N2 and then N1 nodes. All lymph nodes measuring 0.5 cm or larger by EBUS were sampled, independent of PET-CT status.

**Statistical Analysis**

*Prediction model development.* The primary outcome was the highest N stage (N0 vs. N1 vs. N2|3) lymph node with malignancy as determined by EBUS-TBNA. N0, N1 and N2|3 disease groups were compared using Fisher’s exact test for categorical variables and ANOVA for continuous normally distributed variables. Because PET-CT images do not use contrast, and
some patients only had non-contrast CT available, we used N0 or N1 disease as a single variable for CT but they were kept separate for PET as previously reported (see online supplement).(4)

We used univariate ordinal logistic regression to identify variables associated with the outcome variable, highest N stage by EBUS, classified in the following order: N0 < N1 < N2|3. With three ordinal outcomes, two sets of probabilities are calculated. The first is the probability of N stage being equal or greater than one (prN1|2|3) vs. the probability of N0 (prN0) disease. The second is the probability of N stage being N2|3 vs. N0|1. We specified a priori that variables with an overall p-value <0.2 on univariate analysis would be candidate variables for the multivariable ordinal logistic regression model. We checked the proportional odds assumption using the Score test. Different slope parameters were allowed for variables that violated this assumption (see online supplement). We specified a priori that we would use stepwise backward selection with an overall p-value <0.05 for variables to remain in the model.

**Temporal and External Validation**

For temporal validation, data from a completely different cohort of MDACC patients that underwent EBUS-TBNA from September 2016 to January 2019 was used. This data was prospectively collected and constituted the temporal validation cohort. (17, 18)

For external validation, data from three centers (Johns Hopkins, Henry Ford Hospital and Cleveland Clinic Foundation) was used. (17, 18) Consecutive patients were entered using identical definitions, forms and quality control checks as in the development cohort. These patients constitute the external validation cohort and correspond to the external validation cohort in the HAL model.(4)
Model Performance Assessment

We assessed model performance in the development, temporal validation, and external validation cohorts. We used the area under the receiver operating characteristic (ROC) curve (AUC) to assess discrimination. We used the Hosmer-Lemeshow goodness-of-fit test, Brier score, and observed vs. predicted graphs to assess calibration (see online supplement).

We created a calibrated model for the combined data from all three outside institutions and a separate calibrated model for each institution using the general calibration method presented by Steyerberg et al. (see online supplement) as previously reported in the HAL model (see online supplement for additional details).(4, 19)

For the temporal validation cohort, we hypothesized that the model would not require further calibration since the location was the same as the development cohort and we wanted to test model stability over time. Therefore, we pre-specified that we would not calibrate the temporal validation cohort whatsoever, and measured discrimination and calibration using the baseline model.

All statistical analyses use SAS® version 9.4 or STATA® 15.1.

Results

The development cohort consisted of 633 patients. Descriptive statistics for the cohort stratified by final EBUS N stage are in Table 1.

Model Development

Univariate ordinal logistic regression results are in Table E1 and multivariate results are in Table 2. The only candidate variable that violated the proportional odds assumption was N stage by
PET-CT (p<0.001). For this variable, different slope parameters were allowed. Younger age, adenocarcinoma histology, central location, and higher nodal stage by PET-CT were associated with an increased prN1|2|3 (vs. prN0) disease and higher prN2|3 (vs. prN0|1) disease (Table 2). ROC AUC was 0.84 (95% CI=0.81-0.87) for predicting N1|2|3 (vs. N0) disease and 0.85 (95% CI= 0.82-0.89) for predicting N2|3 (vs. N0|1) disease (Figures 1A-B). Model fit was acceptable (Hosmer-Lemeshow test, p=0.960; Brier score, 0.36; observed vs. predicted plots, Figures 2A-B).

**Temporal Validation**

The temporal validation cohort included 473 patients (Table E2). ROC AUC was 0.86 (95%CI= 0.85-0.90) for predicting N1|2|3 (vs. N0) disease and 0.88 (95%CI=0.84-0.92) for predicting N2|3 (vs. N0|1) disease (Figures 1C-D).

Model fit was acceptable (Hosmer-Lemeshow test, p=0.172; Brier score, 0.30; observed vs. predicted graphs, Figures 2C-D). There was no need to calibrate the model, suggesting that the model is stable and accurate over time in this location.

**External Validation**

The external validation cohort included 722 patients (Table E3). Discrimination was good for the combined external validation cohort and for each outside institution when assessed separately. For the combined external validation cohort, AUC was 0.86 (95% CI= 0.84-0.89) for predicting N1|2|3 (vs. N0) disease and 0.88 (95% CI= 0.85-0.90) for predicting N2|3 (vs. N0|1) disease (Figure E2). The AUCs for each outside institution ranged from 0.81 to 0.91 for predicting N1|2|3 (vs. N0) disease and from 0.82 to 0.92 for predicting N2|3 (vs. N0|1) disease (Table 3).
When assessing calibration of the combined external validation cohort, model fit was not acceptable (Hosmer-Lemeshow, p-value <0.001; Brier score, 0.38; observed vs. predicted graphs, Figures 3A-B). When we assessed model calibration in each institution separately, model fit was acceptable in one of the outside institutions (Hosmer-Lemeshow, p-value=0.286; Brier score, 0.34; observed vs. predicted graphs, Figures 4A-B; Table 3) but was off in two of the external validation sites (Hosmer-Lemeshow, p-values <0.001; Brier score range 0.36-0.46; observed vs. predicted graphs, Figures 4C-F; Table 3).

**Calibration of the External Validation Cohorts**

When we calibrated the model to predict outcomes for the combined external validation cohort, two sets of slope and intercept were calculated, one for predicting N1|2|3 (vs. N0) disease and another for N2|3 (vs. N0|1) disease. Both calibration intercepts were the same (0.75), but the slopes were off by 0.01 (1.13 vs. 1.14). Both calibrations were evaluated (intercept=0.75; slopes=1.13 and 1.14), and the one with lower Brier score was selected (intercept=0.75; slope=1.14). After calibration, the Hosmer-Lemeshow test was non-significant (p=0.094), and both the Brier score (0.34) and observed vs. predicted graphs (Figures 3C-D) showed improved model fit.

For model calibration of each outside institution, the slope for all three centers was set to unity (b=1). The pair of intercepts with minimum Brier score and maximum Hosmer-Lemeshow p-value was selected (Table E4). The institution specific calibrated models performed well. Hosmer-Lemeshow tests became non-significant (p-value range from 0.196 to 0.404, Table E4), Brier scores improved (range from 0.29-0.39, Table E4) and observed vs. predicted graphs showed improved model fit (Figure 5).
Discussion and Conclusions

In this study, we report the Help with Oncologic Mediastinal Evaluation for Radiation (HOMER) model, which estimates the probability of N0 vs. N1 vs. N2|3 metastatic nodal disease in NSCLC patients. We demonstrated that HOMER is accurate in outside institutions after calibration using the general calibration method. We also demonstrated that the model maintained good discrimination and calibration over an extended period of time when applied to a single institution using two different data sets.

The ACCP and National Comprehensive Cancer Network lung cancer guidelines suggest using prediction models to estimate the probability of malignancy in solitary pulmonary nodules to help inform decision-making in patients with solitary pulmonary nodules.\(^{(2, 20, 21)}\) Similarly, investigators have developed binary prediction models to estimate prN2|3 (vs. prN0|1) to help inform decision-making in patients with NSLC with regard to staging procedures when surgical treatment is the main option.\(^{(4, 22-26)}\) However, those studies did not distinguish between N0 and N1 disease, which is critical when SABR is a treatment option. One study did use separate binary logistic regression models to identify risk factors for N1 vs. N0 disease and for N2 vs. N0 disease in patients that had surgery.\(^{(27)}\) However, this study did not include non-surgical patients, limiting the generalizability of the findings. More importantly, because the investigators used multiple binary logistic regression models rather than a single ordinal logistic regression model, their model is not valid for clinical prediction. That is because the form of their models are: a) given that a patient has N0 or N1 disease, then the odds are x; b) Given that a patient has N0 or N2 disease, then the odds are y. In real life, physicians cannot know a priori that N1, N2 or N3 disease is definitely absent. Therefore, multiple binary models of this form cannot work for clinical prediction.
This study adds to the existing body of knowledge by using ordinal logistic regression to develop a more generalizable prediction rule to inform decision-making in patients who are candidates for SABR. To our knowledge, HOMER is the first externally and temporally validated model to predict prN0, prN1 and prN2/3 disease developed from a broad population of patients that included both surgical and non-surgical candidates. It incorporates PET imaging which is part of the current standard of care but does not rely on molecular markers. Because it includes both surgical and non-surgical candidates, the model is applicable to patients in whom SABR is the only option and to patients in whom SABR is being considered as an alternative to surgery (e.g. borderline surgical candidates with T1b, N0, M0 disease by PET-CT).(2, 8)

Although there is currently an ongoing prospective study of endosonographic intrathoracic nodal staging of patients being considered for therapy with SABR, there are currently no definitive recommendations on whether to perform EBUS for mediastinal staging prior to SABR.(28, 29) The role of EBUS prior to SABR depends in large part on the context. EBUS could be useful to inform decisions when patients are candidates for both surgery and SABR (e.g. PET-CT N0).(30) A finding of N1 disease would lead to a clear recommendation in such cases. Whether EBUS should be done in this context is a function of the probability of EBUS being positive and the complication rate of EBUS.(31) Given the low rate of complications from EBUS, even a relatively low prN1 disease in these patients might warrant EBUS. Conversely, in patients that are not surgical candidates the need for EBUS is different. Finding N1 disease in such patients would probably lead to definitive radiation therapy while finding N2 disease would lead to multimodal treatment with radiation and chemotherapy instead of treatment with SABR. Whether EBUS is warranted in this context would be a function of the probability of EBUS being positive, the complication rate of EBUS, and the marginal benefit and marginal harm of
treating with chemotherapy and radiation for N2 disease and definitive radiation for N1 disease (as compared to treating with SABR based solely on imaging). (32-50) However, in both cases accurately predicting the probability of EBUS being positive is vital. HOMER can aide in this decision-making process.

Consideration of specific cases may help illustrate these concepts. Consider a 60 year-old patient with adenocarcinoma in the outer two-thirds of the lung, PET-CT N0, who is a surgical candidate but prefers not to have surgery if possible and SABR is being considered. HOMER predicts prN0 is 93% by EBUS (Table E5). When the physician weighs the risk of complications at approximately 1.15% vs. a 7% chance of having prN1|2|3 and changing treatment from SABR to surgery, EBUS seems warranted.(31) However, in other circumstances HOMER might lead to a decision not to do EBUS. Consider an 80 year-old patient with squamous cell carcinoma in the outer two-thirds of the lung, PET-CT N0, who is not a candidate for surgery. HOMER predicts prN0 is 98% by EBUS (Table E5). Given the risks of EBUS, the marginal benefit and marginal harm of treating with chemotherapy and radiation instead of SABR for occult N2 disease, and only a 2% chance of EBUS being positive, proceeding directly to SABR is reasonable. (32-50) The absolute difference in prN0 in these two patients is only 5%, which at first glance seems small. However, given the low risk of EBUS and consideration of the benefits and harms of treatment, the value of information in this context is high.(7, 8, 51)

Another practical application of HOMER is helping to inform decisions in patients with clinical-radiographic N1 disease. Previous studies have pointed out that EBUS in this population may downstage them making them potentially suitable for SABR therapy.(6) If EBUS is performed and the patient is N0 by EBUS, the decision on whether SABR is a reasonable choice depends on knowledge of the post-test probability of nodal disease as well as the benefits and harms of
SABR vs. other treatment alternatives. The probability of nodal disease after a negative EBUS is of course a function of the sensitivity of EBUS. The sensitivity of EBUS varies with PET-CT stage. (6, 52-65) If EBUS sensitivity is 0.8 in the setting of N1 disease by PET-CT, then we can use HOMER to estimate the post-test probability that a patient with a negative EBUS actually has N0, N1 or N2|3 disease. In a 60 year-old patient with adenocarcinoma in the outer two-thirds of the lung, PET-CT N1 disease, assuming EBUS sensitivity is 0.8, the probability of true N0 disease given a negative EBUS is only 61% (see online supplement for calculations). If EBUS sensitivity is 0.9 the probability of true N0 disease is 83%. Conversely, if the patient is 80 years old with a peripheral squamous cell in the outer two-thirds of the lung, PET-CT N1, with a negative EBUS, then the corresponding probabilities for true N0 disease are 91% and 96%. In scenarios where EBUS is negative but HOMER predicts a 17%-39% (or a 4%-9% in the second scenario) probability of having nodal metastasis, other factors related to the benefit and harm of SABR alone vs. alternative strategies become relevant. These other factors include the performance status of the patient, tumor marker status, and the multimodality options being considered. If definitive radiotherapy covering the hilar nodes is used, toxicity will be higher but there will be the possibility of cure. Conversely, if SABR is used and targeted therapy is used to treat subsequent relapses, toxicity will be lower but the treatment will be palliative in nature. By providing the predicted probability, HOMER can help inform this decision, adding nuance to this complex decision process.

By using an ordinal model, HOMER also provides additional insights into the relationship of the predictor variables to N stage that are lacking in binary models and that might fail to capture the entire complexity of a clinical decision problem due to simplification and loss of information. HOMER demonstrates that for older patients with N2|3 disease by PET, the most likely stage by
EBUS is not necessarily N2|3 (and not N1), but rather it can be either N2|3 or N0 disease depending on tumor location (Figure E3C). For younger patients with N2|3 disease by PET, the most likely stage by EBUS is N2|3, with N0 disease being a close second. Quantifying these probabilities using HOMER (Figure E3C) shows us that N1 disease by EBUS is actually a rare finding in PET-CT N2|3 patients, with prN1 being much lower than either prN0 or prN2|3. No binary model can capture these subtleties (see online supplement for further discussion).

As in the HAL model, our study failed to demonstrate a relationship between tumor size and higher N stage by EBUS-TBNA after adjusting for age, tumor location, tumor histology, and PET-CT N stage.(4) This contrasts with other studies that have reported an association between larger tumors and probability of nodal metastatic disease.(22-25) However, none of the prior studies adjusted for PET-CT N stage in their analysis. PET-CT N stage in this dataset is associated with tumor size (p=0.036, see online supplement). Therefore, PET-CT N stage potentially confounds the relationship between tumor size and nodal metastatic disease. This could explain the discordance in findings between studies. To make sure tumor size did not improve model performance, we forced tumor size back into the model, which did not improve discrimination but worsened calibration (see online supplement).

Effective prediction models should be validated on external cohorts and should demonstrate both good discrimination and calibration (see online supplement).(17, 18) HOMER demonstrates good discrimination in both the combined external validation cohort and for each institution when assessed separately. However, observed vs. predicted graphs show that, while prediction of prN1|2|3 (vs. prN0) disease is decent, HOMER overestimates the prN2|3 (vs. prN0|1) disease for two out of the three outside institutions (Figures 3A-B, 4). Application of the general calibration method corrects this (Figures 3C-D, 5).
In regards to the observed vs. predicted graphs for the temporal validation cohort, HOMER slightly overestimates prN1|2|3 (vs. prN0) disease when predicted probability is over 50% (Figure 2C-D). Nonetheless, the Brier score and Hosmer-Lemeshow p-value both show that the model has acceptable fit and does not require further calibration, suggesting that the model has temporal stability. Model stability over time is an important consideration. If the model is temporally stable, then calibration intercepts can be determined for any outside institution and from that point forward the model can make accurate predictions for patients at that institution.

To our knowledge, this is the first study of EBUS-TBNA to externally and temporally validate a prediction model for N0, N1 and N2|3 disease. The HAL model was the first study of EBUS-TBNA to externally validate a prediction model for N2|3 disease.(4) In this study, we externally validate HOMER, but we also temporally validate it, showing that it is stable through time in the institution of model development.

A limitation to the prediction power of our model is the small number of patients that fall in the least common of the three possible outcomes (N1 disease). The proportion of patients identified as having N1 disease by EBUS-TBNA in our data ranges from 7.2% to 10.11% for the development, temporal validation, and external validation cohorts. Our data is consistent with the findings of other investigators, where the proportion of patients with N1 disease ranges from 5.3% to 16%.(53, 54, 66, 67) The least common outcome determines the number of covariates a model can support, so having relatively little N1 disease limits the number of covariates that can be included for model development. Larger cohorts would be required if more covariates were to be introduced. It is possible that a larger sample size would allow identification of additional covariates, which might significantly improve model performance. Furthermore, the validity of the predictions made by HOMER are limited to patients that are between 60 to 80 years of age,
the range of data in which 68% percent of patients from the development cohort are found (see online supplement for details). The model might not perform as well at the extremes of age.

Another limitation is that our predictions are for the observed cytology as identified by EBUS-TBNA, which itself has a varying sensitivity that ranges from 54.5% to 98%, with a recent meta-analysis suggesting a pooled sensitivity of 90%, therefore the model might underestimate the presence of true lymph node metastases.\(^1\), \(^53\), \(^67\)-\(^71\) Since HOMER does not predict the probabilities of N0, N1 or N2|3 disease as determined by thoracotomy, adjustment for EBUS sensitivity may be required depending on how the model is used. A study of surgical patients undergoing thoracotomy would facilitate prediction of the true pretest probability of nodal disease and would help determine the sensitivity and specificity of EBUS-TBNA, but such a design would also have problems with generalizability and possibly selection bias, since non-surgical patients going for SABR and those with significant comorbidities could not be included. In addition predicting pretest probability would not be sufficient by itself to determine whether EBUS-TBNA should be done in a given patient. What we need to know is the chance that EBUS-TBNA will be positive in that same patient (i.e. diagnostic yield). Homer predicts the diagnostic yield of EBUS-TBNA for N0, N1 or N2|3 disease, which is fundamentally different than predicting the pretest probability of disease as determined by the gold-standard (see online supplement). An additional limitation is that our prediction rule is only valid in centers that perform EBUS-TBNA in a similar systematic manner.\(^4\) All three centers of the external validation cohort are high-volume centers. Higher procedural volume is associated with higher diagnostic yield, so results may differ for centers with lower procedural volume.\(^13\) A final limitation is that the Brier score used to assess calibration does not take into account the ordinal nature of the outcomes (see online supplement).
In summary, the HOMER model predicts the probability of finding N0 vs. N1 vs. N2|3 disease on EBUS, and had good performance as assessed by tests of discrimination and calibration in the development cohort. The predictor variables identified were consistent with the previously reported HAL model.\(^{(4)}\) In regards to external validation, the model has good discrimination but requires calibration. After calibration, the model demonstrates sufficient precision to be useful clinically. Performance in the temporal validation cohort suggests that the model is stable through time in the institution where it was developed. The HOMER model is potentially useful for predicting N stage and informing decisions regarding staging and treatment for patients with NSCLC in which SABR is an option. Future studies will need to assess whether calibrated models in outside institutions are temporally stable. If that is indeed the case, then HOMER could potentially be integrated into electronic health records as a decision support tool.
Acknowledgements

The authors thank Joseph Cicenia, MD; Sonali Sethi, MD and Michael Machuzak, MD for their contribution to data collection.
References

1. 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-e250S.

2. Network NCC. Non-small cell lung cancer (Version 2.2019). 2018. Available from: https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf.

3. Reck M, Rabe KF. Precision diagnosis and treatment for advanced non–small-cell lung cancer. *New England Journal of Medicine* 2017; 377: 849-861.

4. O'Connell OJ, Almeida FA, Simoff MJ, Yarmus L, Lazarus R, Young B, Chen Y, Semaan R, Saettele TM, Cicionia J, Bedi H, Kliment C, Li L, Sethi S, Diaz-Mendoza J, Feller-Kopman D, Song J, Gildea T, Lee H, Grosu HB, Machuzak M, Rodriguez-Vial M, Eapen GA, Jimenez CA, Casal RF, Ost DE. A prediction model to help with the assessment of adenopathy in lung cancer: HAL. *American journal of respiratory and critical care medicine* 2017; 195: 1651-1660.

5. Akthar AS, Ferguson MK, Koshy M, Vigneswaran WT, Malik R. Limitations of PET/CT in the detection of occult N1 metastasis in clinical stage I(T1-2aN0) non-small cell lung cancer for staging prior to stereotactic body radiotherapy. *Technology in Cancer Research & Treatment* 2017; 16: 15-21.

6. Vial MR, Khan KA, O'Connell O, Peng SA, Gomez DR, Chang JY, Rice DC, Mehran R, Jimenez CJ, Grosu HB, Ost DE, Eapen GA. Endobronchial ultrasound-guided transbronchial needle aspiration in the nodal staging of stereotactic ablative body radiotherapy patients. *The Annals of thoracic surgery* 2017; 103: 1600-1605.
7. Hunink MGM, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, Glasziou PP. Choosing the best treatment. Decision Making in Health and Medicine: Integrating Evidence and Values, 2 ed. Cambridge: Cambridge University Press; 2014. p. 53-77.

8. Louie AV, Senan S, Patel P, Ferket BS, Lagerwaard FJ, Rodrigues GB, Salama JK, Kelsey C, Palma DA, Hunink MG. When is a biopsy-proven diagnosis necessary before stereotactic ablative radiotherapy for lung cancer?: A decision analysis. *Chest* 2014; 146: 1021-1028.

9. Martinez-Zayas G, Almeida FA, Simoff MJ, Yarmus L, Molina S, Young B, Feller-Kopman D, Sagar AS, Gildea T, Debiane LG, Grosu HB, Casal RF, Arain MH, Eapen GA, Jimenez CA, Noor LZ, Baghaie S, Song J, Li L, Ost DE. Prediction model for malignant N1, N2 or N3 relative to N0 in patients with non-small cell lung cancer. Dallas, Tx.: American Thoracic Society 2019 International Conference; 2019.

10. Ost DE, Ernst A, Lei X, Kovitz KL, Benzaquen S, Diaz-Mendoza J, Greenhill S, Toth J, Feller-Kopman D, Puchalski J, Baram D, Karunakara R, Jimenez CA, Filner JJ, Morice RC, Eapen GA, Michaud GC, Estrada YMRM, Rafeq S, Grosu HB, Ray C, Gilbert CR, Yarmus LB, Simoff M. Diagnostic yield and complications of bronchoscopy for peripheral lung lesions: results of the AQuIRE Registry. *American journal of respiratory and critical care medicine* 2016; 193: 68-77.

11. Ost D, Simoff M, Feller-Kopman D, Kovitz K, Eapen G, Greenhill S, Michaud G, Majid A, HeyM J, Jimenez C, Herth F, Ernst A. Risk adjusted diagnostic yield and outcomes for endobronchial ultrasound guided TBNA: results of the ACCP AQuIRE Registry. *Chest* 2009; 136: 69S.

12. Eapen GA, Shah AM, Lei X, Jimenez CA, Morice RC, Yarmus L, Filner J, Ray C, Michaud G, Greenhill SR, Sarkiss M, Casal R, Rice D, Ost DE. Complications, consequences, and
practice patterns of endobronchial ultrasound-guided transbronchial needle aspiration: results of the AQuIRE Registry. Chest 2013; 143: 1044-1053.

13. Ost DE, Ernst A, Lei X, Feller-Kopman D, Eapen GA, Kovitz KL, Herth FJF, Simoff M, Registry AQB. Diagnostic yield of endobronchial ultrasound-guided transbronchial needle aspiration: results of the AQuIRE Bronchoscopy Registry. Chest 2011; 140: 1557-1566.

14. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The eighth edition lung cancer stage classification. Chest 2017; 151: 193-203.

15. Casal RF, Vial MR, Miller R, Mudambi L, Grosu HB, Eapen GA, Jimenez CA, Morice RC, Cornwell L, Ost D. What Exactly Is a Centrally Located Lung Tumor? Results of an Online Survey. Annals of the American Thoracic Society 2017; 14: 118-123.

16. Casal RF, Sepesi B, Sagar A-ES, Tschirren J, Chen M, Li L, Sunny J, Williams J, Grosu HB, Eapen GA, Jimenez CA, Ost DE. Centrally-located lung cancer and risk of occult nodal disease: an objective evaluation of multiple definitions of tumor centrality with a dedicated imaging software. European Respiratory Journal 2019: 1802220.

17. Austin PC, van Klaveren D, Vergouwe Y, Nieboer D, Lee DS, Steyerberg EW. Validation of prediction models: examining temporal and geographic stability of baseline risk and estimated covariate effects. Diagn Progn Res 2017; 1: 12-12.

18. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. European heart journal 2014; 35: 1925-1931.

19. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York, NY: Springer; 2009.
20. Rivera MP, Mehta AC, Wahidi MM. Establishing the diagnosis of lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; 143: e142S-e165S.

21. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun 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-e120S.

22. Shafazand S, Gould MK. A clinical prediction rule to estimate the probability of mediastinal metastasis in patients with non-small cell lung cancer. *Journal of Thoracic Oncology* 2006; 1: 953-959.

23. 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. *The Journal of Thoracic and Cardiovascular Surgery* 2012; 144: 1360-1364.

24. 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. *The Annals of thoracic surgery* 2013; 96: 1761-1768.

25. 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. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2012; 7: 1246-1251.

26. 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. *Journal of Thoracic Oncology* 2013; 8: 1170-1180.
27. Cho S, Song IH, Yang HC, Kim K, Jheon S. Predictive factors for node metastasis in patients with clinical stage I non-small cell lung cancer. *Ann Thorac Surg* 2013; 96: 239-245.

28. 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.

29. Annema JT. Complete endosonographic intrathoracic nodal staging for lung cancer patients in whom SABR is considered. 2016. Available from: [https://clinicaltrials.gov/ct2/show/NCT02997449](https://clinicaltrials.gov/ct2/show/NCT02997449).

30. Wink KCJ, van Baardwijk A, Troost EGC, De Ruysscher D. Nodal recurrence after stereotactic body radiotherapy for early stage non-small cell lung cancer: Incidence and proposed risk factors. *Cancer treatment reviews* 2017; 56: 8-15.

31. Eapen GA, Shah AM, Lei X, Jimenez CA, Morice RC, Yarmus L, Filner J, Ray C, Michaud G, Greenhill SR, Sarkiss M, Casal R, Rice D, Ost DE, American College of Chest Physicians Quality Improvement Registry E. Complications, consequences, and practice patterns of endobronchial ultrasound-guided transbronchial needle aspiration: Results of the AQuIRE registry. *Chest* 2013; 143: 1044-1053.

32. Ball D, Mai GT, Vinod S, Babington S, Ruben J, Kron T, Chesson B, Herschtal A, Vanevski M, Rezo A, Elder C, Skala M, Wirth A, Wheeler G, Lim A, Shaw M, Schofield P, Irving L, Solomon B. Stereotactic ablative radiotherapy versus standard radiotherapy in stage 1 non-small-cell lung cancer (TROG 09.02 CHISEL): a phase 3, open-label, randomised controlled trial. *The Lancet Oncology* 2019; 20: 494-503.
33. Rowell NP, Williams CJ. Radical radiotherapy for stage I/II non-small cell lung cancer in patients not sufficiently fit for or declining surgery (medically inoperable). *The Cochrane database of systematic reviews* 2001: Cd002935.

34. Dosoretz DE, Katin MJ, Blitzer PH, Rubenstein JH, Salenius S, Rashid M, Dosani RA, Mestas G, Siegel AD, Chadha TT, et al. Radiation therapy in the management of medically inoperable carcinoma of the lung: results and implications for future treatment strategies. *International journal of radiation oncology, biology, physics* 1992; 24: 3-9.

35. Graham PH, Gebski VJ, Langlands AO. Radical radiotherapy for early nonsmall cell lung cancer. *International journal of radiation oncology, biology, physics* 1995; 31: 261-266.

36. Zhang HX, Yin WB, Zhang LJ, Yang ZY, Zhang ZX, Wang M, Chen DF, Gu XZ. Curative radiotherapy of early operable non-small cell lung cancer. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 1989; 14: 89-94.

37. Jeremic B, Shibamoto Y, Acimovic L, Djuric L. Randomized trial of hyperfractionated radiation therapy with or without concurrent chemotherapy for stage III non-small-cell lung cancer. *Journal of Clinical Oncology* 1995; 13: 452-458.

38. Sibley GS, Jamieson TA, Marks LB, Anscher MS, Prosnitz LR. Radiotherapy alone for medically inoperable stage I non-small-cell lung cancer: the Duke experience. *International journal of radiation oncology, biology, physics* 1998; 40: 149-154.

39. O'Rourke N, Roque IFM, Farre Bernado N, Macbeth F. Concurrent chemoradiotherapy in non-small cell lung cancer. *The Cochrane database of systematic reviews* 2010: Cd002140.
40. Blanke C, Ansari R, Mantravadi R, Gonin R, Tokars R, Fisher W, Pennington K, O'Connor T, Rynard S, Miller M, et al. Phase III trial of thoracic irradiation with or without cisplatin for locally advanced unresectable non-small-cell lung cancer: a Hoosier Oncology Group protocol. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 1995; 13: 1425-1429.

41. Cakir S, Egehan I. A randomised clinical trial of radiotherapy plus cisplatin versus radiotherapy alone in stage III non-small cell lung cancer. *Lung Cancer* 2004; 43: 309-316.

42. Clamon G, Herndon J, Cooper R, Chang AY, Rosenman J, Green MR. Radiosensitization with carboplatin for patients with unresectable stage III non-small-cell lung cancer: a phase III trial of the Cancer and Leukemia Group B and the Eastern Cooperative Oncology Group. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 1999; 17: 4-11.

43. Huber RM, Flentje M, Schmidt M, Pollinger B, Gosse H, Willner J, Ulm K. Simultaneous chemoradiotherapy compared with radiotherapy alone after induction chemotherapy in inoperable stage IIIA or IIIB non-small-cell lung cancer: study CTRT99/97 by the Bronchial Carcinoma Therapy Group. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 2006; 24: 4397-4404.

44. Jeremic B, Shibamoto Y, Acimovic L, Milisavljevic S. Hyperfractionated radiotherapy alone for clinical stage I nonsmall cell lung cancer. *International journal of radiation oncology, biology, physics* 1997; 38: 521-525.

45. Jeremic B, Shibamoto Y, Acimovic L, Milisavljevic S. Hyperfractionated radiation therapy with or without concurrent low-dose daily carboplatin/etoposide for stage III non-small-
cell lung cancer: a randomized study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 1996; 14: 1065-1070.

46. Schaake-Koning C, van den Bogaert W, Dalesio O, Festen J, Hoogenhout J, van Houtte P, Kirkpatrick A, Koolen M, Maat B, Nijs A, et al. Effects of concomitant cisplatin and radiotherapy on inoperable non-small-cell lung cancer. *The New England journal of medicine* 1992; 326: 524-530.

47. Soresi E, Clerici M, Grilli R, Borghini U, Zucali R, Leoni M, Botturi M, Vergari C, Luporini G, Scoccia S. A randomized clinical trial comparing radiation therapy v radiation therapy plus cis-dichlorodiammine platinum (II) in the treatment of locally advanced non-small cell lung cancer. *Seminars in oncology* 1988; 15: 20-25.

48. Atagi S, Kawahara M, Tamura T, Noda K, Watanabe K, Yokoyama A, Sugiura T, Senba H, Ishikura S, Ikeda H, Ishizuka N, Saijo N. Standard Thoracic Radiotherapy With or Without Concurrent Daily Low-dose Carboplatin in Elderly Patients with Locally Advanced Non-small Cell Lung Cancer: a Phase III Trial of the Japan Clinical Oncology Group (JCOG9812). *Japanese Journal of Clinical Oncology* 2005; 35: 195-201.

49. Auperin A, Le Pechoux C, Rolland E, Curran WJ, Furuse K, Fournel P, Belderbos J, Clamon G, Ulutin HC, Paulus R, Yamanaka T, Bozonnat MC, Uitterhoeve A, Wang X, Stewart L, Arriagada R, Burdett S, Pignon JP. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2010; 28: 2181-2190.

50. Nyman J, Hallqvist A, Lund J-Å, Brustugun O-T, Bergman B, Bergström P, Friesland S, Lewensohn R, Holmberg E, Lax I. SPACE – A randomized study of SBRT vs
conventional fractionated radiotherapy in medically inoperable stage I NSCLC.

*Radiotherapy and Oncology* 2016; 121: 1-8.

51. Ost DE, Gould MK. Decision making in patients with pulmonary nodules. *American journal of respiratory and critical care medicine* 2012; 185: 363-372.

52. Marchand C, Medford ARL. Relationship between endobronchial ultrasound-guided (EBUS)-transbronchial needle aspiration utility and computed tomography staging, node size at EBUS, and positron emission tomography scan node standard uptake values: A retrospective analysis. *Thoracic cancer* 2017; 8: 285-290.

53. Vial MR, O'Connell OJ, Grosu HB, Hernandez M, Noor L, Casal RF, Stewart J, Sarkiss M, Jimenez CA, Rice D, Mehran R, Ost DE, Eapen GA. Diagnostic performance of endobronchial ultrasound-guided mediastinal lymph node sampling in early stage non-small cell lung cancer: A prospective study. *Respirology (Carlton, Vic)* 2018; 23: 76-81.

54. Yasufuku K, Nakajima T, Waddell T, Keshavjee S, Yoshino I. Endobronchial ultrasound-guided transbronchial needle aspiration for differentiating N0 versus N1 lung cancer. *The Annals of thoracic surgery* 2013; 96: 1756-1760.

55. Leong TL, Loveland PM, Gorelik A, Irving L, Steinfort DP. Preoperative Staging by EBUS in cN0/N1 Lung Cancer: Systematic Review and Meta-Analysis. *Journal of bronchology & interventional pulmonology* 2019; 26: 155-165.

56. Herth FJ, Eberhardt R, Krasnik M, Ernst A. Endobronchial ultrasound-guided transbronchial needle aspiration of lymph nodes in the radiologically and positron emission tomography-normal mediastinum in patients with lung cancer. *Chest* 2008; 133: 887-891.

57. Szlubowski A, Zielinski M, Soja J, Annema JT, Sosnicki W, Jakubiak M, Pankowski J, Cmiel A. A combined approach of endobronchial and endoscopic ultrasound-guided
needle aspiration in the radiologically normal mediastinum in non-small-cell lung cancer staging—a prospective trial. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* 2010; 37: 1175-1179.

58. Sakairi Y, Hoshino H, Fujiwara T, Nakajima T, Yasufuku K, Yoshida S, Yoshino I. Validation of EBUS-TBNA-integrated nodal staging in potentially node-positive non-small cell lung cancer. *General thoracic and cardiovascular surgery* 2013; 61: 522-527.

59. Hwangbo B, Kim SK, Lee HS, Lee HS, Kim MS, Lee JM, Kim HY, Lee GK, Nam BH, Zo JI. Application of endobronchial ultrasound-guided transbronchial needle aspiration following integrated PET/CT in mediastinal staging of potentially operable non-small cell lung cancer. *Chest* 2009; 135: 1280-1287.

60. Oki M, Saka H, Ando M, Kitagawa C, Kogure Y, Seki Y. Endoscopic ultrasound-guided fine needle aspiration and endobronchial ultrasound-guided transbronchial needle aspiration: Are two better than one in mediastinal staging of non-small cell lung cancer? *The Journal of thoracic and cardiovascular surgery* 2014; 148: 1169-1177.

61. Shingyoji M, Nakajima T, Yoshino M, Yoshida Y, Ashinuma H, Itakura M, Tatsumi K, Iizasa T. Endobronchial ultrasonography for positron emission tomography and computed tomography-negative lymph node staging in non-small cell lung cancer. *The Annals of thoracic surgery* 2014; 98: 1762-1767.

62. Dooms C, Tournoy KG, Schuurbiers O, Decaluwe H, De Ryck F, Verhagen A, Beelen R, van der Heijden E, De Leyn P. Endosonography for mediastinal nodal staging of clinical N1 non-small cell lung cancer: a prospective multicenter study. *Chest* 2015; 147: 209-215.
63. Dong X, Qiu X, Liu Q, Jia J. Endobronchial ultrasound-guided transbronchial needle aspiration in the mediastinal staging of non-small cell lung cancer: a meta-analysis. The Annals of thoracic surgery 2013; 96: 1502-1507.

64. Adams K, Shah PL, Edmonds L, Lim E. Test performance of endobronchial ultrasound and transbronchial needle aspiration biopsy for mediastinal staging in patients with lung cancer: systematic review and meta-analysis. Thorax 2009; 64: 757-762.

65. Gu P, Zhao YZ, Jiang LY, Zhang W, Xin Y, Han BH. Endobronchial ultrasound-guided transbronchial needle aspiration for staging of lung cancer: a systematic review and meta-analysis. European journal of cancer (Oxford, England : 1990) 2009; 45: 1389-1396.

66. Evison M, Morris J, Martin J, Shah R, Barber PV, Booton R, Crosbie PAJ. Nodal staging in lung cancer: a risk stratification model for lymph nodes classified as negative by EBUS-TBNA. Journal of Thoracic Oncology 2015; 10: 126-133.

67. Annema JT, van Meerbeeck JP, Rintoul RC, Dooms C, Deschepper E, Dekkers OM, De Leyn P, Braun J, Carroll NR, Praet M, de Ryck F, Vansteenkiste J, Vermassen F, Versteegh MI, Veselic M, Nicholson AG, Rabe KF, Tournoy KG. Mediastinoscopy vs endosonography for mediastinal nodal staging of lung cancer: a randomized trial. Jama 2010; 304: 2245-2252.

68. Ernst A, Eberhardt R, Krasnik M, Herth FJ. Efficacy of endobronchial ultrasound-guided transbronchial needle aspiration of hilar lymph nodes for diagnosing and staging cancer. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer 2009; 4: 947-950.
69. Sakairi Y, Hoshino H, Fujiwara T, Nakajima T, Yasufuku K, Yoshida S, Yoshino I.
Validation of EBUS-TBNA-integrated nodal staging in potentially node-positive non-small cell lung cancer. *General thoracic and cardiovascular surgery* 2013; 61: 522-527.

70. Jhun BW, Um S-W, Suh GY, Chung MP, Kim H, Kwon OJ, Han J, Lee K-J. Clinical value of endobronchial ultrasound findings for predicting nodal metastasis in patients with suspected lymphadenopathy: a prospective study. *Journal of Korean medical science* 2014; 29: 1632-1638.

71. Yan J-H, Pan L, Chen X-L, Chen J-W, Yan L-M, Liu B, Guo Y-Z. Endobronchial ultrasound versus conventional transbronchial needle aspiration in the diagnosis of mediastinal lymphadenopathy: a meta-analysis. *SpringerPlus* 2016; 5: 1716-1716.
Figure Legends

Figure 1. Receiver operating characteristic curves of the prediction model in the institution of model development. The figure plots the area under the curve (AUC) for (A) N1|2|3 (vs. N0) disease (AUC=0.84) and (B) N2|3 (vs. N0|1) disease (AUC=0.85) in the development cohort and for (C) N1|2|3 (vs. N0) disease (AUC=0.86) and (D) N2|3 (vs. N0|1) disease (AUC=0.88) in the temporal validation cohort.

Figure 2. Observed vs. predicted frequencies of the prediction model in the institution of model development. The figure plots the probability of (A) N1|2|3 (vs. N0) disease and (B) N2|3 (vs. N0|1) disease by decile of expected risk in that group of the development cohort and the probability of (C) N1|2|3 (vs. N0) disease and (D) N2|3 (vs. N0|1) disease by decile of expected risk in that group of the temporal validation cohort. The observed probability for each decile is on the vertical axis, the predicted probability on the horizontal axis. A perfect model, where observed=predicted is shown by the line.

Figure 3. Observed vs. predicted frequencies for combined external validation cohort. The figure plots the probability of (A) N1|2|3 (vs. N0) disease and (B) N2|3 (vs. N0|1) disease by decile of expected risk in that group prior to calibration and the probability of (C) N1|2|3 (vs. N0) disease and (D) N2|3 (vs. N0|1) disease by decile of expected risk in that group after calibration. The observed probability for each decile is on the vertical axis, the predicted probability on the horizontal axis. A perfect model, where observed=predicted is shown by the line.

Figure 4. Observed vs. predicted frequencies for each institution of the external validation cohort prior to calibration. The figure plots the probability of (A) N1|2|3 (vs. N0) disease and (B) N2|3 disease (vs. N0|1) disease in Cleveland Clinic Foundation; the probability of (C) N1|2|3 (vs.
N0) disease and (D) N2|3 (vs. N0|1) disease at Johns Hopkins; and the probability of (E) N1|2|3 (vs. N0) disease and (F) N2|3 (vs. N0|1) disease at Henry Ford Hospital. The observed probability for each decile is on the vertical axis, the predicted probability on the horizontal axis. A perfect model, where observed=predicted is shown by the line.

Figure 5. Observed vs. predicted frequencies for each institution of the external validation cohort after calibration. The figure plots the probability of (A) N1|2|3 (vs. N0) disease and (B) N2|3 (vs. N0|1) disease in Cleveland Clinic Foundation; the probability of (C) N1|2|3 (vs. N0) disease and (D) N2|3 (vs. N0|1) disease at Johns Hopkins; and the probability of (E) N1|2|3 (vs. N0) disease and (F) N2|3 (vs. N0|1) disease at Henry Ford Hospital.
Table 1. Descriptive Statistics by N Stage in the Development Cohort (N=633)

|                          | N missing | N0 (N=412) | N1 (N=61) | N2/3 (N=160)* | p-value |
|--------------------------|-----------|------------|-----------|---------------|---------|
| Age (years), mean ±SD    | 0         | 68.99±9.3  | 66.57±10.01 | 65.23 ± 10.49 | 0.001 ‡ |
| Gender, n (%)            | 0         |            |           |               |         |
| Female                   | 194(63.6) | 28(9.2)    | 83(51.9%) |               | 0.549   |
| Male                     | 218(66.5) | 33(10.1)   | 77(23.5)  |               |         |
| Race, n (%)              | 4         |            |           |               |         |
| Asian                    | 13(76.5)  | 2(11.8)    | 2(11.7)   |               | 0.922   |
| Black                    | 31(64.6)  | 4(8.3)     | 13(27.1)  |               |         |
| Hispanic                 | 20(66.7)  | 3(10.0)    | 7(23.3)   |               |         |
| White                    | 347(65.0) | 50(9.4)    | 137(25.7) |               |         |
| ASA Score, n (%)         | 0         | 3(50.0)    | 0(0.0)    | 3(50.0)       | 0.814   |
| 1                        | 31(63.3)  | 4(8.2)     | 14(28.6)  |               |         |
| 2                        | 372(65.1) | 57(10.0)   | 142(24.9) |               |         |
| 4                        | 6(85.7)   | 0(0.0)     | 1(14.3)   |               |         |
| Smoking Status, n (%)    | 0         | 89(66.4)   | 11(8.2)   | 34(25.4)      | 0.748   |
| Current smoker           |           |            |           |               |         |
| Never smoker             | 39(66.1)  | 8(13.6)    | 12(20.3)  |               |         |
| Prior smoker             | 284(64.5) | 42(9.5)    | 114(25.9) |               |         |
| ECOG, n (%)              | 0         | 112(64.3)  | 23(13.2)  | 39(22.4)      | 0.181   |
| 0                        |           | 210(62.9)  | 28(8.4)   | 96(27.7)      |         |
| 1                        | 76(70.4)  | 9(60.8)    | 23(21.3)  |               |         |
| 2                        | 14(82.4)  | 1(5.9)     | 2(11.8)   |               |         |
| Size of the tumor, n (%) | 0         | 186(68.4)  | 23(8.5)   | 63(23.2)      | 0.298   |
| <= 3 cm                  |           | 132(62.9)  | 18(8.6)   | 60(28.6)      |         |
| >3 cm but <= 5 cm        | 94(62.3)  | 20(13.2)   | 37(24.2)  |               |         |
| >5 cm                    |           |            |           |               |         |
| Lobar location of the tumor, n (%) | 0 | 123(68.0) | 18(9.9) | 40(22.1) | 0.386 |
| Left upper lobe or lingula |         | 66(70.2) | 8(8.5) | 20(23.8) |         |
| Left lower lobe          | 140(65.7) | 18(8.5) | 55(25.8) |         |
| Right upper lobe         | 83(57.2)  | 17(11.7)  | 45(31.0)  |         |
| Right lower or middle lobe |       |            |           |               |         |
| Location, n (%)          | 0         | 323(78.4)  | 42(68.9)  | 111(69.4)     | 0.039   |
| Outer 2/3 of lung        |           | 89(21.6)   | 19(31.1)  | 49(30.6)      |         |
| Central 1/3rd of lung    |           |            |           |               |         |
| Histology, n (%)         | 0         | 203(61.3)  | 31(9.4)   | 97(29.3)      | 0.054   |
| Adenocarcinoma           |           | 158(72.8)  | 19(8.8)   | 40(18.4)      |         |
| Squamous cell carcinoma  | 32(55.2)  | 8(13.2)    | 18(31.0)  |         |
| Non-small cell carcinoma | 19(70.4)  | 3(11.1)    | 5(18.5)   |         |
| Other primary lung cancer |         |            |           |               |         |
| CT Characteristics, n (%)| 0         | 20(68.2)   | 3(13.6)   | 4(18.2)       | 0.016   |
| Cavitary                 |           | 32(88.9)   | 0(0.0)    | 4(11.1)       |         |
| Ground glass-semi-solid/infiltrate | 365(63.5) | 58(10.1) | 152(26.4) |         |
| Solid                    |           |            |           |               |         |
| Satellite lesion in same lobe, n (%) | 0 | 392(95.1) | 60(98.4) | 157(98.1) | 0.160 |
| No                       |           | 20(4.9)    | 1(1.6)    | 3(1.9)        |         |
| Yes                      |           |            |           |               |         |
| N stage by PET-CT, n (%) | 0         | 171(95.0)  | 2(1.1)    | 7(3.8)        | <0.001  |
| CT=N0 or N1, PET=N0      |           | 79(86.8)   | 4(4.4)    | 8(8.8)        |         |
| CT=N2 or N3, PET=N0      | 38(48.1)  | 28(35.4)   | 13(16.5)  |         |
| CT=N0 or N1, PET=N1      | 19(48.7)  | 16(41.0)   | 4(10.3)   |         |
| CT=N0 or N3, PET=N1      | 44(68.8)  | 2(3.1)     | 18(28.1)  |         |
| CT=N0 or N1, PET=N2 or N3| 61(33.9)  | 9(5.0)     | 110(61.1) |         |

SD= standard deviation; ASA= American Society of Anesthesiologists; ECOG= Eastern Cooperative Oncology Group; CT= computed tomography; PET= positron emission tomography
N stage as assessed by endobronchial ultrasound-guided transbronchial needle aspiration; *N2 and N3 are combined (N2|3).
P values are for ¡§chi-square test except where otherwise noted; ‡ANOVA
## Table 2. Multivariate Ordinal Logistic Regression Model for Prediction of N0 vs. N1 vs. N2|3 Disease

| Coefficient for N1|2|3 (vs. N0) disease | Odds ratio | 95% CI | p-value | Coefficient for N2|3 (vs. N0|1) disease | Odds ratio | 95% CI | p-value |
|------------------|------------------|------------|--------|---------|------------------|----------------|--------|--------|---------|
| Age (year) *     | -0.029           | 0.97       | 0.95   | 0.99    | 0.003            | -0.029         | 0.97   | 0.95   | 0.99    | 0.003 |
| Tumor location * | Outer 2/3rd of the lung | 0          | 1.00   |         |                  | 0              | 1.00   |        |        |       |
|                  | Central 1/3rd of the lung | 0.486      | 1.62   | 1.03    | 2.55             | 0.034          | 0.486  | 1.62   | 1.03    | 2.55  |
| Tumor histology *| Adenocarcinoma    | 0          | 1.00   |         |                  | 0              | 1.00   |        |        |       |
|                  | Squamous-cell carcinoma | -0.821     | 0.44   | 0.28    | 0.68             | <0.001         | -0.821 | 0.44   | 0.28    | <0.001|
|                  | Non- small cell lung carcinoma | 0.063      | 1.06   | 0.55    | 2.03             | 0.847          | 0.063  | 1.06   | 0.55    | 2.03  |
|                  | Other primary     | -0.409     | 0.66   | 0.25    | 1.75             | 0.409          | -0.409 | 0.66   | 0.25    | 1.75  |
| N stage by PET-CT | CT= N0 or N1, PET=N0 | 0          | 1.00   |         |                  | 0              | 1.00   |        |        |       |
|                  | CT= N0 or N3, PET=N0 | 1.173      | 3.23   | 1.29    | 8.08             | 0.012          | 0.979  | 0.97   | 0.92    | 7.67  |
|                  | CT= N0 or N1, PET=N1 | 3.083      | 21.82  | 9.62    | 49.48            | <0.001         | 1.593  | 1.59   | 1.85    | 13.03 |
|                  | CT= N0 or N3, PET=N1 | 2.990      | 19.89  | 7.76    | 50.93            | <0.001         | 0.932  | 0.93   | 0.69    | 9.24  |
|                  | CT= N2 or N3, PET=N2 or N3 | 2.259      | 9.57   | 4.00    | 22.88            | <0.001         | 2.359  | 2.35   | 4.10    | 27.32 |
|                  | CT= N2 or N3, PET=N2 or N3 | 3.711      | 40.90  | 19.25   | 86.90            | <0.001         | 3.748  | 3.74   | 18.58   | 97.06 |

| Constant‡ | -0.890 | 0.233 | -1.1576 | 0.131 |

CI= confidence interval; CT= computed tomography; PET= positron emission tomography;

N1|2|3= N stage equal or greater than 1; N2|3= N2 or N3 disease; N1|2=N0 or N1 disease

*Variables did not violate the proportional odds assumption in the univariate analysis. Therefore, the coefficients for N1|2|3 (vs. N0) disease are the same to the coefficients of N2|3 (vs. N0|1) disease.

†Variable violated the proportional odds assumption in the univariate analysis. Therefore, two slope parameters were obtained, one for N1|2|3 (vs. N0) disease and one for N2|3 (vs. N0|1) disease. Both coefficients are shown.

‡Two constants were calculated: one for the formula used to predict N1|2|3 (vs. N0) disease and one for the formula used to predict N2|3 (vs. N0|1) disease.
### Table 3. Model Performance at Outside Institutions: Predictions Prior to Calibration

| Institution | Brier score | AUC (95% CI) N1|2|3 (vs. N0) disease | Hosmer-Lemeshow p-value | AUC (95% CI) N2|3 (vs. N0|1) disease |
|-------------|-------------|----------------|-------------------|-------------------------|-----------------------|----------------|
| CCF (N=310) | 0.34        | 0.84 (0.79-0.88) | 0.286             | <0.001                  | 0.87 (0.83-0.91)      |
| JH (N=186)  | 0.46        | 0.82 (0.74-0.89) | <0.001            | 0.82 (0.76-0.89)        |
| HFH (N=226) | 0.36        | 0.91 (0.87-0.95) | <0.001            | 0.92 (0.88-0.95)        |

AUC= Area under the Receiver Operating Characteristics curve; CI= confidence interval; CCF= Cleveland Clinic Foundation; JH= Johns Hopkins; HFH= Henry Ford Hospital; N1|2|3= N stage equal or greater than 1; N2|3= N2 or N3 disease; N1|2=N0 or N1 disease
Figure 1. Receiver operating characteristic curves of the prediction model in the institution of model development. The figure plots the area under the curve (AUC) for (A) N1|2|3 (vs. N0) disease (AUC=0.84) and (B) N2|3 (vs. N0|1) disease (AUC=0.85) in the development cohort and for (C) N1|2|3 (vs. N0) disease (AUC=0.86) and (D) N2|3 (vs. N0|1) disease (AUC=0.88) in the temporal validation cohort.
Figure 2. Observed vs. predicted frequencies of the prediction model in the institution of model development. The figure plots the probability of (A) N1|2|3 (vs. N0) disease and (B) N2|3 (vs. N0|1) disease by decile of expected risk in that group of the development cohort and the probability of (C) N1|2|3 (vs. N0) disease and (D) N2|3 (vs. N0|1) disease by decile of expected risk in that group of the temporal validation cohort. The observed probability for each decile is on the vertical axis, the predicted probability on the horizontal axis. A perfect model, where observed=predicted is shown by the line.
Figure 3. Observed vs. predicted frequencies for combined external validation cohort. The figure plots the probability of (A) N1|2|3 (vs. N0) disease and (B) N2|3 (vs. N0|1) disease by decile of expected risk in that group prior to calibration and the probability of (C) N1|2|3 (vs. N0) disease and (D) N2|3 (vs. N0|1) disease by decile of expected risk in that group after calibration. The observed probability for each decile is on the vertical axis, the predicted probability on the horizontal axis. A perfect model, where observed=predicted is shown by the line.
Figure 4: Observed vs. predicted frequencies for each institution of the external validation cohort prior to calibration. The figure plots the probability of (A) N1|2|3 (vs. N0) disease and (B) N2|3 disease (vs. N0|1) disease in Cleveland Clinic Foundation; the probability of (C) N1|2|3 (vs. N0) disease and (D) N2|3 (vs. N0|1) disease at Johns Hopkins; and the probability of (E) N1|2|3 (vs. N0) disease and (F) N2|3 disease (vs. N0|1) at Henry Ford Hospital. The observed probability for each decile is on the vertical axis, the predicted probability on the horizontal axis. A perfect model, where observed=predicted is shown by the line.
Figure 5: Observed vs. predicted frequencies for each institution of the external validation cohort after calibration. The figure plots the probability of (A) N1|2|3 (vs. N0) disease and (B) N2|3 (vs. N0|1) disease in Cleveland Clinic Foundation; the probability of (C) N1|2|3 (vs. N0) disease and (D) N2|3 (vs. N0|1) disease at Johns Hopkins; and the probability of (E) N1|2|3 (vs. N0) disease and (F) N2|3 (vs. N0|1) disease at Henry Ford Hospital.
Online supplement

**A Prediction Model to Help with Oncologic Mediastinal Evaluation for Radiation: HOMER**

Gabriela Martinez-Zayas; Francisco A. Almeida, MD, MS; Michael J. Simoff, MD; Lonny Yarmus, DO; Sofia Molina, MD; Benjamin Young, MD; David Feller-Kopman, MD; Ala-Eddin S. Sagar, MD; Thomas Gildea, MD, MS; Labib G. Debiane, MD; Horiana B. Grosu, MD, MPH; Roberto F. Casal, MD; Muhammad H. Arain, MBBS; George A. Eapen, MD; Carlos A. Jimenez, MD; Laila Z. Noor, MD; Shiva Baghaie, MPH; Juhee Song, PhD; Liang Li, PhD; David E. Ost, MD, MPH

**Corresponding Author Information:**

David E. Ost, MD, MPH, The University of Texas MD Anderson Cancer Center, Pulmonary Department, 1515 Holcombe Blvd, Unit 1462, Houston, Tx, 77030; Email: 

dost@mdanderson.org; phone: 713-745-8775; fax: 713-749-4922.

**Conflict of Interest:** None to disclosure
Introduction

We used ordinal logistic regression to generate a model which predicts N0 vs. N1 vs. N2 or N3
(N2|3) metastatic nodal disease in patients with non-small cell lung carcinoma (NSCLC) as
staged by endobronchial ultrasound with transbronchial needle aspiration (EBUS-TBNA) to help
drive staging and treatment decision in patients in which SABR is being considered.(1, 2)

Previous studies have focused on creating binary models for predicting metastatic nodal disease
when surgery is being considered as a treatment option. (3-10) However, no ordinal logistic
regression model has been proposed to predict the probability of metastatic nodal disease when
stereotactic ablative radiotherapy (SABR) is being considered as treatment option and where
separating N0 from N1 disease is important.(11, 12)

Methods

Variable definitions

Definitions of all variables were developed prior to data abstraction and provided to all sites. For
computed tomography (CT) scans, abnormal lymph nodes were defined as being greater than or
equal to 1 cm in their short axis. Lymph node N stage was determined by review of the radiology
report and further review by an interventional pulmonologist or an interventional pulmonary
fellow under supervision in order to assign an N stage to the patient. If both contrast and non-
contrast CT were available, the contrast-enhanced images were used to determine CT N stage.

Positive emission tomography (PET) N stage was based on the radiologist’s interpretation of
mediastinal lymph node fluorodeoxyglucose (FDG) activity. In some cases standardized uptake
values (SUV) measurements were recorded. In those cases that SUV measures on lymph nodes
were available, a SUV value equal or greater than 2.5 was considered as positive. Based on the
radiologist’s reading and further review by an interventional pulmonologist or a supervised interventional pulmonary fellow, the PET N stage of the lesion was determined.

We specified a priori that we would use positron emission tomography and computed tomography (PET-CT) N stage using interactions between CT N stage and PET N stage, based on previous work that suggested that sensitivity of PET for mediastinal lymph node involvement is conditional on the size of the node on CT.(13) Because PET-CT images do not use contrast, we combined N0 and N1 disease together for CT but kept N0 and N1 separate for PET.

Tumor location was defined based on the location of the center of the tumor. Tumors defined as central 1/3rd location tumors were those located within the inner 1/3rd of the hemi-thorax (with the hilum being the center) or tumors that had their center within the segmental airways on CT. Tumors located in the outer 2/3rds of the hemithorax required that the center of the tumor not touch the segmental airways and that the tumor center was located outside the central 1/3rd of the lung (Figure E1).(14, 15) The interpretation was made by an interventional pulmonologist or an interventional pulmonary fellow after training and under supervision.

The mean age of the patients in the entire population of the development cohort was 67.8 years with a standard deviation of 9.8. Hence, 68% of the patients from our study are between 60 and 80 years of age (when rounded to the nearest decade) and predictions made by the model are most accurate for patients in this age range.

**Outcome Selection and Rationale**

The outcome variable that we chose to predict was whether or not EBUS-TBNA would find N0, N1 and N2|3 disease in a given patient. Previous studies predicted N2|3 disease as determined by mediastinoscopy or thoracotomy in order to guide surgical decisions. (4-8, 16-20) This
distinction is important because one of the important clinical applications for our prediction rule is informing the decision on whether EBUS-TBNA should be done in patients who are being considered for SABR. Based on current guidelines, mediastinoscopy has been replaced by EBUS-TBNA as the recommended first sampling technique for staging the mediastinum in patients with NSCLC. (21) Prior studies predicted the diagnostic yield of mediastinoscopy for N2|3 disease. With EBUS-TBNA replacing mediastinoscopy as the first sampling technique, O’Connell et al created a prediction model for diagnostic yield of EBUS-TBNA for N2|3 disease (the HAL model). (3, 4) However, mediastinoscopy cannot sample N1 and the HAL model did not include a prediction rule for the diagnostic yield of N1 disease by EBUS-TBNA, so older prediction rules do not apply to fragile patients considered for therapy with SABR.

Arguably, it would be best to predict N0, N1 and N2|3 disease as determined by thoracotomy, since that is the gold standard for staging. That would be necessary to know if we were trying to determine the sensitivity and specificity of EBUS-TBNA or to predict the pretest probability of disease. But this study is not determining sensitivity and specificity of EBUS-TBNA, so use of a surgical gold-standard is not applicable. We are also not predicting pretest probability. All we are doing is predicting the probability that EBUS-TBNA will be positive for N0, N1, or N2|3 disease, meaning that we are predicting diagnostic yield.

So the issue is whether knowing the probability that EBUS-TBNA will yield a positive result for N0, N1 and N2|3 disease is clinically useful. Since the positive predictive value (PPV) of EBUS-TBNA is close to 100%, our assertion is that this knowledge is clinically useful. Below some threshold diagnostic yield, EBUS-TBNA will not be cost-effective.

We should note that the prediction of diagnostic yield for a test is fundamentally different than predicting the results of the gold standard (thoracotomy in this case). A rule which accurately
predicts the pretest probability of nodal disease by thoracotomy at 15% does not necessarily imply a given test is or not warranted. Consider conventional TBNA vs. EBUS-TBNA for this same patient. A rule that accurately predicts the probability of nodal disease by thoracotomy at 15% does not necessarily imply that conventional TBNA is warranted. What we really want to know is the probability the test (conventional TBNA) will be positive. We would also like to know the probability that alternative tests (e.g. EBUS-TBNA) will be positive. A prediction rule based on thoracotomy results would be good for estimating the pretest probability of disease but could not tell us whether or not conventional TBNA or EBUS-TBNA is worthwhile. Knowing the pretest probability is not sufficient to allow us to say that a given test is warranted. To guide the decision on whether or not to do a given test requires information on the diagnostic yield of that specific test in a given patient. This is subtly but significantly different than the probability that there is disease present as assessed by the gold standard (i.e. pretest probability).

If the sensitivity of conventional TBNA or EBUS-TBNA was fixed across all groups, then diagnostic yield would be equal to the probability of nodal disease as determined by thoracotomy (i.e. pretest probability) multiplied by sensitivity, since TBNA specificity is essentially 100%. But if sensitivity varies significantly across strata such an analytic approach, while tempting, will not work unless we know the sensitivity for each individual strata.

However, the sensitivity of EBUS-TBNA sensitivity varies. It has lower sensitivity for N1 and N2|3 disease in certain patient groups with a low pretest probability of N1 and N2|3 disease. A meta-analysis by Gu et al. found that sensitivity of EBUS varies between studies depending on the PET-CT status of the nodes in the patients that were enrolled. (22) So a single pooled estimate of sensitivity is not appropriate for all patients. A more empiric approach is to directly measure and predict diagnostic yield for EBUS-TBNA, which is what we have done. We do...
sacrifice knowledge of true disease prevalence since we could not do thoracotomy in all patients with negative EBUS-TBNA results, but the ability to predict diagnostic yield is practical and clinically relevant. Our estimates of prN1 and prN2|3 are therefore lower than the true probability of N1 and N2|3, but they are close given that EBUS-TBNA has a specificity of close to 100% and in most strata the sensitivity is high.

Model Development

Logistic regression modeling has many applications in diverse areas, including clinical studies and epidemiology. It is used when the dependent variable is dichotomous (“event” or “non-event”) to model the probabilities of events. Binary logistic regression is extendable to more than two response levels. If the response variable takes values that have no ordering (e.g. voting Democratic, Green, Independent, Republican), then the response is nominal. If the response takes values that have an intrinsic order (good, better, best), the outcome is ordinal. (23) In our model, outcomes did have an intrinsic order, therefore we used an ordinal logistic regression model. The ordinal response variable we used had three ordinal response categories (N0 < N1 < N2|3 as assessed by EBUS-TBNA) and was considered the dependent variable.

Ordinal logistic regression’s key assumption is the proportional odds (or parallel lines) assumption. The assumption is that the effects of any of the explanatory variables are consistent (proportional) across the different thresholds (the splits between each pair of categories of the ordinal outcome variable). So if the odds ratio associated with adenocarcinoma is 2 when evaluating N1|2|3 (vs. N0), then the odds ratio is also 2 when evaluating N2|3 (vs. N0|1). In other words, the assumption is that the explanatory variables have the same effect on the odds regardless of the threshold.(24) In this example, the thresholds are (N0 to N1) and (N1 to N2).
The model development cohort used data from The University of Texas MD Anderson Cancer Center (MDACC). We arrived at a model by first conducting univariate analysis. We specified a priori that all variables with an overall p-value less than 0.2 were candidate variables for the multivariate analysis (Table E1). To remain in the final multivariate model, variables had to have a p-value of less than 0.05. Since this is an ordinal logistic regression, we had to check for proportional odds assumption violations. For the variables that had an overall p-value less than 0.2 in the univariate analysis (candidate variables for the multivariate analysis) we checked for proportional odds assumption violations using the Score test. The Score test’s null hypothesis is that the explanatory variable does not violate the proportional odds assumption(25). A p-value less than 0.05 rejects the null, therefore there is a violation of the proportional odds assumption (accepting the alternative hypothesis).(25) For variables that violated the assumption, we specified a priori that we would allow different slope parameters for each outcome. We would therefore have different odds ratios for prediction of N1|2|3 (vs. N0) disease and for N2|3 vs. (N0|1) disease. Allowing different slopes for the variables that have proportional odds assumption violation but having the same slopes for those variables that do not violate it makes our model a partial proportional odds model.(26) By having a partial proportional odds model we obtain two formulas. With one we predict the probability of N1|2|3 (vs. N0) disease and with the second one we predict the probability of N2|3 (vs. N0|1) disease. The formula for predicting the probability of N1|2|3 disease is of the following form:

\[ \text{prN1|2|3} = \frac{\exp(A)}{1+\exp(A)} \]

where

\[ A = -0.89 - 0.0292 \times \text{(age of the patient)} + 0.4864 \times \text{(location of the tumor = central 1/3rd of the lung)} - 0.8217 \times \text{(tumor histology = squamous cell carcinoma)} + 0.0635 \times \text{(tumor histology = non-small cell lung carcinoma)} - 0.4097 \times \text{(tumor histology = other primary lung cancer)} + 1.1738 \times \text{(CT=N2|3,} \]
\[ \Pr(N_{2|3}) = \frac{\exp(C)}{1 + \exp(C)}, \]

where

\[ C = -1.1576 - 0.0292 \times \text{age of the patient} + 0.4864 \times \text{location of the tumor = central 1/3rd of the lung} - 0.8217 \times \text{tumor histology = squamous cell carcinoma} + 0.0635 \times \text{tumor histology = non-small cell lung carcinoma} - 0.4097 \times \text{tumor histology = other primary lung cancer} + 0.9798 \times \text{CT=N2|3, PET=N0} + 1.5937 \times \text{CT=N0|1, PET=N1} + 0.9323 \times \text{CT=N2|3, PET=N1} + 2.3599 \times \text{CT=N0|1, PET=N2|3} + 3.7486 \times \text{CT=N2|3, PET=N2|3}. \]

**Model Performance Assessment**

Assessing performance of a statistical prediction model includes measuring whether the model has the ability to correctly separate subjects with and without the outcome (discrimination) and measuring how closely the predicted probabilities are to the actual observed frequency of outcomes (calibration).\(^{(27, 28)}\) We assessed discrimination using the area under the receiver operating characteristic (ROC) curve (AUC), which is a measure of a test’s discriminatory power. It reflects the probability that a prediction rule will be able to discriminate correctly which of two patients is at a higher risk. Note that for logistic regression this is akin to the prediction rule accurately assessing the rank order of probabilities for any two patients. It does not necessarily inform us about the ability of a test to predict absolute risk. We therefore also assessed calibration using the Brier score, the Hosmer-Lemeshow test and observed vs. predicted graphs.\(^{(27)}\) These methods assess the agreement between observed and predicted outcomes on an absolute scale.
The null hypothesis of the Hosmer-Lemeshow goodness-of-fit test is that the model fits the data well, while the alternative hypothesis is that there is lack of fit. Thus, a p-value < 0.05 rejects the null hypothesis that the model fits the data. (29)

The Brier score ranges from 0 to 2 for categorical predictions with three categories, where 0 represents a perfect score and 2 the worst score. (30) However, one limitation of the Brier score is that it fails to capture the magnitude of the inaccuracy when applied to ordinal categorical data with more than two outcomes. For example, incorrectly predicting a patient has N1 disease when they have N2 disease is considered the same as predicting the patient has N0 disease when they have N2 disease.

We therefore also used observed vs. predicted graphs. The graphs plot the predicted probability of disease by decile of expected risk (x-axis) versus the observed probability of disease (y-axis) for that decile. A perfect model where observed is equal to predicted is shown by a 45 degree line. (27, 28)

Since model performance in the development cohort usually overestimates performance, model validation requires assessment using different data sources. This can be done by using patients from the same institution from a different time-period (temporal validation cohort) or from other hospitals (external validation cohorts). (28, 31) Therefore, we externally and temporally validated our model.

For external validation, we used data from three other centers: Johns Hopkins, Henry Ford Hospital, and the Cleveland Clinic Foundation. We validated the model using all of three centers together (combined external validation cohort) and for each particular institution separately.

After assessing for calibration in the entire external validation cohort, the model was calibrated
for the combined data and for each institution separately using the general calibration method presented by Steyerberg et. al. (see following subsection).

**General calibration method**

After assessing model performance in the external validation cohort, the model was recalibrated for the combined external validation cohort and for each outside institution using the general calibration method model proposed by Steyerberg et. al. In this method, a logistic regression model is first fitted to the linear predictor (log odds) generated by the model as the only covariate. The logistic regression model gives an intercept \(a\) and a slope \(b\). When the linear predictor is multiplied by \(b\) (slope) and \(a\) (intercept) is added, the raw predicted probabilities are adjusted and there is improvement in model fit with no change in model discrimination.

Since the HOMER model is an ordinal logistic regression model with partial proportional odds assumption violations, we had two formulas, one for predicting \(pN1|2|3\) (vs. \(pN0\)) disease and another for predicting \(pN2|3\) (vs. \(pN0|1\)) disease. Therefore, when calibrating for the combined external validation cohort and for each institution, we fitted two logistic regressions, one for the linear predictor of \(pN1|2|3\) (vs. \(pN0\)) disease as the only covariate and another for the linear predictor of \(pN2|3\) (vs. \(pN0|1\)) disease as the only covariate.

To determine \(a\) and \(b\) for the combined external validation cohort, the logistic regression model for the combined external validation cohort was of the following form, log odds (observed \(N1|2|3\) disease) = \(a + b*(\text{linear predictor for} \ N1|2|3 \text{ disease})\); and log odds (observed \(N2|3\) disease) = \(a' + b'*\text{(linear predictor for} \ N2|3 \text{ disease})\). Since two logistic regressions were fitted, two sets of
slope and intercept were obtained. Both sets of slope and intercept were evaluated and the one with the lowest Brier score was selected.

For calibration of each center, slopes were set to unity for all institutions ($b=1$), and only the pair of intercepts for each institution were estimated by doing the two logistic regression models. The logistic regression model for each center was of the following form: log odds (observed N1|2|3 disease) = a+1*(linear predictor for N1|2|3 disease); and log odds (observed N2|3 disease)=a’+1*(linear predictor for N2|3 disease). For each institution, the adjustment intercepts obtained in the fitted logistic regression were added to their respective linear predictors (one for N1|2|3 (vs. N0) disease and another for N2|3 (vs. N0|1) disease). Performance was then reassessed using the Brier Score, Hosmer-Lemeshow goodness-of-fit test and AUC. The process of fitting a logistic regression to the linear predictors (with the slope set to unity) followed by addition of the obtained calibration intercepts (to their respective linear predictors) and assessment of model performance (using the Brier Score, Hosmer-Lemeshow p-value and AUC) was repeated until the pair of intercepts with the lowest Brier score and highest Hosmer-Lemeshow p-value was found (Table E4).

**Results**

**Model Development Cohort**

Table 1 shows the descriptive statistics for the development cohort, stratified by EBUS N stage. Table E1 shows the univariate ordinal logistic regression results. Of the candidate variables, only N stage by PET-CT had proportional odds assumption violations. This variable remained in the final multivariate model, along with age, location of the tumor, and tumor histology (Table 2).

Since we allowed different slopes for the variable that violated the proportional odds assumption
(N stage by PET-CT) but we had the same slopes for the variables that did not violate it (age, tumor location and tumor histology), our model is a partial proportional odds model.

The model had acceptable discrimination, with a ROC AUC of 0.84 (95% CI=0.81-0.87) for predicting N1|2|3 (vs. N0) disease and 0.85 (95% CI=0.82-0.89) for predicting N2|3 (vs. N0|1) disease (Figure 1A-B). Model fit was acceptable as assessed by the Hosmer-Lemeshow test (p=0.960), observed vs. predicted plots (Figure 2A-B), and Brier Score (0.36). Table E5 shows the predicted probabilities of N0, N1 and N2|3 disease for a given change in the explanatory variables for patients between 60 to 80 years of age.

**Temporal and External Validation Cohort**

Table E2 shows the descriptive statistics for the temporal validation cohort. Figures 2C-D and 3C-D show model discrimination and calibration as assessed by observed vs. predicted graphs respectively.

Table E3 shows the descriptive statistics for the external validation cohort. Figure E2 shows model discrimination for the combined external validation cohort. Figures 3A-B and 4 show model calibration as assessed by observed vs. predicted graphs for the combined external validation cohort and for each institution, respectively. Table 3 provides model performance characteristics for each institution prior to model calibration.

Subsequent model calibration was done by the general calibration model described by Steyerberg.(32) For the combined external validation cohort, both calibration intercepts were the same (0.75) and the slopes were off by 0.01 (1.13 vs. 1.14). The two different calibrations were evaluated (intercept=0.75 and slopes 1.13 and 1.14), and the one with a lower Brier score was selected (intercept=0.75 and slope=1.14). Model fit was acceptable after calibration (multiplying
by 1.14 and adding 0.75) with a Hosmer-Lemeshow p-value of 0.094 and improvement in observed vs. predicted graph (Figures 3C-D). After calibration to each center, model fit was acceptable for every institution as assessed by the Brier scores, Hosmer-Lemeshow test and observed vs. predicted graphs (Figure 5 and Table E4).

**Discussion**

With an ordinal logistic regression model we can calculate probabilities of N0, N1 and N2|3 disease by EBUS-TBNA. By knowing the odds of N1|2|3 (vs. N0) disease, we know the prN0 and prN1|2|3 disease. By knowing the odds of N2|3 (vs. N0|1) disease we know the prN2|3 disease. We can then calculate the prN1 by subtracting the prN2|3 disease from the prN1|2|3 disease. Our Help with Oncologic Mediastinal Evaluation for Radiation (HOMER) model has the ability to distinguish the prN0 from prN1 and prN2|3 disease; this is particularly important for patients that are candidates for treatment with stereotactic ablative radiotherapy (SABR), since SABR would be a suboptimal treatment for patients with N1 disease. (11, 12)

The HOMER model provides substantive additional insights as compared to the HAL model and other previous publications that have used binary prediction models (predicting prN0|1 vs. prN2|3 disease; prN0 vs. prN2 disease, or prN0 vs. prN1|2|3 disease). (3-7, 10, 33) By using ordinal logistic regression, HOMER provides additional insights into each of the predictive variables for every N stage that binary models lack. For instance, in the HAL paper centrally located tumors had higher odds of having N2|3 (relative N0 or N1) disease, but whether or not central tumors were a risk factor N1 disease could not be answered. HOMER demonstrates that centrally located tumors do have a higher risk of having N1 disease (relative to N0) disease and that the odds ratio for N1 vs. N0 is roughly the same as for N2|3 vs. N0|1. Figure E3 provides further insight by quantifying the predicted prN0, prN1 and prN2|3 disease according to location.
of the tumor for a given N stage by PET. The HOMER model provides insights that are not clinically obvious. For example, in 80 year-old patients with N2|3 disease by PET, the most likely stage by EBUS is either N2|3 or N0 depending on whether the tumor is centrally located. It also shows that N1 disease is always the least likely EBUS stage by a large margin in patients with PET N2|3 disease (young or old). Physicians often believe that lung cancer spreads regionally going from N0 to N1 to N2 to N3. If a patient has N2|3 disease by PET, then the clinically intuitive answer is that most likely the patient will have N2|3 disease by EBUS. If the patient does not have N2 disease, the paradigm suggest that N1 would be the second most probable answer, and N0 would be the least likely answer. However, HOMER demonstrates that this is untrue (Figure E3C). Note the ordinal nature of the outcomes for the HOMER model is not predicated on assumptions about the pathophysiology of how lung cancer metastasizes, but rather it is based on survival probability of different N stages. (34-36)

Tumor Size

The HOMER model failed to demonstrate a relationship between tumor size and higher N stage by EBUS-TBNA after adjusting for age, tumor location and histology, and N status by PET-CT. In contrast, prior studies reported an association between larger tumors and probability of nodal disease. (4-7) However, none of those studies adjusted for PET-CT N stage in their analysis. If PET-CT N stage is associated with tumor size, it can act as a confounder for the relationship between tumor size and nodal metastatic disease. We conducted one-way analysis of variance to test the relation between PET-CT N stage and tumor size. We found that indeed there is an association between PET-CT N stage and tumor size (F test (5,628) =2.4, p=0.036). We wanted to further test if by adding tumor size to the HOMER model (even if it was not a candidate variable for the multivariate analysis) model performance would improve and observed that
model performance did not significantly improve. With size forced into the model, AUC was 0.84 (95% CI= 0.81-0.87) for predicting prN1|2|3 (vs. prN0) disease and 0.86 (95% CI=0.83-0.89) for predicting prN2|3 (vs. prN0|1) disease in the development cohort. This is unchanged from HOMER. The Brier Score showed a slight improvement (0.35 with size included vs. 0.36 in the HOMER model). However, when size was included in the model the Hosmer Lemeshow goodness of fit-test had a significant p-value of 0.015, indicating poor calibration. This shows us that, if we force tumor size into the HOMER model, the result is similar in terms of discrimination but model fit is not acceptable.(29)

**PET-CT alone**

Hypothesizing that the main factor driving the predictions made by the HOMER model was PET-CT, we tested model performance of only PET-CT. We observed that indeed the majority of discrimination of the HOMER model was provided by PET-CT, which had an AUC of 0.81 (95%CI= 0.78-0.85) for predicting N1|2|3 (vs. N0) disease and 0.83 (95% CI=0.80-0.87) for predicting N2|3 (vs. N0|1) disease in the development cohort. However, if only PET-CT was used, model calibration was poor (Hosmer Lemeshow p-value (<0.001).

**N1 Disease Scenario and using HOMER to Estimate Negative Predictive Value**

If we know the sensitivity of EBUS, we can use HOMER to estimate the post-test probability that a patient with a negative EBUS actually has N0 disease. This is predicated on the specificity of EBUS being 100%. Given these assumptions, then:

\[
\text{Prevalence} \times \text{sensitivity} = \text{Patients who Test+; provided specificity is 100%}.
\]

\[
\text{Prevalence} = \frac{\text{EBUS+}}{\text{sensitivity}}
\]
Therefore the prevalence of N2|3 disease = probability EBUS+ N2|3 / sensitivity.

HOMER calculates the probability of EBUS being positive for N2|3, so the predicted probability divided by our sensitivity gives an estimate of the true prevalence of N2|3 disease.

We then repeat this for N1 disease.

Now we know that true prevalence of N0 + N1 + N2|3 = 1, so

Prevalence of N0 = 1 – prevalence N1 – Prevalence N2|3

Applying this to a 60 year-old patient with adenocarcinoma in the outer two-thirds of the lung, PET-CT N1 disease then the corresponding probabilities for true N0 disease are:

Probability that EBUS will be positive and show N2|3 disease (prN2|3) is 0.211 as provided in Table E5. Let sensitivity = 0.8.

Prevalence of N2|3 = prN2|3 / sensitivity = 0.211/0.8 = 0.264

Prevalence of N1 = prN1 / sensitivity = 0.397/0.8 = 0.496

Prevalence of N0 = 1 – prevalence N1 – Prevalence N2|3 = 1 – 0.264 – 0.496 = 0.24

Now we know from Table E5 that about 39.2% of patients will have a negative EBUS result in all lymph node (prN0). Remembering that specificity is assumed to be 1, we calculate the probability of truly having N0 disease given that the EBUS result was negative:

Prevalence of N0 disease / prN0 = 0.24/0.392 = 0.61.

Application to Outside Institutions

The HAL model was the first study to externally validate a prediction model for N2|3 vs. N0|1 disease.(3) However, the HOMER model is, to our knowledge, the first study of EBUS-TBNA to
both externally and temporally validate a prediction model for N0, N1 and N2|3 disease. Our findings suggest that the HOMER model is stable through time in the institution where it was created. We hypothesize that it can be used in any outside institution in a practical fashion. First collect institution specific data and apply HOMER to arrive at individual level predictions. Measure discrimination using the AUC. If it is good, then assess calibration by comparing the observed vs. expected outcomes, measure the Brier score and assess the Hosmer-Lemeshow test. Most likely calibration will be required. Calibrate with the general calibration method as described above to arrive at an institution specific intercept.(32) Then collect an additional new validation cohort from the same institution. Use the calibrated HOMER model (i.e. using the intercept that you derived from the first cohort) and reassess model discrimination and calibration. In future studies we will study the sample size required for center-specific calibration. We will also temporally validate the model in the three outside hospitals used for validation to assess if indeed the calibrated model as derived in this study is temporally stable through time at these other centers. If the calibrated model performs, then HOMER is potentially scalable to many institutions, since the data to run it is readily available in most electronic health records.
References

1. Hunink MGM, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, Glasziou PP. Choosing the best treatment. Decision Making in Health and Medicine: Integrating Evidence and Values, 2 ed. Cambridge: Cambridge University Press; 2014. p. 53-77.

2. Louie AV, Senan S, Patel P, Ferket BS, Lagerwaard FJ, Rodrigues GB, Salama JK, Kelsey C, Palma DA, Hunink MG. When is a biopsy-proven diagnosis necessary before stereotactic ablative radiotherapy for lung cancer?: A decision analysis. Chest 2014; 146: 1021-1028.

3. O’Connell OJ, Almeida FA, Simoff MJ, Yarmus L, Lazarus R, Young B, Chen Y, Semaan R, Saettele TM, Cicion J, Bedi H, Kliment C, Li L, Sethi S, Diaz-Mendoza J, Feller-Kopman D, Song J, Gildea T, Lee H, Grosu HB, Machuzak M, Rodriguez-Vial M, Eapen GA, Jimenez CA, Casal RF, Ost DE. A prediction model to help with the assessment of adenopathy in lung cancer: HAL. American journal of respiratory and critical care medicine 2017; 195: 1651-1660.

4. Shafazand S, Gould MK. A clinical prediction rule to estimate the probability of mediastinal metastasis in patients with non-small cell lung cancer. Journal of Thoracic Oncology 2006; 1: 953-959.

5. 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. The Journal of Thoracic and Cardiovascular Surgery 2012; 144: 1360-1364.

6. 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. The Annals of thoracic surgery 2013; 96: 1761-1768.
7. 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. *Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer* 2012; 7: 1246-1251.

8. 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. *Journal of Thoracic Oncology* 2013; 8: 1170-1180.

9. Cho S, Song IH, Yang HC, Kim K, Jheon S. Predictive factors for node metastasis in patients with clinical stage I non-small cell lung cancer. *Ann Thorac Surg* 2013; 96: 239-245.

10. Zang R-C, Qiu B, Gao S-G, He J. A model predicting lymph node status for patients with clinical stage T1aN0-2M0 non-small cell lung cancer. *Chinese medical journal* 2017; 130: 398-403.

11. Akthar AS, Ferguson MK, Koshy M, Vigneswaran WT, Malik R. Limitations of PET/CT in the detection of occult N1 metastasis in clinical stage I(T1-2aN0) non-small cell lung cancer for staging prior to stereotactic body radiotherapy. *Technology in Cancer Research & Treatment* 2017; 16: 15-21.

12. Vial MR, Khan KA, O’Connell O, Peng SA, Gomez DR, Chang JY, Rice DC, Mehran R, Jimenez CJ, Grosu HB, Ost DE, Eapen GA. Endobronchial ultrasound-guided transbronchial needle aspiration in the nodal staging of stereotactic ablative body radiotherapy patients. *The Annals of thoracic surgery* 2017; 103: 1600-1605.

13. 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. *Annals of internal medicine* 2003; 139: 879-892.

14. Casal RF, Vial MR, Miller R, Mudambi L, Grosu HB, Eapen GA, Jimenez CA, Morice RC, Cornwell L, Ost D. What Exactly Is a Centrally Located Lung Tumor? Results of an Online Survey. *Annals of the American Thoracic Society* 2017; 14: 118-123.

15. Casal RF, Sepesi B, Sagar A-ES, Tschirren J, Chen M, Li L, Sunny J, Williams J, Grosu HB, Eapen GA, Jimenez CA, Ost DE. Centrally-located lung cancer and risk of occult nodal disease: an objective evaluation of multiple definitions of tumor centrality with a dedicated imaging software. *European Respiratory Journal* 2019: 1802220.

16. 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. *Journal of thoracic disease* 2015; 7: 576-584.

17. 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. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* 2015; 47: 1022-1026.

18. 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.

19. 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: the official publication of the International Cancer Imaging Society

20. Takenaka T, Yano T, Morodomi Y, Ito K, Miura N, Kawano D, Shoji F, Baba S, Abe K, Honda H, Maehara Y. Prediction of true-negative lymph node metastasis in clinical IA non-small cell lung cancer by measuring standardized uptake values on positron emission tomography. Surgery today 2012; 42: 934-939.

21. 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-e250S.

22. Gu P, Zhao YZ, Jiang LY, Zhang W, Xin Y, Han BH. Endobronchial ultrasound-guided transbronchial needle aspiration for staging of lung cancer: a systematic review and meta-analysis. European journal of cancer (Oxford, England : 1990) 2009; 45: 1389-1396.

23. Sas Institute Inc. Ordinal Response Modeling with the LOGISTIC Procedure. 2013.

24. Using Statistical Regression Methods in Education Research. 2011 [cited 2019 January 24, 2019]. Available from: http://www.restore.ac.uk/srme/www/fac/soc/wie/research-new/srme/glossary/index52d5.html?selectedLetter=P#proportional-odds-assumption.

25. UCLA: Statistical Consulting Group. Ordered logistic regression - SAS annotated output. Available from: https://stats.idre.ucla.edu/sas/output/ordered-logistic-regression/.

26. Williams R. Generalized ordered logit/partial propotional odds model for ordinal dependent variables. The Stata Journal 2006; 6: 58-82.
27. 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.

28. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European heart journal* 2014; 35: 1925-1931.

29. Fagerland MW, Hosmer DW. How to test for goodness of fit in ordinal logistic regression models. *Stata Journal* 2017; 17: 668-686.

30. Brier GW. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 1950; 78: 1-3.

31. Austin PC, van Klaveren D, Vergouwe Y, Nieboer D, Lee DS, Steyerberg EW. Validation of prediction models: examining temporal and geographic stability of baseline risk and estimated covariate effects. *Diagn Progn Res* 2017; 1: 12-12.

32. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY: Springer; 2009.

33. Evison M, Morris J, Martin J, Shah R, Barber PV, Booton R, Crosbie PAJ. Nodal staging in lung cancer: a risk stratification model for lymph nodes classified as negative by EBUS-TBNA. *Journal of Thoracic Oncology* 2015; 10: 126-133.

34. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The eighth edition lung cancer stage classification. *Chest* 2017; 151: 193-203.

35. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WEE, Nicholson AG, Groome P, Mitchell A, Bolejack V, Goldstraw P, Rami-Porta R, Asamura H, Ball D, Beer DG, Beyruti R, Bolejack V, Chansky K, Crowley J, Detterbeck F, Erich Eberhardt WE, Edwards J, Galateau-Sallé F, Giroux D, Gleeson F, Groome P, Huang J, Kennedy C,
Kim J, Kim YT, Kingsbury L, Kondo H, Krasnik M, Kubota K, Lerut A, Lyons G, Marino M, Marom EM, van Meerbeeck J, Mitchell A, Nakano T, Nicholson AG, Nowak A, Peake M, Rice T, Rosenzweig K, Ruffini E, Rusch V, Saijo N, Van Schil P, Sculier J-P, Shemanski L, Stratton K, Suzuki K, Tachimori Y, Thomas CF, Travis W, Tsao MS, Turrisi A, Vansteenkiste J, Watanabe H, Wu Y-L, Baas P, Erasmus J, Hasegawa S, Inai K, Kernstine K, Kindler H, Krug L, Nackaerts K, Pass H, Rice D, Falkson C, Filosso PL, Giaccone G, Kondo K, Luchi M, Okumura M, Blackstone E, Abad Cavaco F, Ansótegui Barrera E, Abal Arca J, Parente Lamelas I, Arnaud Obrer A, Guijarro Jorge R, Ball D, Bascom GK, Blanco Orozco AI, González Castro MA, Blum MG, Chimondeguy D, Cvijanovic V, Defranchi S, de Olaiz Navarro B, Escobar Campuzano I, Macía Videirea I, Fernández Araujo E, Andreo García F, Fong KM, Francisco Corral G, Cerezo González S, Freixinet Gilart J, García Arangüena L, García Barajas S, Girard P, Goksel T, González Budiño MT, González Casaurrán G, Gullón Blanco JA, Hernández Hernández J, Hernández Rodríguez H, Herrero Collantes J, Iglesias Heras M, Izquierdo Elena JM, Jakobsen E, Kostas S, León Atance P, Núñez Ares A, Liao M, Losanovscky M, Lyons G, Magaroles R, De Esteban Júlvez L, Mariñán Gorospe M, McCaughan B, Kennedy C, Melchor Íñiguez R, Miravet Sorribes L, Naranjo Gozalo S, Álvarez de Arriba C, Núñez Delgado M, Padilla Alarcón J, Peñalver Cuesta JC, Park JS, Pass H, Pavón Fernández MJ, Rosenberg M, Ruffini E, Rusch V, Sánchez de Cos Escuín J, Saura Vinuesa A, Serra Mitjans M, Strand TE, Subotic D, Swisher S, Terra R, Thomas C, Tournoy K, Van Schil P, Velasquez M, Wu YL, Yokoi K. The IASLC Lung Cancer Staging Project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. Journal of Thoracic Oncology 2016; 11: 39-51.
36. Asamura H, Chansky K, Crowley J, Goldstraw P, Rusch VW, Vansteenkiste JF, Watanabe H, Wu Y-L, Zielinski M, Ball D, Rami-Porta R. The International Association for the Study of Lung Cancer Lung Cancer Staging Project: proposals for the revision of the N descriptors in the forthcoming 8th edition of the TNM classification for lung cancer. 

*Journal of Thoracic Oncology* 2015; 10: 1675-1684.
Figure Legends

Figure E1. Tumors located within the inner one-third of the hemi-thorax or with the center of the tumor within the segmental airways on CT were defined as a central location tumors (shown in red). Tumors that were located in the outer two-thirds of the hemithorax required that their center did not touch the segmental airways and that their center was located outside the central one-third to be considered as outer location tumors (shown in yellow).

Figure E2. Receiver operating characteristic curve for the prediction model in the combined external validation cohort for (A) N1|2|3 (vs. N0) disease (area under the curve=0.86) and (B) N2|3 (vs. N0|1) disease (area under the curve=0.88).

Figure E3. The predicted probability of N0, N1 or N2|3 disease for central vs. peripheral tumor locations as a function of age according to PET N stage (A) N stage by PET=N0, (B) N stage by PET=N1 and C) N stage by PET=N2 or N3.
Table E1. Univariate Ordinal Logistic Regression to for Prediction of N0 vs. N1 vs. N2|3

### Disease

|                          | Odds ratio of having N1|2|3 (vs. N0) disease | 95% CI | p-value | Odds ratio of having N2|3 (vs. N0|1) disease | 95% CI | p-value |
|-------------------------|------------------------|-------------------|--------|---------|------------------------|----------------|--------|---------|
| Age (years) ‡           | 0.97 0.95 0.98         | <0.001            | 0.965 0.95 0.98 | <0.001 |
| Gender                  |                        |                   |        |         |                        |                   |        |         |
| Female                  | 1.00                   | 1.00              |        |         |                        |                   |        |         |
| Male                    | 0.86 0.63 1.19         | 0.373             | 0.86 0.63 1.19 | 0.373  |
| Race                    |                        |                   |        |         |                        |                   |        |         |
| Asian                   | 0.54 0.17 1.69         | 0.316             | 0.54 0.17 1.69 | 0.316  |
| Black                   | 1.04 0.57 1.89         | 0.473             | 1.04 0.57 1.89 | 0.473  |
| Hispanic                | 0.92 0.42 1.97         | 0.808             | 0.92 0.42 1.97 | 0.808  |
| White                   | 1.00                   | 1.00              |        |         |                        |                   |        |         |
| ASA Score               |                        |                   |        |         |                        |                   |        |         |
| 1                       | 1.00                   | 1.00              |        |         |                        |                   |        |         |
| 2                       | 0.47 0.09 2.41         | 0.368             | 0.47 0.09 2.41 | 0.368  |
| 3                       | 0.42 0.09 1.96         | 0.271             | 0.42 0.09 1.96 | 0.271  |
| 4                       | 0.14 0.01 1.83         | 0.134             | 0.14 0.01 1.83 | 0.134  |
| Smoking Status          |                        |                   |        |         |                        |                   |        |         |
| Never used cigarettes   | 0.94 0.50 1.78         | 0.851             | 0.94 0.50 1.78 | 0.851  |
| Prior use of cigarettes | 1.07 0.72 1.60         | 0.739             | 1.07 0.72 1.60 | 0.739  |
| Current use of cigarettes | 1.00                 | 1.00              |        |         |                        |                   |        |         |
| ECOG                    |                        |                   |        |         |                        |                   |        |         |
| 0                       | 1.00                   | 1.00              |        |         |                        |                   |        |         |
| 1                       | 1.15 0.79 1.68         | 0.465             | 1.15 0.79 1.68 | 0.465  |
| 2                       | 0.80 0.48 1.34         | 0.400             | 0.80 0.48 1.34 | 0.400  |
| 3                       | 0.41 0.11 1.46         | 0.168             | 0.41 0.11 1.46 | 0.168  |
| Size of the tumor       |                        |                   |        |         |                        |                   |        |         |
| <= 3 cm                 | 1.00                   | 1.00              |        |         |                        |                   |        |         |
| >3 cm but <= 5 cm       | 1.30 0.89 1.88         | 0.172             | 1.30 0.89 1.88 | 0.172  |
| >5 cm                   | 1.24 0.82 1.87         | 0.304             | 1.24 0.82 1.87 | 0.304  |
| Lobar location of the tumor |                    |                   |        |         |                        |                   |        |         |
| Left upper lobe or lingula | 1.00                 | 1.00              |        |         |                        |                   |        |         |
| Left lower lobe         | 0.91 0.54 1.56         | 0.736             | 0.91 0.54 1.56 | 0.736  |
| Right upper lobe        | 1.14 0.75 1.72         | 0.544             | 1.14 0.75 1.72 | 0.544  |
| Right lower or middle lobe | 1.58                 | 1.01 2.46         | 0.044 1.58 1.01 2.46 | 0.044  |
| Location ‡             |                        |                   |        |         |                        |                   |        |         |
| Outer 2/3rd of lung     | 1.00                   | *                 | 1.00 *  |        |                        |                   |        |         |
| Central 1/3rd of lung   | 1.57 1.10 2.26         | 0.014             | 1.57 1.10 2.26 | 0.014  |
| Histology ‡            |                        |                   |        |         |                        |                   |        |         |
| Adenocarcinoma          | 1.00                   | *                 | 1.00 *  |        |                        |                   |        |         |
| Squamous cell carcinoma | 0.58 0.40 0.84         | 0.004             | 0.58 0.40 0.84 | 0.004  |
| Non-small cell carcinoma | 1.21                 | 0.70 2.10         | 0.489 1.21 0.70 2.10 | 0.489  |
| Other primary lung cancer | 0.64                 | 0.27 1.49         | 0.298 0.64 0.27 1.49 | 0.298  |
| CT Characteristics      |                        |                   |        |         |                        |                   |        |         |
| Cavitary               | 1.00                   | I                 | 1.00 I  |        |                        |                   |        |         |
| Ground glass, semi-solid, infiltrate | 0.30 0.08 1.16 | 0.080 0.30 0.08 1.16 | 0.080  |
| Solid                  | 1.31 0.53 3.26         | 0.562             | 1.31 0.53 3.26 | 0.562  |
| Satellite lesion in same lobe |             |                   |        |         |                        |                   |        |         |
| No                     | 1.00                   | I                 | 1.00 I  |        |                        |                   |        |         |
| Yes                    | 0.37 0.13 1.08         | 0.068             | 0.37 0.13 1.08 | 0.068  |
| PET-CT § |  |  | * | 1.00 |  |  | * |
| CT=N0 or N1, PET=N0 | 1.00 |  | * | 1.00 |  | 6.79 | 0.104 |
| CT=N2 or N3, PET=N0 | 2.89 | 1.17 | 7.13 | 0.022 | 2.38 | 0.84 | * |
| CT=N0 or N1, PET=N1 | 20.50 | 9.19 | 45.74 | <0.001 | 4.87 | 1.86 | 12.73 | 0.001 |
| CT=N2 or N3, PET=N1 | 20.00 | 7.98 | 50.11 | <0.001 | 2.82 | 0.78 | 10.17 | 0.112 |
| CT=N0 or N1, PET=N2 or N3 | 8.64 | 3.68 | 20.28 | <0.001 | 9.67 | 3.81 | 24.55 | <0.001 |
| CT=N2 or N3, PET=N2 or N3 | 37.07 | 17.72 | 77.58 | <0.001 | 38.84 | 17.23 | 87.55 | <0.001 |

ASA=American Society of Anesthesiologists; ECOG= Eastern Cooperative Oncology Group; CT= computed tomography; PET= positron emission tomography; CI=Confidence interval; N1|2|3= N stage equal or greater than 1; N2|3= N2 or N3 disease; N1|2=N0 or N1 disease

* Overall p-value is <0.2 and the variable was collected for MD Anderson Cancer Center data; † Overall p-value is <0.2 but the variable was not collected for data outside MD Anderson Cancer Center.

‡ Variables had a p-value <0.2 and did not violate the proportional odds assumption after Score test. Therefore, the odds ratio N1|2|3 (vs. N0) disease are the same to the odds ratio of having N2|3 (vs. N0|1) disease.
§ Variables had a p-value of less than 0.2 and violated the proportional odds assumption after Score test. Therefore, the odds ratio of N1|2|3 (vs. N0) disease are different to the odds ratio of N2|3 (vs. N0|1) disease. Both odds ratios are shown.
Table E2. Descriptive Statistics by N Stage for the Temporal Validation Cohort (N=473)

|                  | N missing | N0 (N=347) | N1 (N=34) | N2|3 (N=92) | p-value* |
|------------------|-----------|------------|-----------|---|---------|----------|
| Age (years), mean ±SD | 0         | 70.05 ± 9.34 | 68 ± 10.20 | 67.51 ± 10.04 | 0.051Ɨ |
| Size of the tumor, n (%) | 0         | 203(79.9) | 12(4.7) | 39(15.4) | 0.007 |
| <= 3 cm     |           |           |           |           |         |
| >3 cm but <= 5 cm |           | 84(64.1)  | 12(9.2)  | 35(26.7) |         |
| >5 cm       |           | 60(68.2)  | 10(11.4) | 18(20.5) |         |
| Location, n (%) | 0         | 262(76.8) | 25(7.3) | 54(15.8) | 0.006 |
| Outer 2/3rd of lung |           |           |           |           |         |
| Central 1/3rd of lung |           | 85(64.4)  | 9(6.8)  | 38(28.8) |         |
| Histology, n (%) | 0         | 203(73.6) | 15(5.4) | 58(21.0) | 0.043 |
| Adenocarcinoma      |           |           |           |           |         |
| Squamous cell carcinoma |         | 108(76.1) | 13(9.2) | 21(16.8) |         |
| Non-small cell carcinoma |         | 28(71.8)  | 2(5.1)  | 9(23.1)  |         |
| Other primary lung cancer |         | 8(50.0)   | 4(25.0) | 4(25.0)  |         |
| N stage by PET-CT, n (%) | 0         | 221(96.1) | 4(1.7)  | 5(2.2)   | <0.001 |
| CT=N0 or N1, PET=N0 |           |           |           |           |         |
| CT=N2 or N3, PET=N0 |           | 27(93.1)  | 0(0.0)  | 2(6.9)   |         |
| CT=N0 or N1, PET=N1 |           | 30(61.2)  | 14(28.6) | 5(10.2)  |         |
| CT=N2 or N3, PET=N1 |           | 8(57.1)   | 4(28.6) | 2(14.3)  |         |
| CT=N0 or N1, PET=N2 or N3 |       | 5(38.5)   | 1(7.7)  | 7(53.9)  |         |
| CT=N2 or N3, PET=N2 or N3 |       | 56(40.6)  | 11(8.0%)| 71(51.5) |         |

SD= standard deviation; CT= computed tomography; PET= positron emission tomography
N stage as assessed by endobronchial ultrasound-guided transbronchial needle aspiration; *N2 and N3 are combined (N2|3).
P values are for* Chi-square test except where otherwise noted; Ɨ Anova test
Table E3. Descriptive Statistics by N Stage for the External Validation Cohort (N=722)

|                                | N missing | N0 (N=353)        | N1 (N=73)        | N2|3 (N=296) | p-value* |
|--------------------------------|-----------|-------------------|-----------------|---------|-----------|----------|
| Age (years)                    | 0         | 69.97 ± 9.57      | 68.11 ±10.7     | 66.11 ± 10.76 | <0.001    |
| Size of the tumor, n (%)       | 0         |                   |                 |          |           |          |
| <= 3 cm                        | 173(53.)  | 26(8.0)           | 127(39.0)       | 0.232   |           |          |
| >3 cm but <= 5 cm              | 104(45.4) | 26(11.4)          | 99(43.2)        |          |           |          |
| >5 cm                          | 76(45.5)  | 21(12.6)          | 70(41.9)        |          |           |          |
| Location, n (%)                | 0         |                   |                 |          |           |          |
| Outer 2/3rd of lung            | 277(54.5) | 40(7.9)           | 191(37.6)       | <0.001  |           |          |
| Central 1/3rd of lung           | 76(35.5)  | 33(15.4)          | 105(49.1)       |          |           |          |
| Histology, n (%)               | 0         |                   |                 |          |           |          |
| Adenocarcinoma                 | 179(47.0) | 28(7.3)           | 174(45.7)       | 0.001   |           |          |
| Squamous cell carcinoma        | 142(53.6) | 35(13.2)          | 88(33.2)        |          |           |          |
| Non-small cell carcinoma       | 11(27.5)  | 3(7.5)            | 26(65)          |          |           |          |
| Other primary lung cancer      | 21(58.3)  | 7(19.4)           | 8(22.2)         |          |           |          |
| N stage by PET-CT, n (%)       | 0         |                   |                 |          |           |          |
| CT=N0 or N1, PET=N0            | 153(93.3) | 3(1.8)            | 8(4.9)          | <0.001  |           |          |
| CT=N2 or N3, PET=N0            | 79(85.9)  | 4(4.3)            | 9(9.8)          |          |           |          |
| CT=N0 or N1, PET=N1            | 27(37.5)  | 34(47.2)          | 11(15.3)        |          |           |          |
| CT=N2 or N3, PET=N1            | 11(30.6)  | 16(44.4)          | 9(25)           |          |           |          |
| CT=N0 or N1, PET=N2 or N3      | 33(50.8)  | 3(4.6)            | 29(44.6)        |          |           |          |
| CT=N2 or N3, PET=N2 or N3      | 50(17.1)  | 13(4.4)           | 230(78.5)       |          |           |          |

SD= standard deviation; CT= computed tomography; PET= positron emission tomography
N stage as assessed by endobronchial ultrasound-guided transbrachial needle aspiration; *N2 and N3 are combined (N2|3).

* Chi-square test except where otherwise noted; Ɨ Anova test
Table E4. Intercepts for Calibration for Each Institution of the External Validation Cohort

| Calibration intercepts for Johns Hopkins | Calibration intercepts for Cleveland Clinic Foundation | Calibration intercepts for Henry Ford Hospital |
|-----------------------------------------|-------------------------------------------------------|-----------------------------------------------|
| N1|2|3 (vs. N0) disease                      | N1|2|3 (vs. N0) disease                          | N1|2|3 (vs. N0) disease                            |
| N2|3 (vs. N0|1) disease                            | Brier score                                 | AUC (95% CI)                                 | AUC (95% CI)                                 | AUC (95% CI)                                 |
| ---------------------------------------|-------------------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| -0.00                                  | +0.00                                                 | 0.456                                         | <0.001                                       | 0.812 (0.740-0.884)                            | 0.823 (0.758-0.887)                           |
| +0.90                                   | +0.90                                                 | 0.394                                         | 0.041                                        | 0.812 (0.740-0.884)                            | 0.823 (0.758-0.887)                           |
| +1.00                                   | +1.00                                                 | 0.393                                         | 0.037                                        | 0.812 (0.740-0.884)                            | 0.823 (0.758-0.887)                           |
| +1.10                                   | +1.10                                                 | 0.394                                         | 0.023                                        | 0.812 (0.740-0.884)                            | 0.823 (0.758-0.887)                           |
| +1.10                                   | +0.90                                                 | 0.392                                         | 0.259                                        | 0.812 (0.740-0.884)                            | 0.823 (0.758-0.887)                           |
| +0.00                                  | +0.00                                                 | 0.337                                         | 0.287                                        | 0.837 (0.792-0.882)                            | 0.867 (0.825-0.909)                           |
| +0.18                                   | +0.18                                                 | 0.334                                         | 0.551                                        | 0.837 (0.792-0.882)                            | 0.867 (0.825-0.909)                           |
| +0.22                                   | +0.22                                                 | 0.334                                         | 0.586                                        | 0.837 (0.792-0.882)                            | 0.867 (0.825-0.909)                           |
| +0.26                                   | +0.26                                                 | 0.334                                         | 0.608                                        | 0.837 (0.792-0.882)                            | 0.867 (0.825-0.909)                           |
| +0.18                                   | +0.26                                                 | 0.333                                         | 0.404                                        | 0.837 (0.792-0.882)                            | 0.867 (0.825-0.909)                           |
| +0.00                                  | +0.00                                                 | 0.361                                         | <0.001                                       | 0.908 (0.866-0.949)                            | 0.914 (0.875-0.954)                           |
| +0.90                                   | +0.90                                                 | 0.286                                         | 0.080                                        | 0.908 (0.866-0.949)                            | 0.914 (0.875-0.954)                           |
| +0.98                                   | +0.98                                                 | 0.286                                         | 0.082                                        | 0.908 (0.866-0.949)                            | 0.914 (0.875-0.954)                           |
| +1.06                                   | +1.06                                                 | 0.286                                         | 0.036                                        | 0.908 (0.866-0.949)                            | 0.914 (0.875-0.954)                           |
| +1.06                                   | +0.90                                                 | 0.285                                         | 0.196                                        | 0.908 (0.866-0.949)                            | 0.914 (0.875-0.954)                           |

AUC= Area under the Receiver Operator Characteristics curve; CI= confidence interval; N1|2|3= N stage equal or greater than 1; N2|3= N2 or N3 disease; N1|2= N0 or N1 disease
For each given pair of intercepts, Brier score, Hosmer-Lemeshow p-value and AUC are shown. The pair of intercepts with the lowest Brier score was selected (shown in bold). There is no change in AUC for a given change in intercept.
Table E5. Predicted Probabilities of N0, N1 and N2|3 Disease for a Given Change in the Explanatory Variables in the Development Cohort

| PET-CT            | Histology       | Location       | Age | Predicted probability of N0 | Predicted Probability of N1 | Predicted Probability of N2|3 |
|-------------------|-----------------|----------------|-----|-----------------------------|----------------------------|-----------------------------|
| CT=N0 or N1, PET=N0 | Adenocarcinoma | Central 1/3rd  | 60  | 0.896                       | 0.022                      | 0.081                       |
| CT=N0 or N1, PET=N0 | Adenocarcinoma | Outer 2/3rd    | 60  | 0.934                       | 0.015                      | 0.052                       |
| CT=N0 or N1, PET=N0 | Squamous cell   | Central 1/3rd  | 60  | 0.952                       | 0.011                      | 0.037                       |
| CT=N0 or N1, PET=N0 | Squamous cell   | Outer 2/3rd    | 60  | 0.970                       | 0.007                      | 0.023                       |
| CT=N0 or N1, PET=N0 | Non-small cell  | Central 1/3rd  | 60  | 0.890                       | 0.024                      | 0.086                       |
| CT=N0 or N1, PET=N0 | Non-small cell  | Outer 2/3rd    | 60  | 0.930                       | 0.016                      | 0.055                       |
| CT=N0 or N1, PET=N0 | Other lung primary | Central 1/3rd | 60  | 0.929                       | 0.016                      | 0.055                       |
| CT=N0 or N1, PET=N0 | Other lung primary | Outer 2/3rd  | 60  | 0.955                       | 0.010                      | 0.035                       |
| CT=N0 or N1, PET=N0 | Adenocarcinoma  | Central 1/3rd  | 70  | 0.921                       | 0.017                      | 0.062                       |
| CT=N0 or N1, PET=N0 | Adenocarcinoma  | Outer 2/3rd    | 70  | 0.950                       | 0.011                      | 0.039                       |
| CT=N0 or N1, PET=N0 | Squamous cell   | Central 1/3rd  | 70  | 0.963                       | 0.008                      | 0.028                       |
| CT=N0 or N1, PET=N0 | Squamous cell   | Outer 2/3rd    | 70  | 0.977                       | 0.005                      | 0.018                       |
| CT=N0 or N1, PET=N0 | Non-small cell  | Central 1/3rd  | 70  | 0.916                       | 0.018                      | 0.066                       |
| CT=N0 or N1, PET=N0 | Non-small cell  | Outer 2/3rd    | 70  | 0.947                       | 0.012                      | 0.041                       |
| CT=N0 or N1, PET=N0 | Other lung primary | Central 1/3rd| 70  | 0.946                       | 0.012                      | 0.042                       |
| CT=N0 or N1, PET=N0 | Other lung primary | Outer 2/3rd  | 70  | 0.966                       | 0.008                      | 0.026                       |
| CT=N0 or N1, PET=N0 | Adenocarcinoma  | Central 1/3rd  | 80  | 0.940                       | 0.014                      | 0.047                       |
| CT=N0 or N1, PET=N0 | Adenocarcinoma  | Outer 2/3rd    | 80  | 0.962                       | 0.009                      | 0.029                       |
| CT=N0 or N1, PET=N0 | Squamous cell   | Central 1/3rd  | 80  | 0.972                       | 0.006                      | 0.021                       |
| CT=N0 or N1, PET=N0 | Squamous cell   | Outer 2/3rd    | 80  | 0.983                       | 0.004                      | 0.013                       |
| CT=N0 or N1, PET=N0 | Non-small cell  | Central 1/3rd  | 80  | 0.936                       | 0.014                      | 0.050                       |
| CT=N0 or N1, PET=N0 | Non-small cell  | Outer 2/3rd    | 80  | 0.960                       | 0.009                      | 0.031                       |
| CT=N0 or N1, PET=N0 | Other lung primary | Central 1/3rd| 80  | 0.959                       | 0.009                      | 0.032                       |
| CT=N0 or N1, PET=N0 | Other lung primary | Outer 2/3rd  | 80  | 0.974                       | 0.006                      | 0.020                       |
| CT=N0 or N1, PET=N1 | Adenocarcinoma  | Central 1/3rd  | 60  | 0.284                       | 0.413                      | 0.303                       |
| CT=N0 or N1, PET=N1 | Adenocarcinoma  | Outer 2/3rd    | 60  | 0.392                       | 0.397                      | 0.211                       |
| CT=N0 or N1, PET=N1 | Squamous cell   | Central 1/3rd  | 60  | 0.474                       | 0.365                      | 0.161                       |
| CT=N0 or N1, PET=N1 | Squamous cell   | Outer 2/3rd    | 60  | 0.595                       | 0.300                      | 0.105                       |
| CT=N0 or N1, PET=N1 | Non-small cell  | Central 1/3rd  | 60  | 0.271                       | 0.412                      | 0.317                       |
| CT=N0 or N1, PET=N1 | Non-small cell  | Outer 2/3rd    | 60  | 0.377                       | 0.401                      | 0.222                       |
| CT=N0 or N1, PET=N1 | Other lung primary | Central 1/3rd| 60  | 0.374                       | 0.402                      | 0.224                       |
| CT=N0 or N1, PET=N1 | Other lung primary | Outer 2/3rd  | 60  | 0.493                       | 0.356                      | 0.151                       |
| CT=N0 or N1, PET=N1 | Adenocarcinoma  | Central 1/3rd  | 70  | 0.347                       | 0.408                      | 0.245                       |
| CT=N0 or N1, PET=N1 | Adenocarcinoma  | Outer 2/3rd    | 70  | 0.464                       | 0.370                      | 0.166                       |
| CT=N0 or N1, PET=N1 | Squamous cell   | Central 1/3rd  | 70  | 0.547                       | 0.328                      | 0.125                       |
| CT=N0 or N1, PET=N1 | Squamous cell   | Outer 2/3rd    | 70  | 0.663                       | 0.257                      | 0.081                       |
| CT=N0 or N1, PET=N1 | Non-small cell  | Central 1/3rd  | 70  | 0.333                       | 0.410                      | 0.257                       |
| CT=N0 or N1, PET=N1 | Non-small cell  | Outer 2/3rd    | 70  | 0.448                       | 0.377                      | 0.175                       |
| CT=N0 or N1, PET=N1 | Other lung primary | Central 1/3rd | 70 | 0.445 | 0.378 | 0.177 |
| CT=N0 or N1, PET=N1 | Other lung primary | Outer 2/3rd | 70 | 0.566 | 0.317 | 0.117 |
| CT=N0 or N1, PET=N1 | Adenocarcinoma | Central 1/3rd | 80 | 0.416 | 0.389 | 0.195 |
| CT=N0 or N1, PET=N1 | Adenocarcinoma | Outer 2/3rd | 80 | 0.537 | 0.334 | 0.130 |
| CT=N0 or N1, PET=N1 | Squamous cell | Central 1/3rd | 80 | 0.618 | 0.286 | 0.096 |
| CT=N0 or N1, PET=N1 | Squamous cell | Outer 2/3rd | 80 | 0.725 | 0.214 | 0.062 |
| CT=N0 or N1, PET=N1 | Non-small cell | Central 1/3rd | 80 | 0.401 | 0.394 | 0.205 |
| CT=N0 or N1, PET=N1 | Non-small cell | Outer 2/3rd | 80 | 0.521 | 0.342 | 0.137 |
| CT=N0 or N1, PET=N1 | Other lung primary | Central 1/3rd | 80 | 0.517 | 0.344 | 0.139 |
| CT=N0 or N1, PET=N1 | Other lung primary | Outer 2/3rd | 80 | 0.636 | 0.274 | 0.090 |
| CT=N0 or N1, PET=N2 or N3 | Adenocarcinoma | Central 1/3rd | 60 | 0.475 | 0.042 | 0.483 |
| CT=N0 or N1, PET=N2 or N3 | Adenocarcinoma | Outer 2/3rd | 60 | 0.595 | 0.040 | 0.365 |
| CT=N0 or N1, PET=N2 or N3 | Squamous cell | Central 1/3rd | 60 | 0.673 | 0.036 | 0.292 |
| CT=N0 or N1, PET=N2 or N3 | Squamous cell | Outer 2/3rd | 60 | 0.770 | 0.028 | 0.202 |
| CT=N0 or N1, PET=N2 or N3 | Non-small cell | Central 1/3rd | 60 | 0.459 | 0.042 | 0.499 |
| CT=N0 or N1, PET=N2 or N3 | Non-small cell | Outer 2/3rd | 60 | 0.580 | 0.040 | 0.380 |
| CT=N0 or N1, PET=N2 or N3 | Other lung primary | Central 1/3rd | 60 | 0.577 | 0.040 | 0.383 |
| CT=N0 or N1, PET=N2 or N3 | Other lung primary | Outer 2/3rd | 60 | 0.689 | 0.035 | 0.276 |
| CT=N0 or N1, PET=N2 or N3 | Adenocarcinoma | Central 1/3rd | 70 | 0.548 | 0.041 | 0.411 |
| CT=N0 or N1, PET=N2 or N3 | Adenocarcinoma | Outer 2/3rd | 70 | 0.663 | 0.036 | 0.301 |
| CT=N0 or N1, PET=N2 or N3 | Squamous cell | Central 1/3rd | 70 | 0.734 | 0.031 | 0.235 |
| CT=N0 or N1, PET=N2 or N3 | Squamous cell | Outer 2/3rd | 70 | 0.817 | 0.024 | 0.159 |
| CT=N0 or N1, PET=N2 or N3 | Non-small cell | Central 1/3rd | 70 | 0.532 | 0.041 | 0.427 |
| CT=N0 or N1, PET=N2 or N3 | Non-small cell | Outer 2/3rd | 70 | 0.649 | 0.037 | 0.314 |
| CT=N0 or N1, PET=N2 or N3 | Other lung primary | Central 1/3rd | 70 | 0.646 | 0.037 | 0.317 |
| CT=N0 or N1, PET=N2 or N3 | Other lung primary | Outer 2/3rd | 70 | 0.748 | 0.030 | 0.222 |
| CT=N0 or N1, PET=N2 or N3 | Adenocarcinoma | Central 1/3rd | 80 | 0.619 | 0.039 | 0.343 |
| CT=N0 or N1, PET=N2 or N3 | Adenocarcinoma | Outer 2/3rd | 80 | 0.725 | 0.032 | 0.243 |
| CT=N0 or N1, PET=N2 or N3 | Squamous cell | Central 1/3rd | 80 | 0.787 | 0.027 | 0.187 |
| CT=N0 or N1, PET=N2 or N3 | Squamous cell | Outer 2/3rd | 80 | 0.857 | 0.019 | 0.124 |
| CT=N0 or N1, PET=N2 or N3 | Non-small cell | Central 1/3rd | 80 | 0.604 | 0.039 | 0.357 |
| CT=N0 or N1, PET=N2 or N3 | Non-small cell | Outer 2/3rd | 80 | 0.712 | 0.033 | 0.255 |
| CT=N0 or N1, PET=N2 or N3 | Other lung primary | Central 1/3rd | 80 | 0.710 | 0.033 | 0.257 |
| CT=N0 or N1, PET=N2 or N3 | Other lung primary | Outer 2/3rd | 80 | 0.799 | 0.026 | 0.176 |
| CT=N2 or N3, PET=N0 | Adenocarcinoma | Central 1/3rd | 60 | 0.728 | 0.081 | 0.191 |
| CT=N2 or N3, PET=N0 | Adenocarcinoma | Outer 2/3rd | 60 | 0.813 | 0.060 | 0.126 |
| CT=N2 or N3, PET=N0 | Squamous cell | Central 1/3rd | 60 | 0.859 | 0.047 | 0.094 |
| CT=N2 or N3, PET=N0 | Squamous cell | Outer 2/3rd | 60 | 0.908 | 0.032 | 0.060 |
| CT=N2 or N3, PET=N0 | Non-small cell | Central 1/3rd | 60 | 0.715 | 0.084 | 0.201 |
| CT=N2 or N3, PET=N0 | Non-small cell | Outer 2/3rd | 60 | 0.803 | 0.063 | 0.134 |
| CT=N2 or N3, PET=N0 | Other lung primary | Central 1/3rd | 60 | 0.801 | 0.064 | 0.135 |
| CT=N2 or N3, PET=N0 | Other lung primary | Outer 2/3rd | 60 | 0.868 | 0.045 | 0.088 |
| CT=N2 or N3, PET=N0 | Adenocarcinoma | Central 1/3rd | 70 | 0.782 | 0.069 | 0.149 |
| CT=N2 or N3, PET=N0 | Tumor Type          | Location     | Percent | z-score | p-value |
|---------------------|---------------------|--------------|---------|---------|---------|
| Adenocarcinoma      | Outer 2/3rd         | 70           | 0.854   | 0.049   | 0.098   |
| Squamous cell       | Central 1/3rd       | 70           | 0.891   | 0.037   | 0.072   |
| Squamous cell       | Outer 2/3rd         | 70           | 0.930   | 0.025   | 0.045   |
| Non-small cell      | Central 1/3rd       | 70           | 0.771   | 0.071   | 0.158   |
| Non-small cell      | Outer 2/3rd         | 70           | 0.846   | 0.051   | 0.103   |
| Other lung primary  | Central 1/3rd       | 70           | 0.844   | 0.052   | 0.104   |
| Other lung primary  | Outer 2/3rd         | 70           | 0.898   | 0.035   | 0.067   |
| Adenocarcinoma      | Central 1/3rd       | 80           | 0.828   | 0.056   | 0.116   |
| Adenocarcinoma      | Outer 2/3rd         | 80           | 0.887   | 0.039   | 0.075   |
| Squamous cell       | Central 1/3rd       | 80           | 0.916   | 0.029   | 0.055   |
| Squamous cell       | Outer 2/3rd         | 80           | 0.947   | 0.019   | 0.034   |
| Non-small cell      | Central 1/3rd       | 80           | 0.818   | 0.059   | 0.123   |
| Non-small cell      | Outer 2/3rd         | 80           | 0.880   | 0.041   | 0.079   |
| Other lung primary  | Central 1/3rd       | 80           | 0.879   | 0.041   | 0.080   |
| Other lung primary  | Outer 2/3rd         | 80           | 0.922   | 0.027   | 0.051   |
| Adenocarcinoma      | Central 1/3rd       | 60           | 0.303   | 0.513   | 0.183   |
| Adenocarcinoma      | Outer 2/3rd         | 60           | 0.414   | 0.464   | 0.121   |
| Squamous cell       | Central 1/3rd       | 60           | 0.497   | 0.413   | 0.090   |
| Squamous cell       | Outer 2/3rd         | 60           | 0.617   | 0.326   | 0.057   |
| Non-small cell      | Central 1/3rd       | 60           | 0.290   | 0.517   | 0.193   |
| Non-small cell      | Outer 2/3rd         | 60           | 0.399   | 0.473   | 0.128   |
| Other lung primary  | Central 1/3rd       | 60           | 0.396   | 0.474   | 0.130   |
| Other lung primary  | Outer 2/3rd         | 60           | 0.516   | 0.400   | 0.084   |
| Adenocarcinoma      | Central 1/3rd       | 70           | 0.368   | 0.488   | 0.144   |
| Adenocarcinoma      | Outer 2/3rd         | 70           | 0.487   | 0.420   | 0.093   |
| Squamous cell       | Central 1/3rd       | 70           | 0.570   | 0.361   | 0.069   |
| Squamous cell       | Outer 2/3rd         | 70           | 0.683   | 0.273   | 0.043   |
| Non-small cell      | Central 1/3rd       | 70           | 0.354   | 0.495   | 0.152   |
| Non-small cell      | Outer 2/3rd         | 70           | 0.471   | 0.430   | 0.099   |
| Other lung primary  | Central 1/3rd       | 70           | 0.468   | 0.432   | 0.100   |
| Other lung primary  | Outer 2/3rd         | 70           | 0.588   | 0.348   | 0.064   |
| Adenocarcinoma      | Central 1/3rd       | 80           | 0.439   | 0.450   | 0.111   |
| Adenocarcinoma      | Outer 2/3rd         | 80           | 0.560   | 0.369   | 0.071   |
| Squamous cell       | Central 1/3rd       | 80           | 0.640   | 0.308   | 0.052   |
| Squamous cell       | Outer 2/3rd         | 80           | 0.743   | 0.224   | 0.033   |
| Non-small cell      | Central 1/3rd       | 80           | 0.423   | 0.459   | 0.118   |
| Non-small cell      | Outer 2/3rd         | 80           | 0.544   | 0.380   | 0.076   |
| Other lung primary  | Central 1/3rd       | 80           | 0.541   | 0.383   | 0.077   |
| Other lung primary  | Outer 2/3rd         | 80           | 0.657   | 0.295   | 0.049   |
| Adenocarcinoma      | Central 1/3rd       | 60           | 0.175   | 0.036   | 0.790   |
| Adenocarcinoma      | Outer 2/3rd         | 60           | 0.256   | 0.046   | 0.698   |
| Squamous cell       | Central 1/3rd       | 60           | 0.325   | 0.052   | 0.623   |
| Squamous cell       | Outer 2/3rd         | 60           | 0.439   | 0.057   | 0.504   |
| CT=N2 or N3, PET=N2 or N3 | Tumor Type | Location | N | p-value | q-value | Statistic |
|---------------------------|------------|----------|---|---------|---------|-----------|
| CT=N2 or N3, PET=N2 or N3 | Non-small cell | Central 1/3rd | 60 | 0.166 | 0.034 | 0.800 |
| CT=N2 or N3, PET=N2 or N3 | Non-small cell | Outer 2/3rd | 60 | 0.244 | 0.045 | 0.711 |
| CT=N2 or N3, PET=N2 or N3 | Other lung primary | Central 1/3rd | 60 | 0.242 | 0.045 | 0.714 |
| CT=N2 or N3, PET=N2 or N3 | Other lung primary | Outer 2/3rd | 60 | 0.341 | 0.054 | 0.605 |
| CT=N2 or N3, PET=N2 or N3 | Adenocarcinoma | Central 1/3rd | 70 | 0.221 | 0.042 | 0.737 |
| CT=N2 or N3, PET=N2 or N3 | Adenocarcinoma | Outer 2/3rd | 70 | 0.316 | 0.052 | 0.633 |
| CT=N2 or N3, PET=N2 or N3 | Squamous cell | Central 1/3rd | 70 | 0.392 | 0.056 | 0.552 |
| CT=N2 or N3, PET=N2 or N3 | Squamous cell | Outer 2/3rd | 70 | 0.512 | 0.057 | 0.431 |
| CT=N2 or N3, PET=N2 or N3 | Non-small cell | Central 1/3rd | 70 | 0.210 | 0.041 | 0.749 |
| CT=N2 or N3, PET=N2 or N3 | Non-small cell | Outer 2/3rd | 70 | 0.302 | 0.051 | 0.647 |
| CT=N2 or N3, PET=N2 or N3 | Other lung primary | Central 1/3rd | 70 | 0.299 | 0.050 | 0.650 |
| CT=N2 or N3, PET=N2 or N3 | Other lung primary | Outer 2/3rd | 70 | 0.410 | 0.057 | 0.533 |
| CT=N2 or N3, PET=N2 or N3 | Adenocarcinoma | Central 1/3rd | 80 | 0.275 | 0.048 | 0.676 |
| CT=N2 or N3, PET=N2 or N3 | Adenocarcinoma | Outer 2/3rd | 80 | 0.382 | 0.056 | 0.563 |
| CT=N2 or N3, PET=N2 or N3 | Squamous cell | Central 1/3rd | 80 | 0.463 | 0.058 | 0.479 |
| CT=N2 or N3, PET=N2 or N3 | Squamous cell | Outer 2/3rd | 80 | 0.584 | 0.055 | 0.361 |
| CT=N2 or N3, PET=N2 or N3 | Non-small cell | Central 1/3rd | 80 | 0.263 | 0.047 | 0.690 |
| CT=N2 or N3, PET=N2 or N3 | Non-small cell | Outer 2/3rd | 80 | 0.367 | 0.055 | 0.578 |
| CT=N2 or N3, PET=N2 or N3 | Other lung primary | Central 1/3rd | 80 | 0.364 | 0.055 | 0.581 |
| CT=N2 or N3, PET=N2 or N3 | Other lung primary | Outer 2/3rd | 80 | 0.482 | 0.057 | 0.460 |

CT= computed tomography; PET= positron emission tomography; N2|3= N2 or N3 disease
Figure E1. Tumors located within the inner one-third of the hemi-thorax or with the center of the tumor within the segmental airways on CT were defined as a central location tumors (shown in red). Tumors that were located in the outer two-thirds of the hemithorax required that their center did not touch the segmental airways and that their center was located outside the central one-third to be considered as outer location tumors (shown in yellow).
Figure E2. Receiver operating characteristic curve for the prediction model in the combined external validation cohort for (A) N1|2|3 (vs. N0) disease (area under the curve=0.86) and (B) N2|3 (vs. N0|1) disease (area under the curve=0.88).
Figure E3. The predicted probability of N0, N1 or N2|3 disease for central vs. peripheral tumor locations as a function of age according to PET N stage (A) N stage by PET=N0, (B) N stage by PET=N1 and C) N stage by PET=N2 or N3.