Incorporation of a Molecular Prognostic Classifier Improves Conventional Non-Small Cell Lung Cancer Staging

Johannes R. Kratz, MD, a Greg J. Haro, MD, a Nancy R. Cook, ScD, b Jianxing He, MD, c Stephen K. Van Den Eeden, PhD, d Gavitt A. Woodard, MD, a Matthew A. Gubens, MD, a Thierry M. Jahan, MD, a Kirk D. Jones, MD, a Il-Jin Kim, PhD, a Biao He, PhD, a David M. Jablons, MD, a Michael J. Mann, MD,a,*

aUniversity of California, San Francisco, San Francisco, California
bDivision of Preventive Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts
cThe First Affiliated Hospital of Guangzhou Medical University, State Key Laboratory of Respiratory Disease, Guangzhou, People’s Republic of China
dKaiser Permanente Department of Research, Northern California; Oakland, California

Received 22 October 2018; revised 6 March 2019; accepted 26 March 2019
Available online - 5 April 2019

ABSTRACT

Introduction: Despite adoption of molecular biomarkers in the management of NSCLC, the recently adopted eighth edition of the TNM staging system utilized only clinicopathologic characteristics and validated improvement in risk stratification of early-stage disease has remained elusive. We therefore evaluated the integration of a clinically validated molecular prognostic classifier into conventional staging.

Methods: A novel staging system, the TNMB (with the B denoting biology) system, which integrates a 14-gene molecular prognostic classifier into the eighth edition of the TNM staging system, was developed by using data from 321 patients with NSCLC at the University of California, San Francisco. The TNMB staging system was subsequently validated in an independent, multicenter cohort of 1373 patients, and its implementation was compared with adoption of the seventh and eighth edition staging systems utilizing metrics of reclassification.

Results: Compared with staging according to the eighth edition of the TNM system, the TNMB staging system enhanced the identification of high-risk patients, with a net reclassification improvement of 0.33 (95% confidence interval [CI]: 0.24–0.41). It better predicted differences in survival, with a relative integrated discrimination improvement of 22.1% (95% CI: 8.8%–35.3%), and it improved agreement between observed and predicted survival, with a decrease in the reclassification calibration statistic of from 39 to 21. The seventh and eighth editions failed to change the net reclassification improvement (0.01 [95% CI: −0.04 to 0.03] and 0.03 [95% CI: 0.00 to 0.06], respectively) or relative integrated discrimination improvement (2.1% [95% CI: −5.8 to 9.9] and −2.5% [95% CI: −17.6 to 12.4], respectively); in addition, the eighth edition worsened calibration, with an increase in the reclassification calibration statistic from 23 to 25.

Conclusions: Incorporation of a molecular prognostic classifier significantly improved identification of high-risk patients and survival predictions compared with when conventional staging is used. The TNMB staging system may lead to improved survival of early-stage disease through more effective application of adjuvant therapy.

© 2019 International Association for the Study of Lung Cancer. Published by Elsevier Inc. All rights reserved.

*Corresponding author.

Disclosure: Drs. Kratz, Jablons, and Mann report having a consulting relationship with Razor Genomics, Inc, the company that established a Clinical Laboratory Improvement Amendments-certified laboratory and developed the molecular assay based on University of California, San Francisco, technology; they also report being inventors of related technology owned by the University of California, for which the University submitted patent applications that have been licensed to Razor Genomics, Inc. The remaining authors declare no conflict of interest.

Address for correspondence: Michael J. Mann, MD, Division of Adult Cardiothoracic Surgery, University of California, San Francisco. 1600 Divisadero St., Room A-743, San Francisco, CA 94143-1724, 415-353-7502. E-mail: michael.mann@ucsf.edu

© 2019 International Association for the Study of Lung Cancer. Published by Elsevier Inc. All rights reserved.

ISSN: 1556-0864

https://doi.org/10.1016/j.jtho.2019.03.015
Keywords: Non–small cell lung cancer; Eighth edition staging system; Molecular prognostic classifier; Reclassification statistics

Introduction

In 1907, Halsted et al described the natural history of solid tumors as growth in size, followed by spread to the lymph nodes and, ultimately, metastasis to other organs. The descriptors of tumor size, nodal status, and presence of metastases were integrated and codified as the TNM staging system by Denoix in 1953. Despite significant advances in our understanding of molecular pathways that drive fundamental tumor biology, these same basic descriptors still form the basis of both our prognostic and treatment frameworks more than 100 years later.

One of the greatest limitations to conventional staging in NSCLC has been its inability to differentiate high-risk patients who are likely to harbor occult metastasis after primary tumor resection from low-risk patients who have more likely been cured by surgery alone. This limitation is reflected in the persistently high mortality observed in early-stage disease and in the fact that many of those patients who later prove to have metastatic disease never received potentially lifesaving, early postoperative intervention.

The TNM staging system in NSCLC has been revised twice within the past decade; the eighth edition was recently adopted in 2018. The changes in the eighth edition include increasing the number of tumor size cutoff points, including more stage subdivisions, and distinguishing between different forms of metastasis. Despite the eighth edition’s comprehensive training data set, multiple external validations have yielded disappointing results, and 5-year overall survival continues to be as low as 55% to 60% among patients with pathological stage I NSCLC. The difficulty in improving risk stratification for early-stage NSCLC is likely related to the limitations of using the same set of basic TNM descriptors. Little further refinement to current NSCLC staging may be possible without adoption of more biologically sophisticated predictors.

In the past decade, molecular genetics has fundamentally changed our understanding of tumor biology and how it relates to both progression of disease and patient outcomes. Multiple studies have demonstrated that biomarkers can be both prognostic of survival and predictive of responses to specific systemic therapies. This has led to the suggestion that the next major advance in cancer staging will be the addition of molecular descriptors of tumor biology to conventional TNM staging systems.

Our group has published the largest and most comprehensive independent validation of a molecular prognostic classifier of tumor biology in NSCLC. We hypothesized that the addition of this biologic signature to conventional TNM staging would lead to improvements in risk stratification. To evaluate this inquiry, we first developed and compared novel staging systems that integrated the molecular prognostic classifier with the eighth edition staging. The staging system with the best statistical performance was then validated in an independent, multicenter cohort of 1373 patients and compared with conventional TNM staging by using metrics of reclassification.

Materials and Methods

Patient Cohorts

A total of 1694 patients who underwent resection of nonsquamous NSCLC as previously described were included in this study. In all, 321 patients with stage I to IIIC disease who underwent resection at the University of California, San Francisco (UCSF), between 1997 and 2007 and for whom molecular prognostic and staging data were available were included in the model development cohort. A total of 1373 patients with stage I to IIIC disease resected at community hospitals belonging to the Kaiser Permanente Northern California system (henceforth referred to as Kaiser) between 1998 and 2005 and three leading academic institutions in the China Clinical Trials Consortium (CCTC) between 2000 and 2008 were included in the validation cohort. As these patients were initially staged according to the sixth and seventh editions of the TNM staging system, they were restaged according to the new eighth edition staging system for NSCLC by utilizing pathologic tumor size and nodal status. The size of the lepidic component and invasion of surrounding structures were not used to restage patients, as they were not recorded in the pathology reports at the time of resection. Follow-up, vital status, and date of death were obtained from review of medical records and verified by the Kaiser Permanente Northern California Cancer Registry, California Death Records, Social Security Death Master File, and direct patient contact.

The study was approved by the institutional review boards of UCSF and the Kaiser Permanente Division of Research. Each of the Chinese institutions complied with the International Guidelines for Good Clinical Practice, and their respective institutional review boards permitted research with clinical data handled in blinded fashion.

Molecular Prognostic Classifier

The molecular prognostic classifier integrates expression levels of 11 cancer-related target genes (BCL2 associated athanogene gene [BAG1], BRCA1, DNA
repair associated gene [BRCA1], DNA repair associated gene 2 [BRCA1], cell division cycle 6 gene [CDC6], cyclin dependent kinase 2 associated protein 1 gene [CDK2AP1], erb-b2 receptor tyrosine kinase 3 gene [ERBB3], fucosyltransferase 3 (Lewis blood group) gene [FUT3], interleukin 11 gene [IL11], LCK proto-oncogene, SRC family tyrosine kinase gene [LCK], and Wnt family member 3A gene [WNT3A]) against a background of three reference genes (esterase D gene [ESD], TATA-box binding protein gene [TBP], and YAP associated protein 1 gene [YAP1]). The fully specified algorithm and cutoff points for low, intermediate, and high risk of recurrence were published previously. All assays were performed in a Clinical Laboratory Improvement Amendments–certified laboratory on samples of formalin-fixed, paraffin-embedded tumor tissue, as previously described. This certified molecular prognostic classifier has undergone extensive technical validation and has been demonstrated to be highly reproducible.

Of note, the molecular prognostic classifier was developed specifically for nonsquamous NSCLC, as previous work in our laboratory and the laboratories of others had suggested fundamental differences in molecular biology between squamous and nonsquamous NSCLC.

Model Development

We devised a method of supervised reclassification that was utilized to integrate tumor biology into the eighth edition staging system. Supervised reclassification maintains the order and priority of the current eighth edition but allows for modifications based on the molecular prognostic classifier’s risk of recurrence: low, intermediate, or high. The UCSF development cohort was split (by using a 3:1 ratio) into training (n = 240) and testing (n = 81) groups, respectively. Various degrees of upstaging and downstaging were assessed in the training group, such as by upstaging by one stage for high risk of recurrence, downstaging by one stage for low risk of recurrence, upstaging by two stages for high risk of recurrence, or downstaging by two stages for low risk of recurrence. The top three performing staging systems were selected for analysis and were then evaluated in the testing group by using the reclassification statistics described later in this article. The model with the best performance was independently validated in the Kaiser/CCTC cohort and compared with conventional staging.

Statistics

The primary outcome measure was overall survival after surgical resection. Overall survival was defined as death from any cause. Of note, overall survival was used instead of disease-free survival because it was verifiable by multiple sources and not as susceptible to incomplete medical records or interpretation bias. Survival time was defined as the time from surgical resection to death or censoring. Patients were administratively censored 5 years after surgical resection, and it was assumed that censoring was independent of survival time. Kaplan-Meier analysis and Cox proportional hazard models were used to estimate observed and predicted risk of survival at 5 years based on the different staging systems with use of the stcoxgr package in Stata 15.1 software (StataCorp LP; College Station, TX).

Time-dependent area under the receiver operating characteristic curve concordance index (C-index) was calculated to assess the overall model performance, and Royston’s modification of Nagelkerke’s $R^2$ statistic was calculated by using the str2ph package in Stata 15.1 (StataCorp LP; College Station, TX).

The categorical net reclassification improvement (NRI) was calculated from the survival probabilities of the ordered staging systems among those moving up and down stage classifications.

$$NRI = \left[ P(\text{event|up}) - P(\text{event}) \right] \times P(\text{up}) + \left[ P(\text{event|down}) - P(\text{event|down}) \right] \times P(\text{down})/P(\text{event})$$

The integrated discrimination improvement (IDI) was calculated as the difference in the proportion of explained variation at time $t$ ($R^2(t)$) between staging systems, and the relative IDI (rIDI) was calculated as the percentage improvement between staging systems using str2ph package Stata 15.1 (StataCorp LP; College Station, TX).

$$\text{IDI}(t) = \left( R^2(t)_{\text{NewStage}} - R^2(t)_{\text{OldStage}} \right)$$

$$\text{rIDI}(t) = \left( R^2(t)_{\text{NewStage}} - R^2(t)_{\text{OldStage}} \right)/R^2(t)_{\text{OldStage}}$$

The reclassification calibration (RC) statistic with utilization of the Greenwood-Nam-D’Agostino goodness-of-fit test was calculated by using observed and predicted risk of survival at 5 years in reclassification table cells with at least five events. The calculation was performed with R software (version 3.4.4, R Foundation) as described by Demler et al. Predicted risk was calculated in the validation cohort by using Cox proportional hazard models based on the ordered staging systems. The 95% confidence intervals (CIs) were constructed by using bootstrap resampling. For all statistical tests, a prespecified two-sided $\alpha$ value of 0.05 was considered.
significant. All analyses were performed with Stata 15.1 software (StataCorp LP, College Station, TX) unless otherwise noted. Of note, adoption of the seventh edition of the TNM staging system refers to a comparison of the seventh versus sixth editions and adoption of the eighth edition refers to a comparison of the eighth versus seventh editions.

Results
A total of 1694 patients who underwent resection of nonsquamous NSCLC were included in this study (Table 1). There were no substantial differences between the UCSF, Kaiser, and CCTC cohorts. More specifically, there were no differences between the cohorts in terms of sex, smoking history, NSCLC subtype, or 5-year mortality. The CCTC patients were slightly younger and had a somewhat more advanced disease stage than the UCSF and Kaiser patients did.

We devised a method of supervised reclassification that integrated tumor biology with the eighth edition staging and generated various staging systems in the UCSF development cohort (training group n = 240, testing group n = 81). Supervised reclassification maintained the order and priority of the current eighth edition staging system but allowed for modifications based on the molecular prognostic classifier’s risk of recurrence: low, intermediate, or high. The top three performing staging systems developed in the training group were then evaluated in the testing group (Table 2). All supervised reclassification staging systems outperformed the eighth edition and improved the prediction of overall survival. However, the –1/0/+1 staging system better identified high-risk patients with the largest change in the NRI and better predicted differences in survival with the largest change in discrimination while maintaining model fit comparable to that of the other staging systems. For these reasons, the –1/0/+1 staging system was chosen for validation and renamed the TNMB staging system (with the B denoting biology [Table 3]).

The TNMB staging system was subsequently applied to the independent Kaiser/CCTC validation cohort (n = 1373 [Supplementary Table 1]). In all, 78% of patients were reclassified with the TNMB staging system. The predominant effect of incorporating tumor biology was upward classification of patients into more advanced disease stage than the UCSF and Kaiser patients did.

### Table 1. Summary of Patient and Tumor Characteristics in the Development and Validation Cohorts (N = 1694)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>UCSF (Development Cohort [n = 321])</th>
<th>Kaiser (Validation Cohort [n = 418])</th>
<th>CCTC (Validation Cohort [n = 955])</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age at resection ± SD, y</td>
<td>67.5 ± 10.6</td>
<td>66.5 ± 9.3</td>
<td>58.3 ± 10.7</td>
</tr>
<tr>
<td>Females, n (%)</td>
<td>191 (59.5)</td>
<td>191 (45.7)</td>
<td>597 (62.5)</td>
</tr>
<tr>
<td>Smoking history, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>215 (67.0)</td>
<td>354 (84.7)</td>
<td>489 (51.2)</td>
</tr>
<tr>
<td>No</td>
<td>55 (17.1)</td>
<td>35 (8.4)</td>
<td>394 (41.3)</td>
</tr>
<tr>
<td>Unknown</td>
<td>51 (15.9)</td>
<td>29 (6.9)</td>
<td>72 (7.5)</td>
</tr>
<tr>
<td>Survivor follow-up, mo (IQR)</td>
<td>69.5 [47.0–91.0]</td>
<td>102.0 [85.0–120.0]</td>
<td>55.0 [37.8–70.5]</td>
</tr>
<tr>
<td>5-Year mortality, n (%)</td>
<td>127 (42.5)</td>
<td>179 (43.9)</td>
<td>401 (47.1)</td>
</tr>
<tr>
<td>Lung cancer subtype, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>264 (82.2)</td>
<td>324 (77.5)</td>
<td>869 (91.0)</td>
</tr>
<tr>
<td>Large cell carcinoma</td>
<td>17 (5.3)</td>
<td>15 (3.6)</td>
<td>17 (1.8)</td>
</tr>
<tr>
<td>Adenosquamous</td>
<td>10 (3.1)</td>
<td>15 (3.6)</td>
<td>46 (4.8)</td>
</tr>
<tr>
<td>Not otherwise specified</td>
<td>30 (9.4)</td>
<td>64 (15.3)</td>
<td>23 (2.4)</td>
</tr>
<tr>
<td>Eighth edition staging, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA1</td>
<td>12 (3.7)</td>
<td>18 (4.3)</td>
<td>10 (1.1)</td>
</tr>
<tr>
<td>IA2</td>
<td>82 (25.6)</td>
<td>137 (32.8)</td>
<td>127 (13.3)</td>
</tr>
<tr>
<td>IA3</td>
<td>74 (23.1)</td>
<td>128 (30.6)</td>
<td>148 (15.6)</td>
</tr>
<tr>
<td>IB</td>
<td>46 (14.3)</td>
<td>72 (17.2)</td>
<td>94 (9.8)</td>
</tr>
<tr>
<td>Stage II</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IIA</td>
<td>21 (6.5)</td>
<td>35 (8.4)</td>
<td>62 (6.5)</td>
</tr>
<tr>
<td>IIB</td>
<td>44 (13.7)</td>
<td>14 (3.4)</td>
<td>208 (21.8)</td>
</tr>
<tr>
<td>Stage III</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IIIA</td>
<td>40 (12.5)</td>
<td>14 (3.4)</td>
<td>256 (26.8)</td>
</tr>
<tr>
<td>IIIB</td>
<td>2 (0.6)</td>
<td>0 (0.0)</td>
<td>50 (5.2)</td>
</tr>
<tr>
<td>IIIC</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

UCSF, University of California, San Francisco; Kaiser, Kaiser Permanente Northern California; CCTC, China Clinical Trials Consortium; IQR, interquartile range.
Table 2. Performance and Improvement Measures from the Supervised Reclassification Staging Systems in the Testing Group of the Development Cohort (n = 81)

<table>
<thead>
<tr>
<th>Measures</th>
<th>Eighth Edition</th>
<th>-1/0/+1 (TNMB)</th>
<th>-2/0/+1</th>
<th>-2/+1/+2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.14</td>
<td>0.19</td>
<td>0.19</td>
<td>0.18</td>
</tr>
<tr>
<td>C-index</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>Improvement Measures Compared to with in the eighth edition</td>
<td>0.51</td>
<td>0.51</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Net reclassification improvement</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Integrated discrimination improvement</td>
<td>34.5%</td>
<td>29.7%</td>
<td>29.3%</td>
<td></td>
</tr>
</tbody>
</table>

Note: Revised staging systems were generated by maintaining the order and priority of the eighth edition staging of NSCLC but allowing for modifications based on the molecular prognostic classifier’s risk of recurrence: low, intermediate, or high (i.e., -1/0/+1 (TNMB)), down-stage by one stage if low risk/no change in stage if intermediate risk/ up-stage by one stage if high risk; -2/0/+1, down-stage by two stages if low risk/no change in stage if intermediate risk/up-stage by one stage if high risk; -2/+1/+2, down-stage by two stages if low risk/up-stage by one stage if intermediate risk/up-stage by two stages if high risk). In the testing group, all staging systems outperformed the eighth edition. The -1/0/+1 staging system had the most improvement in reclassification and discrimination, while maintaining model fit comparable to that of the other staging systems. It was chosen for validation and renamed the TNMB staging system (with the $B$ denoting biology).

$R^2$ denotes Royston’s modification of Nagelkerke’s $R^2$; C-index denotes the concordance index/area under the receiver operating characteristic curve.

Overall model fit was then evaluated from adoption of the seventh edition, eighth edition, and TNMB staging systems (Fig. 1 and Tables 4 and 5). The modified Nagelkerke’s $R^2$ statistic demonstrated better model fit in TNMB staging than with adoption of the seventh edition or eighth edition staging ($R^2 = 0.17$ with TNMB staging, $R^2 = 0.14$ with staging using the seventh and eighth editions [see Table 4]). Kaplan-Meier analysis of overall survival demonstrated superior survival curve separation in TNMB staging than with conventional staging (see Fig. 1). These results were corroborated through evaluation of the chi-square statistic of the log-rank test and C-index. There was minimal change in the chi-square statistic of the log-rank test for trend by stage in conventional staging (159 according to the seventh edition versus 155 according to the eighth edition). In contrast, incorporation of the molecular prognostic classifier into TNMB staging led to an improved chi-square statistic of 194. This finding illustrates improved survival discrimination in TNMB staging compared with in ongoing modifications of TNM criteria in conventional staging. Moreover, the C-index or area under the receiver operating characteristic curve was larger with TNMB staging than with staging using the seventh or eighth edition (a C-index of 0.66 with TNMB staging versus 0.64 with staging using the seventh and eighth editions [see Table 4]).

In recent years, more sophisticated reclassification metrics have been applied to better assess improvements in prognostic systems. The TNMB staging system was found to be superior to conventional staging by these more modern and clinically relevant reclassification metrics (see Table 5). Incorporation of tumor biology into TNMB staging yielded an NRI of 0.33 (95% CI: 0.24–0.41) when TNMB staging is compared with...
staging according to the eighth edition. This finding demonstrates that adoption of TNMB staging improved the identification of high-risk patients. In contrast, there was no change in NRI from adoption of the seventh or eighth editions (0.01 [95% CI: –0.04 to 0.03] and 0.03 [95% CI: –0.00 to 0.06], respectively [see Table 5]).

The IDI and rIDI quantify the relative improvement in the ability of the new staging system to distinguish between those who survive and those who die (see Table 5). A comparison of TNMB staging with staging according to the eighth edition resulted in an IDI of 0.03 (95% CI: 0.01–0.05) and an rIDI of 22.1% (95% CI: 8.8%–35.3%), whereas adoption of the seventh or eighth editions failed to demonstrate any benefit (0.00 [95% CI: –0.01 to 0.01] versus 2.1% [95% CI: –5.8% to 9.9%] and 0.00 [95% CI: –0.02 to 0.02] versus –2.5% [95% CI: –17.6% to 12.4%], respectively). This finding displays improved discrimination between survivors and non-survivors from adoption of TNMB staging.

Agreement in observed and predicted risks of death at 5 years among reclassified patients for each staging system can be directly compared by using the RC statistic (Fig. 2 and see also Table 5). Counterintuitively, improved calibration is captured by a smaller RC statistic and a p value greater than 0.05. Whereas adoption of the seventh edition decreased the RC statistic from 6 to 5 (an increase in p value from 0.64 to 0.78), adoption of the eighth edition worsened calibration and increased the RC statistic from 23 to 25 (a decrease in p value from 0.05 to 0.02). In contrast, adoption of TNMB staging in place of eighth edition TNM staging improved the agreement between predicted and observed outcomes and decreased the RC statistic from 39 to 21 (an increase in p value from <0.01 to 0.23).

### Discussion

The potential of molecular signatures to refine prognosis in NSCLC has been shown by multiple independent groups. In 2009, the year the seventh edition staging system was adopted, leaders from the International Association for the Study of Lung Cancer Staging Committee anticipated incorporation of molecular predictors of prognosis as part of future TNM staging efforts. However, molecular predictors were not incorporated into the eighth edition staging system.
Multiple methods of integrating molecular prognostic biomarkers into staging are possible. The method of integration is vitally important as it is likely to influence widespread adoption of the proposed model. The major barriers to clinical adoption include the lack of model generalizability to patients outside of the study and the use of complex, non-clinically intuitive models. The first barrier exists for several reasons, including model overfitting, use of small patient cohorts for model development, and lack of proper validation cohorts. Of note, external validation is critical in evaluation and comparison of model performance. Despite the eighth edition’s comprehensive training data set, multiple external validations, including the National Cancer Database, have yielded disappointing results, and survival continues to be low among early-stage patients. Complexity serves as an additional barrier to adoption as clinicians struggle both to understand and to apply the proposed model. Unfortunately, the proposal of complex models is common, as they tend to perform better in small development cohorts and provide more striking results for publication despite their potentially lower clinical relevance.

The supervised reclassification method proposed in this study is simple, transparent, and clinically intuitive. It maintains the order and priority of conventional staging but allows for modifications based on the molecular prognostic classifier. The incorporation of a fourth category representing biology is easy for clinicians to understand and apply, and it allows for flexibility to test and incorporate future refinements from novel characterizations of tumor biology. The resulting TNMB staging system was independently validated in an integrated population of two large-scale international cohorts drawn from community and academic hospitals.

In our patient cohort, adoption of the TNMB staging system significantly improved prediction of overall survival greater than adoption of the seventh or eighth edition staging systems. The addition of the molecular prognostic classifier improved identification of high-risk patients, discrimination between survivors and nonsurvivors, and calibration of predicted and observed outcomes. In contrast, adoption of the seventh or eighth editions offered no benefit in identification of high-risk patients or discrimination. Whereas calibration improved after adoption of the seventh edition, the agreement between predicted and observed outcomes worsened after adoption of the eighth edition. The TNMB staging system supports the notion that a substantial, clinically relevant improvement to staging is best achieved by incorporating molecular classifiers into staging rather than by continued fine-tuning of conventional TNM pathologic characteristics.

Improved staging is clinically important for all cancer, but it represents a vital unmet need in NSCLC. Survival of patients with early-stage NSCLC remains unacceptably low. As a result of understaging, many of these patients die from unrecognized metastatic disease without an attempt at potentially curative adjuvant therapy. It has been suggested by multiple randomized clinical trials that postoperative patients with NSCLC and worse prognoses are more likely to benefit from adjuvant therapy, whereas patients with better prognoses are less likely to benefit. In fact, prognostic information beyond TNM staging is currently endorsed by the National Comprehensive Cancer Network as a basis for decisions regarding the use of adjuvant therapy in early-stage NSCLC, despite little supporting data related to the proposed prognostic indicators. The molecular prognostic classifier used in this study has been demonstrated to provide a more reliable

### Table 5. Improvement Measures from Adoption of the Seventh Edition and Eighth Edition of TNM Staging and TNMB Staging for NSCLC in the Validation Cohort (n = 1373)

<table>
<thead>
<tr>
<th>Improvement Measuresa</th>
<th>Seventh Edition vs. Sixth Edition, value (95% CI)</th>
<th>Eighth Edition vs. Seventh Edition, value (95% CI)</th>
<th>TNMB vs. Eighth Edition, value (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net reclassification improvement</td>
<td>-0.01 (0.04 to 0.03)</td>
<td>0.03 (0.01 to 0.06)</td>
<td>0.33 (0.24-0.41)</td>
</tr>
<tr>
<td>Integrated discrimination improvement</td>
<td>0.00 (0.01 to 0.01)</td>
<td>-0.00 (0.02 to 0.02)</td>
<td>0.03 (0.01-0.05)</td>
</tr>
<tr>
<td>Relative integrated discrimination improvement</td>
<td>2.1% (5.8% to 9.9%)</td>
<td>-2.5% (-17.6% to 12.4%)</td>
<td>22.1% (8.8%-35.3%)</td>
</tr>
<tr>
<td>Reclassification calibration statistic</td>
<td>6.1 → 4.8</td>
<td>22.5 → 25.1</td>
<td>39.2 → 20.8</td>
</tr>
<tr>
<td>Change in p value</td>
<td>0.64 → 0.78</td>
<td>.05 → 0.02</td>
<td>&lt;0.01 → 0.23</td>
</tr>
</tbody>
</table>

Note: TNMB, a novel staging system proposed in this study, integrates a validated 14-gene molecular prognostic classifier into the eighth edition Staging. TNMB had improved model fit, reclassification, discrimination, and calibration compared to the eighth edition. There were no changes in model fit, reclassification, or discrimination and calibration worsened with continued fine-tuning of pathologic characteristics in conventional staging.

aNet reclassification improvement = (P(event|up) – P(event|down)) × P(up) – P(event|down) / P(event) × (1 - P(event)). Integrated discrimination improvement = (R²(OldStage) - R²(NewStage) / R²(OldStage)). Relative integrated discrimination improvement = (R²(OldStage) - R²(NewStage))/R²(OldStage). Reclassification calibration statistic utilizing the Greenwood-Nam-D’Agostino goodness-of-fit test was applied at 5 years (smaller value denotes improvement).

bChange to larger value denotes improvement.

cI, confidence interval.
prognosis than the National Comprehensive Cancer Network–recommended clinicopathologic factors. The upward classification of patients into more advanced disease stages observed in this study indicates that a significant number of patients may benefit from additional adjuvant therapy. In fact, recently published data have provided initial confirmatory support that patients within stage I identified by this molecular prognostic classifier as high risk do benefit from postoperative intervention that might otherwise have been withheld on the basis of staging and clinicopathologic criteria alone.

The greatest impact of any modification to current staging systems is likely achieved when the updated

Figure 2. Calibration of seventh edition, eighth edition, and TNMB staging of NSCLC in the validation cohort (n = 1373). Solid lines denote predicted risk of survival, and individual points denote observed risk of survival with 95% confidence intervals. The tables under the graphs depict observed and predicted risk of 5-year survival. TNMB, a novel staging system proposed in this study that integrates a molecular prognostic classifier with the eighth edition staging system, has agreement between predicted and observed outcomes superior to that provided by conventional staging.
system is incorporated into future randomized clinical studies. Research based on enhanced risk stratification can lead to substantial improvements in clinical care at the population level. The results of this study suggest that improved prognostication may be better achieved through the incorporation of a well-validated molecular risk profile than through ongoing reorganization of conventional clinicopathologic information.

Limitations of this study include its retrospective nature and the lack of information about other potential prognostic and predictive biomarkers in our patient cohort, such as $EGFR$, $KRAS$, and echinoderm microtubule associated protein like 4 gene ($EML4$)–$ALK$ receptor tyrosine kinase gene ($ALK$). In addition, restaging of patients from the seventh to eighth edition was performed by utilizing pathologic tumor size and nodal status, and we did not have information regarding the size of the lepidic component or invasion of surrounding structures. A small number of patients may therefore have been staged inaccurately; however, any study that restages patients from a previous methodology will have similar information gaps, as these details are not recorded in the pathology reports at the time of resection.

We also recognize that additional tumor specimen testing incurs additional costs. However, a more accurate prognostic staging system has the potential to utilize health care resources more efficiently by directing relatively inexpensive adjuvant therapy to the patient populations that are most likely to benefit and avoiding recurrence and more expensive late-stage treatments that are significantly less effective. Molecularly tailored adjuvant therapy, therefore, may ultimately reduce recurrence, improve survival, and reduce costs.

The field of molecular biomarkers in thoracic oncology is rapidly advancing. This study offers the first demonstration of the successful incorporation of a molecular prognostic classifier into a coherent lung cancer staging system.

Supplementary Data
Note: To access the supplementary material accompanying this article, visit the online version of the Journal of Thoracic Oncology at www.jto.org and at: https://doi.org/10.1016/j.jtho.2019.03.015.

References


