

# Application of the revised lung cancer staging system (IASLC Staging Project) to a cancer center population

Edmund S. Kassis, MD,<sup>a</sup> Ara A. Vaporciyan, MD,<sup>a</sup> Stephen G. Swisher, MD,<sup>a</sup> Arlene M. Correa, PhD,<sup>a</sup> B. Nebiyu Bekele, PhD,<sup>b</sup> Jeremy J. Erasmus, MD,<sup>c</sup> Wayne L. Hofstetter, MD,<sup>a</sup> Ritsuko Komaki,<sup>d</sup> Reza J. Mehran,<sup>a</sup> Cesar A. Moran,<sup>e</sup> Katherine M. Pisters,<sup>f</sup> David C. Rice, MD,<sup>a</sup> Garrett L. Walsh, MD,<sup>a</sup> and Jack A. Roth, MD<sup>a</sup>

**Objective:** The International Association for the Study of Lung Cancer (IASLC) proposed a revision to the Union Internationale Contre le Cancer (UICC-6) staging system for non-small cell lung cancer. The goal of our study was to compare these systems in patients undergoing surgery for non-small cell lung cancer to determine whether one system is superior in staging operable disease.

**Methods:** Pathologic stages in 1154 patients undergoing complete resection over a 9-year period were analyzed. Patients were assigned a stage based on both IASLC and UICC-6 systems. We tested for statistically meaningful differences between the two staging systems using the Wilcoxon signed rank test and the permutation test.

**Results:** The IASLC system is more effective than the UICC-6 system at ordering and differentiating patients ( $P = .009$ ). Application of the IASLC system resulted in 202 (17.5%) patients being reassigned to a different stage ( $P = .012$ ), with the most common shifts occurring from IB to IIA and IIIB to IIIA. The 5-year and median survivals of the IASLC IIIA patients including those shifted from the UICC-6 IIIB were 37% and 35 months, respectively. Reclassifying UICC-6 IIIB to IASLC IIIA did not reduce survival for the newly characterized IIIA cohort.

**Conclusion:** Our data confirm that the proposed IASLC staging system is more effective at differentiating stage than the UICC-6 system. Reclassifying patients from UICC-6 IIIB to IASLC IIIA will shift some patients from a stage previously considered unresectable to a stage frequently offered surgical resection. Further study and validation of the IASLC system are warranted.

Supplemental material is available online.

Despite the overall poor prognosis of patients with lung cancer, there are subsets of patients who benefit from treatment.<sup>1-4</sup> Effective staging systems stratify patient survival and can be used to assess outcome of defined patient subgroups after treatment. The sixth edition of the Union Inter-

nationale Contre le Cancer (UICC-6) and the American Joint Committee on Cancer (AJCC) has served as the current tumor, node, metastases (TNM) staging system for non-small cell lung cancer (NSCLC) since 2002.<sup>5</sup> The UICC-6 system is derived from the 1997 staging system proposed by Mountain.<sup>6</sup> This staging system was based on 5319 patients treated for primary lung cancer at The University of Texas—M. D. Anderson Cancer Center (UTMDACC) (4351 patients) from 1975 to 1988 or by the National Cancer Institute Cooperative Lung Cancer Study Group (968 patients) from 1977 to 1982. This represents primarily a single-institution experience from a single country. The current staging system has considerable intrastage heterogeneity with groups within a stage varying widely in prognosis.

In 1998 the International Association for the Study of Lung Cancer (IASLC) staging project was initiated to develop the next revision of the current UICC-6 system.<sup>7-11</sup> The proposed revision represents data collected from 100,869 patients from Europe, Australia, Asia, and North America. The data were analyzed by Cancer Research and Biostatistics and the IASLC International Staging Committee. The revised system proposes changes to the T and M classifications (Table 1) and overall stage groupings (Table 2). The revised TNM staging has been submitted for approval to the UICC. The IASLC system has yet to be independently evaluated.

From the Departments of Thoracic and Cardiovascular Surgery,<sup>a</sup> Biostatistics,<sup>b</sup> Radiology,<sup>c</sup> Radiation Oncology,<sup>d</sup> Pathology,<sup>e</sup> and Thoracic/Head and Neck Medical Oncology,<sup>f</sup> The University of Texas M. D. Anderson Cancer Center, Houston, Tex.

This work was partially supported by grants from the National Cancer Institute and the National Institute of Health: Specialized Program of Research Excellence (SPORE) in Lung Cancer (2P50-CA70907); by The University of Texas M. D. Anderson Cancer Center Support Core Grant (CA 16672); by a grant from the Tobacco Settlement Funds as appropriated by the Texas State Legislature (Project 8); by the W. M. Keck Foundation; and a sponsored research agreement with Introgen Therapeutics, Inc.

Read at the Eighty-eighth Annual Meeting of The American Association for Thoracic Surgery, San Diego, Calif, May 10–14, 2008.

Received for publication May 8, 2008; revisions received Oct 13, 2008; accepted for publication Jan 13, 2009; available ahead of print May 29, 2009.

Address for reprints: Jack A. Roth, MD, Department of Thoracic and Cardiovascular Surgery, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd, Box 445, Houston, TX 77030-4095 (E-mail: jroth@mdanderson.org).

J Thorac Cardiovasc Surg 2009;138:412-8  
0022-5223/\$36.00

Copyright © 2009 by The American Association for Thoracic Surgery  
doi:10.1016/j.jtcvs.2009.01.033

<b>Abbreviations and Abstracts</b>	
AJCC	= American Joint Committee on Cancer
IASLC	= International Association for the Study of Lung Cancer
NSCLC	= non–small cell lung cancer
TNM	= tumor, node, metastasis
UICC	= Union Internationale Contre le Cancer
UTMDACC	= University of Texas M. D. Anderson Cancer Center

The goal of our study was to apply the proposed changes to the current UICC-6 staging system to a cancer center population undergoing surgery for NSCLC and to directly compare the proposed IASLC and UICC-6 staging systems with respect to discrimination, monotonicity and intrastage heterogeneity.

PATIENTS AND METHODS

Population

This study analyzed data from a prospectively collected database of 1154 patients who underwent an R0 surgical resection for NSCLC at UTMDACC between 1998 and 2006. UTMDACC was a contributor of patient data for the IASLC study. Less than 5% of the patients in our study were the same

TABLE 1. Comparison of T and M stage of UICC-6 and IASLC staging systems

<b>A. UICC-6 staging system</b>	
Tx: Primary tumor cannot be assessed, <i>or</i> tumor proven by the presence of malignant cells in sputum or bronchial washings but not visualized by imaging or bronchoscopy	
T0: No evidence of primary tumor	
Tis: Carcinoma in situ	
T1: Tumor 3 cm or less in greatest dimension, surrounded by lung or visceral pleura, without evidence of invasion more proximal than the lobar bronchus	
T2: Tumor more than 3 cm in diameter; or tumor with <i>any</i> of the following features:	
<ul style="list-style-type: none"><li>• Involves main bronchus, 2 cm or more distal to the carina</li><li>• Invades visceral pleura</li><li>• Associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung</li></ul>	
T3: Tumor more than 7 cm or	
<ul style="list-style-type: none"><li>• Direct invasion any of the following: chest wall (including superior sulcus tumours), diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium</li><li>• Tumor in the main bronchus less than 2 cm distal to the carina without carinal invasion</li><li>• Associated atelectasis or obstructive pneumonitis of the entire lung</li></ul>	
T4: Tumor of any size that invades any of the following:	
<ul style="list-style-type: none"><li>• Mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body or carina</li><li>• Separate tumor nodule(s) in the ipsilateral primary lobe</li><li>• Malignant pleural effusion</li></ul>	
M1: Distant metastases	
<b>B. IASLC staging system</b>	
Tx: Primary tumor cannot be assessed, <i>or</i> tumor proven by the presence of malignant cells in sputum or bronchial washings but not visualized by imaging or bronchoscopy	
T0: No evidence of primary tumor	
Tis: Carcinoma in situ	
T1: Tumor 3 cm or less in greatest dimension, surrounded by lung or visceral pleura, without evidence of invasion more proximal than the lobar bronchus	
T1a: Tumor 2 cm or less in greatest dimension	
T1b: Tumor more than 2 cm but not more than 3 cm in greatest dimension	
T2: Tumor more than 3 cm but not more than 7 cm or tumor with <i>any</i> of the following features:	
<ul style="list-style-type: none"><li>• Involves main bronchus, 2 cm or more distal to the carina</li><li>• Invades visceral pleura</li><li>• Associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung</li></ul>	
T2a: Tumor more than 3 cm but not more than 5 cm in greatest dimension	
T2b: Tumor more than 5 cm but not more than 7 cm in greatest dimension	
T3: Tumor more than 7 cm or	
<ul style="list-style-type: none"><li>• Direct invasion any of the following: chest wall (including superior sulcus tumours), diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium</li><li>• Tumor in the main bronchus less than 2 cm distal to the carina</li><li>• Associated atelectasis or obstructive pneumonitis of the entire lung</li><li>• Separate tumor nodule(s) in the same lobe</li></ul>	
T4: Tumor of any size that invades any of the following: mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body or carina	
Separate tumor nodule(s) in a different ipsilateral lobe	
M1a: Tumor nodule in contralateral lung, tumor with pleural nodules, malignant effusion	
M1b: Distant metastases	

UICC, Union Internationale Contre le Cancer; IASLC, International Association for the Study of Lung Cancer.

**TABLE 2. Comparison of TNM stage groupings of IASLC versus UICC-6 staging systems**

Stage	UICC-6	IASLC
Stage IA	T1 N0 M0	T1a N0 M0 T1b N0 M0
Stage IB	T2 N0 M0	T2a N0 M0
Stage IIA	T1 N1 M0	T1a N1 M0 T1b N1 M0 T2a N1 M0 T2b N0 M0
Stage IIB	T2 N1 M0 T3 N0 M0	T2b N1 M0 T3 N0 M0
Stage IIIA	T3 N1 M0 T1-3 N2 M0	T1a N2 M0 T1b N2 M0 T2a N2 M0 T2b N2 M0 T3 N1 M0 T3 N2 M0 T4 N0 M0 T4 N1 M0
Stage IIIB	T4, Any N, M0 Any T, N3 M0	T4 N2 M0 Any T, N3, M0
Stage IV	Any T, Any N, M1	Any T, Any N, M1 a/b

UICC, Union Internationale Contre le Cancer; IASLC, International Association for the Study of Lung Cancer.

patients as those in the IASLC study, and of those patients who were the same, the length of follow-up was different in the two studies. The two populations can be treated as two distinct data sets.

Institutional review board approval was obtained for this study, and informed consent was obtained from each participant.

Patients were excluded from analysis if histologic type was small cell carcinoma, neuroendocrine carcinoma, or predominantly bronchoalveolar carcinoma. Chemotherapy or radiation therapy administration did not exclude patients from analysis. A histologic classification of adenocarcinoma with bronchoalveolar features was included in the analysis. Pathologic staging of resected specimens was based on the application of the UICC-6 TNM staging system. Nodal (N) classification for each patient was determined either by systematic lymph node dissection or by lymph node sampling. Each specimen was then reassigned a surgical TNM classification and overall stage on the basis of the IASLC system.

Data on pathologic TNM classification, overall stage, and outcome for patients treated at UTMDACC were collected prospectively using Internet-based data collection entered by the health provider at the point of care. The data are subjected to periodic reviews for quality control.

### Statistical Analyses

We tested for statistically meaningful differences between the UICC-6 and IASLC staging systems with respect to stage assignment using a Wilcoxon signed rank test. Statistical analysis was performed with SPSS software (SPSS, Inc, Chicago, Ill).

Survival probabilities were assessed by the Kaplan–Meier method calculated from the date of surgery until death or most recent follow-up. Operative mortality was excluded from survival analysis to allow for assessment of long-term rather than short-term mortality. Each patient was assigned a T classification and overall stage grouping on the basis of both UICC-6 and IASLC staging systems (Tables 1 and 2). The prognostic significance of overall stage using both the UICC-6 and IASLC systems was determined by univariate analysis. The effectiveness of each staging system was evaluated statistically by a log–rank trend test. In addition, we also assessed whether one staging system is more effective than the other via a permutation test in which we construct

**TABLE 3. Patient characteristics**

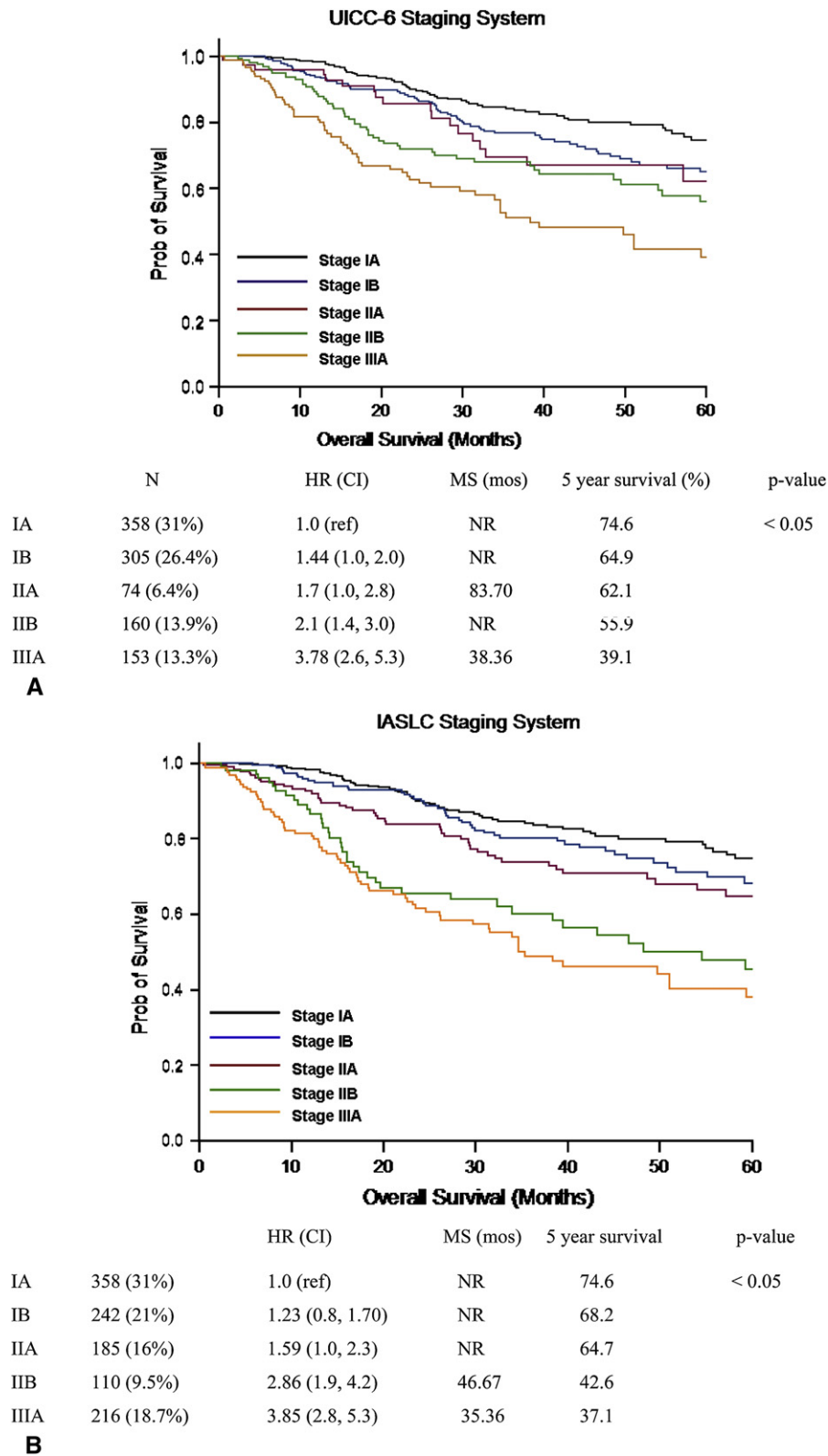
Age, y (median, range)	66 (32–90)
Sex (N, %)	
Male	607 (52.6)
Female	547 (47.4)
Histology	
Adenocarcinoma	658 (57%)
Squamous cell	388 (33.6%)
NSCLC (NOS)	62 (5.4%)
Large cell	24 (2.1%)
Adenosquamous	22 (1.9%)
Procedure	
Lobectomy/bilobectomy	947 (82.1%)
Pneumonectomy	94 (8.1%)
Wedge resection	65 (5.6%)
Segmentectomy	48 (4.2%)

NSCLC, Non–small cell lung cancer; NOS, not otherwise specified.

differences in log–rank trend test statistics under random rearrangements (ie, permutations) of the original labels assigned to the observed data. By repetition of this process many times (eg, 10,000 times), a null distribution is created and used to assess the difference between the log–rank trend statistic under the original labels and the randomly permuted labels.<sup>12</sup> Please see the included [appendix](#) for complete description of the statistical analysis.

### RESULTS

The population studied included all patients undergoing curative resection for NSCLC at UTMDACC between 1998 and 2006. The demographics of the study group are shown in Table 3. Each patient was assigned a pathologic T classification and overall stage on the basis of both the UICC-6 and IASLC staging systems. In 952 patients, application of the IASLC staging system resulted in no change from the UICC-6 assigned stage grouping. However, 202 (17.5%) patients were assigned a different stage grouping when the IASLC staging system was applied to their pathologic stage (Wilcoxon signed rank test;  $P = .012$ ). Of these 202 patients with a change in stage grouping, 73 (36.2%) patients were upstaged and 129 (63.8%) patients were downstaged by application of the IASLC staging system (Figure 1). Patients assigned a higher stage by the IASLC staging system included patients shifted from UICC-6 IB to IASLC IIA ( $n = 44$ , 60.2%) and IASLC IIB ( $n = 19$ , 26%) or UICC-6 IIB to IASLC IIIA ( $n = 10$ , 13.7%). Of the 73 patients who were assigned a higher stage by IASLC, none was assigned a stage that would preclude surgical resection. Application of the IASLC system resulted in downstaging patients from UICC-6 stages IIB, IIIB, and IV. Sixty-seven (53.2%) patients classified as being in UICC-6 IIB were downstaged to IASLC stage IIA. One hundred four (9%) patients were staged by the UICC-6 as having advanced locoregional disease precluding surgical resection. Of these 104 patients with stage IIIB or IV disease, 59 (56%) were restaged by the IASLC system as having potentially respectable disease. Of these, 59 patients in UICC-6 IIIB were downstaged with 51 (40.4%) patients shifted to



**FIGURE 1.** Kaplan–Meier survival by overall stage for UICC-6 (A) and IASLC (B). *HR*, Hazard ratio; *CI*, confidence interval; *MS*, median survival; *NR*, not reached. *UICC*, Union Internationale Contre le Cancer; *IASLC*, International Association for the Study of Lung Cancer.

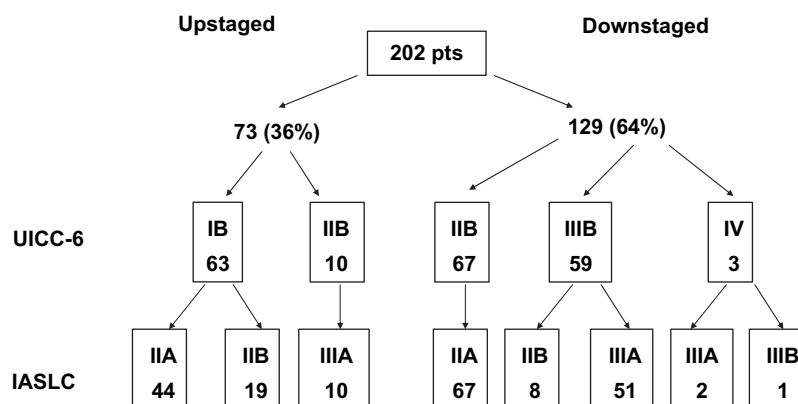


FIGURE 2. Shifting of stage after application of IASLC system. IASLC, International Association for the Study of Lung Cancer.

IASLC stage IIIA and 8 (6.3%) patients to IASLC stage IIB. Three patients were downstaged from stage IV to IIIA ( $n = 2$ ) and IIIB ( $n = 1$ ). The 5-year and median survivals of the IIIA patients in the IASLC system, including those shifted from UICC-6 IIIB, were 37% and 35 months, respectively. Shifting of patients from UICC stage IIIB, usually considered unresectable, to IASLC stage IIIA, in which patients are frequently offered surgical resection, did not result in a decrease in median or overall survival for IASLC stage IIIA patients. The 5-year survival of IASLC stage IIA including those shifted from UICC-6 IB was 64.7% (Figure 2).

### Statistical Comparison of UICC6 and IASLC Staging Systems

We assessed each staging system's ability to discriminate between lower stage and higher stage patients with respect to overall survival and monotonicity as assessed by strong inverse relationship between stage and overall survival using a permutation test described in the appendix. The IASLC staging system is significantly more effective with respect to discrimination and monotonicity than the UICC-6 system. Application of the permutation test on patients with operable disease showed that the IASLC staging system is more effective than the UICC-6 system at predicting overall survival of patients with operable disease on the basis of stage ( $P = .009$ ).

### DISCUSSION

Our study aim was to apply the IASLC T classification and overall stage groupings to a population of patients who underwent complete surgical resection for NSCLC at UTMDACC between 1998 and 2006. The proposed changes to the staging system represent a major change that will result in significant shifts of patients into higher or lower stages. We asked the question whether application of the IASLC staging system to a cancer center population replicates the findings of the IASLC International Staging Committee and whether the new IASLC staging system is an

improvement over the UICC-6 staging system. Our findings confirmed the ordering of stages reported for the IASLC staging system. Furthermore, direct comparison of the IASLC and UICC-6 staging systems using the permutation test demonstrated that the IASLC system is more effective at differentiating high, mid, and low stage groupings than the current UICC-6 system at a high level of statistical significance. This increased effectiveness of the IASLC staging system may help identify those patients at higher risk for recurrence. The improved stratification of survival and limitation of heterogeneity among patients within a stage may have important implications with regard to clinical research and adjuvant treatment decisions.

Our data also demonstrate a significant shifting of patients between stages when the IASLC system was applied to their pathologic stage. Of the 1154 patients in the study, 202 (17.5%) changed stage when the IASLC system was applied. One hundred four (9%) patients were staged by the UICC-6 as having advanced locoregional disease precluding surgical resection. Of these 104 patients with stage IIIB or IV disease, 59 (56%) were restaged by the IASLC system as having potentially resectable disease. Of the 73 patients who were assigned a higher stage by IASLC, none was assigned a stage that would preclude resection.

There are several limitations to this analysis. The study was based on a single-institution experience with a relatively small number of patients. The data were entered into a database prospectively but the patients were not entered into this study on a protocol. The type of surgical resection and extent of nodal dissection were left to the discretion of the operating surgeon. Also, there was no centralized pathologic review, and each pathologic specimen was evaluated on a case-by-case basis. The strength of the study lies in the prospective data collection, uniformity of the staging procedures for this patient population, and the independent evaluation by our statistician using a novel application of the permutation test to evaluate the two staging systems at ordering and differentiating patients between stages.



Several controversies will undoubtedly arise with the adoption of the IASLC staging system.

In the current UICC-6 system, T4 lesions are staged as IIIB regardless of lymph node status and are considered unresectable except in special circumstances. In this study, shifting of stage with application of the IASLC may potentially alter the management of 134 (11.6%) patients. Sixty-three of these patients were upstaged from a stage where surgery alone is the recommended treatment to a stage where adjuvant chemotherapy may be considered.<sup>13-15</sup> Additionally, 10 patients were upstaged to a stage where neoadjuvant chemotherapy is frequently offered (stage II to IIIA). The role of adjuvant and neoadjuvant chemotherapy in these patient populations may need to be re-evaluated. The IASLC system T4 lesions would be considered as IIIA or IIIB and the designation is based on the presence or absence of mediastinal nodal metastases. Satellite nodules in the ipsilateral primary lobe are considered unresectable T4 (stage IIIB) disease by UICC-6 criteria but T3 (stage IIB or IIIA) and potentially resectable by IASLC. Additionally, a satellite nodule in the ipsilateral lung but outside the primary lobe is unresectable M1 (stage IV) in the UICC-6 system and potentially resectable T4 (stage IIIA or IIIB) by IASLC. The optimal treatment strategy for these stages needs to be re-evaluated. Further study and validation of IASLC staging system and its effects on patient care are warranted.

## References

1. Naruke T, Goya T, Tsuchiya R, Suemasu K. Prognosis and survival in resected lung carcinoma based on the new international staging system. *J Thorac Cardiovasc Surg.* 1988;96:440-7.
2. Thomas PA, Piantadosi S. Lung Cancer Study Group. Postoperative T1 N0 non-small cell lung cancer. *J Thorac Cardiovasc Surg.* 1978;94:349-54.
3. Martini N, Burt ME, Bains MS, McCormack PM, Rusch VW, Ginsberg RJ. Survival after resection of stage II non-small cell lung cancer. *Ann Thorac Surg.* 1992;54:460-6.
4. Burt ME, Pomerantz AH, Bains MS, McCormack PM, Kaiser LR, Hilaris BS, et al. Results of surgical treatment of stage III lung cancer invading the mediastinum. *Surg Clin North Am.* 1987;67:987-1000.
5. AJCC cancer staging handbook/American Joint Committee on Cancer. 6th ed. Philadelphia: Lippincott-Raven; 2002.
6. Mountain CF. Revisions in the international staging for lung cancer. *Chest.* 1997; 111:1710-7.
7. Groome PA, Bolejack V, Crowley JJ, Kennedy C, Krasnik M, Sobin LH, et al. The IASLC Lung Cancer Staging Project: validation of the proposals for revision of the T, N, and M descriptors and consequent stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumors. *J Thorac Oncol.* 2007;2:694-705.
8. Rami-Porta R, Ball D, Crowley J, Giroux DJ, Jett J, Travis WD, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the T descriptors in the forthcoming (seventh) edition of the TNM classification of malignant tumors. *J Thorac Oncol.* 2007;2:593-602.
9. Goldstraw P, Crowley J, Chansky K, Giroux D, Groome PA, Rami-Porta R, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumors. *J Thorac Oncol.* 2007;2:706-14.
10. Rusch VW, Crowley J, Giroux DJ, Goldstraw P, Im JG, Tsuboi M, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the N descriptors in the forthcoming (seventh) edition of the TNM classification of malignant tumors. *J Thorac Oncol.* 2007;2:603-12.
11. Postmus PE, Brambilla E, Chansky K, Crowley J, Goldstraw P, Patz EF, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the M descriptors in the forthcoming (seventh) edition of the TNM classification of malignant tumors. *J Thorac Oncol.* 2007;2:686-93.
12. Hofstetter W, Correa AM, Bekele N, Ajani JA, Phan A, Komaki RR, et al. Proposed modification of nodal status in AJCC esophageal cancer staging system. *Ann Thorac Surg.* 2007;84:365-75.
13. The International Adjuvant Lung Cancer Trial Collaborative Group. Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. *N Engl J Med.* 2004;350:351-60.
14. Douillard J, Rossel R, De Lena M, Carpagnano F, Ramlau R, González-Larriba JL, et al. Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB-IIIa non-small cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial. *Lancet Oncol.* 2006;7:719-27.
15. Rosell R, Gómez-Codina J, Camps C, Maestre J, Padille J, Cantó A, et al. A randomized trial comparing preoperative chemotherapy plus surgery with surgery alone in patients with non-small-cell lung cancer. *N Engl J Med.* 1994;330:153-8.

## Discussion

**Dr Bryan Fitch Meyers (St. Louis, Mo).** I congratulate Dr Kassis on his presentation. Because there were not any photographs of aortas or mitral valves, I am going to use some of my time to underscore some of his findings and elaborate a little bit, then follow up with two short questions.

First of all, the old system was created using 5000 patients, 90% of them from a single center, M. D. Anderson, where Dr Kassis is working right now. The new system was created using more than 100,000 patients collected worldwide, and it really is an international effort to represent uniform staging for patients with lung cancer. Twenty-eight thousand of those patients actually had a thoracotomy. This major revision has expanded the ability to make clear statements about staging.

If we assume that these proposed changes are meaningful and important changes based on 28,000 operated patients, then what conclusions would be drawn on the basis of the results of a confirmatory study like this? First, we would either be reassured or raise some doubt that the M. D. Anderson method of selecting patients and treating patients is consistent with those done in the rest of the world. If we assume that the M. D. Anderson system is representative of North America or American techniques, if these results were discrepant with the findings of the international group, we would wonder whether North America or America is distinct from other groups. Fortunately, your results are reassuring in that they confirm the recommendations by the international group.

It turns out that the findings that interest us most in a staging system are three things. What we want to see when one makes a change in a staging system, particularly a long awaited change like this, is that patients do shift from one stage to another. Dr Kassis has mentioned that 17% of the patients were shifted from one stage to another, and whether you use a *P* value or not, that is a substantial and clinically important change in the way that patients were assigned a stage. That number in itself tells us this was a meaningful change. The other aspects that are important are the distinctiveness of the stage groups and the heterogeneity within each stage group.

If you look at Dr Kassis' slide of the old system, with the stage IIA curve crossing over the IIB and then crossing over the IB, there is deficiency in the distinctiveness of the curves. However, if you move to the next slide where they applied the new staging system to their own data, you see that that distinctiveness has been improved with the new system. So, again, his presentation shows the superiority of the new system.

One area that was lacking in your presentation and discussion pertains to the problem of heterogeneity within each stage group. Is there a way that you could add to these results that reassures us that the patients who moved were moved from a group where they were less representative to a group that they now are more representative and homogeneous in their new stage group?

**Dr Kassis.** We did not definitively look at heterogeneity between stages, although that is a very good question and something that we could certainly evaluate at a future time. What we did do, though, is attempt to look at Kaplan–Meier curves in terms of stratifying these patients based on their ability to differentiate patients on the basis of stage. We took it one step further to try and do a statistical assessment of these patients by using a model called the permutation test to help quantify the differences that we see based on the Kaplan–Meier curve. However, your question regarding heterogeneity is something that we need to look at in the future.

**Dr Meyers.** The other question I had for you is that when we do a model, either a predictive or a descriptive model, and we want to validate it, we often use cases that were not used to create the model, and this was not mentioned in your presentation or the paper. Did the M. D. Anderson patients who were presented here play a role in the 28,000 operated patients who were used to create this new model? What are your thoughts on the impact of your answer on the importance of these results?

**Dr Kassis.** Less than 10% of our patient population of the 1154 patients was analyzed by the IASLC system, and the length of follow-up was such that in our estimation they are two completely separate data sets.

**Dr Frank C. Detterbeck** (*New Haven, Conn*). I just have a comment. I think that we need to be careful about what we

are trying to get from this. There are many purposes to a staging system. One is to have a nomenclature so that when one person is talking about certain patients in one study and another person is talking about them in another institution, the same group of patients are being discussed. Another is to determine prognosis, and that is clearly what was chosen as the primary goal in the IASLC staging project. And clearly it meets that goal. I think a third one, which is to select appropriate treatment for patients, is a bit of a slippery slope. That is not what the IASLC staging project was designed to do and it is not what staging systems are designed to do. Now, we use the language to help us talk about it, but it is really clinical trials that define what the appropriate treatment is for patients. We cannot just identify a stage (eg, stage II) and base treatment merely on that. We have to look at which patients we are talking about and what the clinical trials have shown us that we should be doing.

I think that you are taking the stage classification system to a different realm here than what it was intended to do, and I am not sure that that is really appropriate.

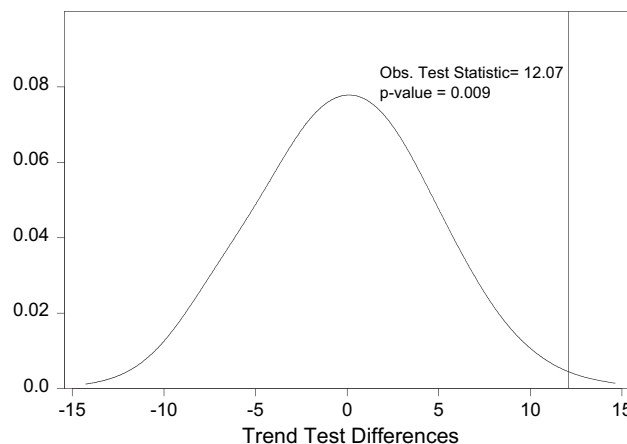
**Dr Kassis.** Thank you for your comments. I do agree with you. I do not think that we should be altering patient management on the basis of the stage shifting that we have demonstrated here. What I am trying to demonstrate is that the shifting of patients may lead to further studies so that we can better assess and better determine what to do with these patients with a satellite nodule in the ipsilateral lobe that was formerly T4 in stage IIIB disease. In the current system, if they are N0, they are going to be stage IIB; if they are N1, they are going to be stage IIIA. I think we need further studies to evaluate how we are going to manage these patients now that are assigned different stage groupings.

## APPENDIX

The main purpose of this section is to provide a formal statistical test for comparing two staging systems (the standard AJCC nodal staging system for esophageal cancer and a modified AJCC nodal staging system) with respect to assessing the discrimination between lower and higher stage disease and assessing the monotonic relationship between stage and survival. Before we can compare these two staging systems, we must first define what makes a staging system effective. We also must define how to quantify this effectiveness, and last, we must have a way of statistically comparing the effectiveness of the two staging systems.

The characteristic that defines the effectiveness of any staging system is its ability to differentiate, within a given disease, between patients with low-stage patients (those patients who survive a long time), mid-stage patients (patients who survive a moderate amount of time, and high-stage patients (patients who survive a relatively short amount of time). Thus an effective staging system is characterized by (1) the ability to discriminate between lower and higher stage patients with respect to survival and (2) a monotone decreasing relationship between stage and survival; this monotone relationship is quantified *graphically* by Kaplan–Meier curves. Moreover, this type of monotone relationship between stage and survival may be quantified numerically by a log–rank trend test.<sup>A1</sup> This statistic is used because it characterizes the effectiveness of a staging system as defined above: the more effective the staging system (ie, the stronger the relationship between stage and survival), the larger the value of the log–rank trend test statistic. Inasmuch as the log–rank trend test is available and can easily be calculated,<sup>A2</sup> we chose this statistic as a metric for measuring the strength of the association between stage and survival.

Once we have quantified the strength of the relationship between stage and survival of each of the two staging systems (ie, assessed the effectiveness of each staging system) using the log–rank trend test statistic, we also need to assess whether one staging system has a stronger relationship between stage and survival than the other staging system. The complicating factor in assessing the difference in effectiveness of the two staging systems is that the same set of patients are categorized under both systems inducing correlation between the two log–rank trend test statistics. We address this complicating factor by assessing differences between staging systems in the strength of the relationship between stage and survival via a permutation (randomization) test. A permutation test is a type of hypothesis test in which the null distribution is obtained by calculating possible values of the test statistic under random rearrangements (ie, permutations) of the original labels assigned to the observed data.<sup>A3</sup> By repeating this process many times (eg, 1000 times), we create a null distribution. Creation of a null distribution in this way only differs from null distribu-



**APPENDIX FIGURE 1.** Distribution of differences under the null distribution.

tions derived from statistical theory (eg, standard normal,  $\chi^2$ , F) in how the null distribution is obtained but does not differ in how they are used or interpreted. An added benefit of using null distributions derived from permutation tests is that they can be used in situations in which the null distribution is difficult to construct analytically (as in this case).

Each of the staging systems under consideration in this analysis has five staging categories. We call these five categories stage 1A, stage 1B, stage IIA, stage IIB, and stage IIIA. Under the null hypothesis for our permutation test, we assume that the two staging systems are exchangeable. This means that under the null hypothesis we assume that there is no difference between the two staging systems with respect to the strength of each staging system's relationship between stage and survival. The alternative hypothesis is that there is difference in the two staging systems with respect to the strength of the relationship between stage and survival. To assess these hypotheses, we construct our test statistic, which is the difference in the two trend tests calculated under each staging system. For the observed data, the difference in the two staging system log–rank test statistics is 12.07 (53.88 for the UICC-6 staging system and 65.95 for the IASLC staging system). These log–rank trend tests tell us that both methods show a strong relationship between stage and survival. However, it appears that the IASLC staging system is better inasmuch as the observed test statistic is larger for this staging system and larger test statistics imply stronger evidence that the null hypothesis (ie, no relationship between stage and survival) should be rejected. To *formally test* whether the IASLC staging system is indeed *statistically better* (and assess whether the differences in test statistics may only be due to chance), we construct our null distribution to which this observed test statistic will be compared by performing the following steps:

1. For each patient with 50% probability, we randomly rearrange (ie, permute) the staging system labels originally assigned to that patient. That is, for a given patient the stage



- assigned under the standard staging system is switched and the stage assignment is now considered to have been assigned under the modified system and vice versa).
2. Once all patients have been permuted, we calculate the log-rank trend test statistic for the two staging systems and record the permuted difference in log-rank trend tests.
  3. We repeat steps 1 and 2 a total of 1000 times.

The null distribution we construct using this method is given in [Appendix Figure 1](#).

As shown, the differences in the trend statistics under the null distribution are centered around 0 as one would expect

if there were no difference between staging systems in their ability to differentiate between low, middle, and high stage patients. Moreover, the probability of observing a difference in trend statistics is rare inasmuch as 12.07 is only 0.009 under the null hypothesis of no difference in the two staging systems.

### Appendix References

- A1. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. Berlin, New York: Springer-Verlag; 1997.
- A2. Cantor A. Extending SAS survival analysis techniques for medical research. Cary (NC): SAS Institute; 1997.
- A3. Efron B, Tibshirani R. An introduction to the bootstrap. London, New York: Chapman & Hall Ltd; 1993.