ORIGINAL ARTICLE

The IASLC Lung Cancer Staging Project: Methodology and Validation Used in the Development of Proposals for Revision of the Stage Classification of NSCLC in the Forthcoming (Eighth) Edition of the TNM Classification of Lung Cancer

Frank C. Detterbeck, MD,[a,*] Kari Chansky, MPH,[b] Patti Groome, PhD,[c] Vanessa Bolejack, MPH,[b] John Crowley, PhD,[b] Lynn Shemanski, PhD,[b] Catherine Kennedy, RMRA,[d] Mark Krasnik, MD,[e] Michael Peake, MD,[f] Ramón Rami-Porta, MD,[g] on behalf of the IASLC Staging and Prognostic Factors Committee, Advisory Boards, and Participating Institutions[**]

[a]*Department of Surgery, Yale University, New Haven, Connecticut*
[b]*Cancer Research And Biostatistics, Seattle, Washington*
[c]*Queen's Cancer Research Institute, Kingston, Ontario, Canada*
[d]*University of Sydney, Strathfield Private Hospital Campus, Strathfield, New South Wales, Australia*
[e]*Gentofte University Hospital, Copenhagen, Denmark*
[f]*University of Leicester, Leicester, United Kingdom*
[g]*Thoracic Surgery Service, Hospital Universitari Mutua Terrassa and Centros de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES) Lung Cancer Group, Terrassa, Barcelona, Spain*

Received 13 June 2016; revised 23 June 2016; accepted 24 June 2016
Available online - 21 July 2016

ABSTRACT

Introduction: Stage classification provides a consistent language to describe the anatomic extent of disease and is therefore a critical tool in caring for patients. The Staging and Prognostic Factors Committee of the International Association for the Study of Lung Cancer developed proposals for revision of the classification of lung cancer for the eighth edition of the tumor, node, and metastasis (TNM) classification, which takes effect in 2017.

Methods: An international database of 94,708 patients with lung cancer diagnosed in 1999–2010 was assembled. This article describes the process and statistical methods used to refine the lung cancer stage classification.

Results: Extensive analysis allowed definition of tumor, node, and metastasis categories and stage groupings that demonstrated consistent discrimination overall and within multiple different patient cohorts (e.g., clinical or pathologic stage, R0 or R-any resection status, geographic region). Additional analyses provided evidence of applicability over time, across a spectrum of geographic regions, histologic types, evaluative approaches, and follow-up intervals.

Conclusions: An extensive analysis has produced stage classification proposals for lung cancer with a robust degree of discriminatory consistency and general applicability. Nevertheless, external validation is encouraged to identify areas of strength and weakness; a sound validation should have discriminatory ability and be based on an independent data set of adequate size and sufficient follow-up with enough patients for each subgroup.

Keywords: Lung cancer; NSCLC; TNM classification; Lung cancer staging; Validation; Prognosis

*Corresponding author.

**See Appendix for the members of the International Association for the Study of Lung Cancer Staging and Prognostic Factors Committee, Advisory Boards, and Participating Institutions.

Disclosure: Dr. Peake reports personal fees from Roche Pharmaceuticals, Bristol-Myers Squibb Oncology, and Lilly Pharmaceuticals outside the submitted work. The remaining authors declare no conflict of interest.

Address for correspondence: Frank C. Detterbeck, MD, Yale University, 330 Cedar St., New Haven, CT 06520-8062. E-mail: frank.detterbeck@yale.edu

ISSN: 1556-0864
http://dx.doi.org/10.1016/j.jtho.2016.06.028

Introduction

Stage classification is a fundamental cornerstone in the management of patients with cancer. It provides a nomenclature to communicate effectively about the anatomic extent of the cancer. This is useful in many ways (e.g., grouping patients who have a similar extent of disease, providing a basis for reporting results and for comparing one treatment with another). Indeed, tumor stage is a prominent criterion for inclusion or exclusion in clinical trials. Knowing the extent of disease (and the selected treatment) has a major impact in predicting prognosis for a group of patients with similar characteristics (in addition to other patient, tumor, and environmental factors).[1,2] Thus, stage classification is used as an important tool in estimating prognosis, selecting treatment, and conducting and reporting on research studies. The ability to communicate clearly about the anatomic extent of disease is crucial to all of these.

The stage classification system is determined by the American Joint Committee on Cancer (AJCC) and Union for International Cancer Control (UICC). These are separate organizations, but they work collaboratively to define a single, consistent stage classification system for use around the world. This system is periodically updated, with the eighth edition taking effect on January 1, 2017. For lung cancer, the AJCC and UICC have relied primarily on the extensive work done by the Staging and Prognostic Factors Committee (SPFC) of the International Association for the Study of Lung Cancer (IASLC). The IASLC initiative provided a quantum leap forward for the seventh and eighth editions with regard to development, analysis, and validation of proposed stage classification elements (Fig. 1). This article describes the process and methodology used by the SPFC in developing proposals for the eighth edition of the stage classification for NSCLC. The actual tumor (T), node (N), and metastasis (M) stage grouping proposals, as well as related articles concerning tumor size measurement and the classification of multiple pulmonary sites of lung cancer, are described in specific articles.[3–11]

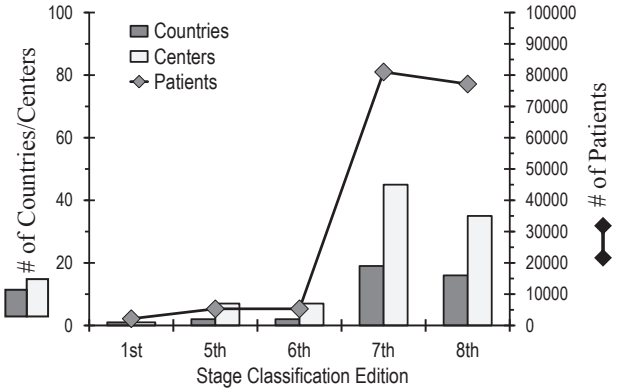

Figure 1. Features of data used for editions of the lung cancer stage classification system. A comparison of the number of cases, contributing institutions, and countries for various editions of the lung cancer stage classification system.

Patients and Methods

Process of Development of Proposals

Structure of the SPFC. The Lung Cancer Domain of the SPFC was divided into several subcommittees (Methodology, Tumor [T] Descriptors, Node [N] Descriptors, Metastasis [M] Descriptors, SCLC, Neuroendocrine Tumors, and Prognostic Factors) and ad hoc workgroups (T Coding and Size Measurement in Preinvasive and Lepidic Adenocarcinoma and Classification of Multiple Pulmonary Sites of Lung Cancer [see Appendix]). Each of these subcommittees were tasked with analyzing the data and developing proposals within their area. Annually, the chairs of the subcommittees, other members involved in data analysis, and the biostatisticians of Cancer Research And Biostatistics assembled for face-to-face meetings, and the entire SPFC and the advisory boards assembled every 2 years in conjunction with the World Conference on Lung Cancer meetings. Proposals for classification were first developed within the subcommittees and then presented and discussed at joint meetings. After revision the proposals were provided to the entire SPFC for further input and approval according to a formal process.

General Approach and Guiding Principles. Stage classification should coalesce tumors sharing similarities into groups that in turn should be distinct from one another. The definition of groups is based on the anatomic extent of the tumor (i.e., the characteristic the stage classification is designed to address). Of course, many other features of a tumor or patient (e.g., histologic type, genetic characteristics, age, sex, performance status, comorbidities, socioeconomic status) that may also be useful to classify patients for various purposes can be identified. However, stage classification relates only to the anatomic extent of disease.

On what basis should one consider that patients within a group are similar, or that groups are distinct from one another? There is no single "absolute" answer to this question, and how one addresses this depends on clinical judgment and viewpoint as much as it does on science. The SPFC chose to use prognosis of patients in the IASLC database as a tool to identify tumors that bear similarities and to distinguish groups. One must remember, however, that prognosis is determined by many factors (e.g., tumor-related, patient-related, environment-related, and treatment-related factors).[2]

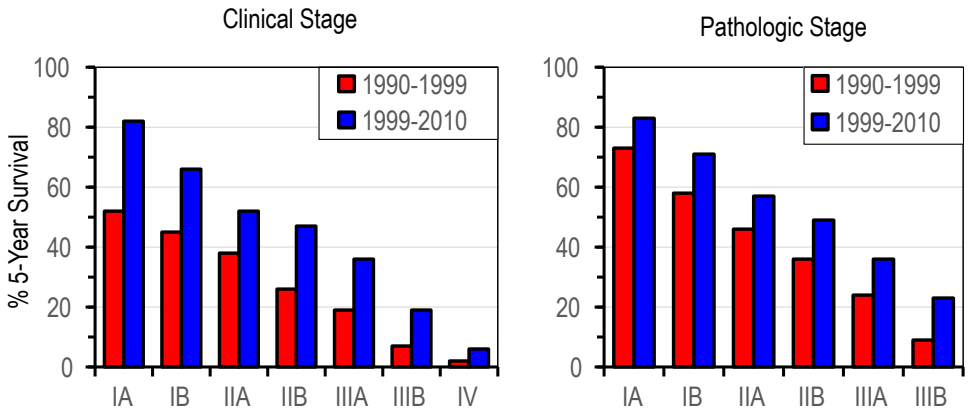

Figure 2. Comparison of survival in the 1990–1999 versus 1999–2000 data sets. Five-year overall survival (%) in the 1990–1999 versus 1999–2000 data sets (using the seventh-edition classification). The graph demonstrates that although discrimination is maintained in each data set, the actual prognosis (i.e., calibration) is markedly different.

Furthermore, prognosis varies substantially over time (e.g., comparing the 1999–2010 data set with the 1990–1999 data set [Fig. 2]), by geographic region of the institutions providing the source data, and by the source database type (Fig. 3*A–C*). Thus, prognosis per se could not be used; rather, we used the presence (or absence) of a *difference in prognosis* that was *consistent across multiple comparisons* (clinical, pathologic, R0, R-any, N0,

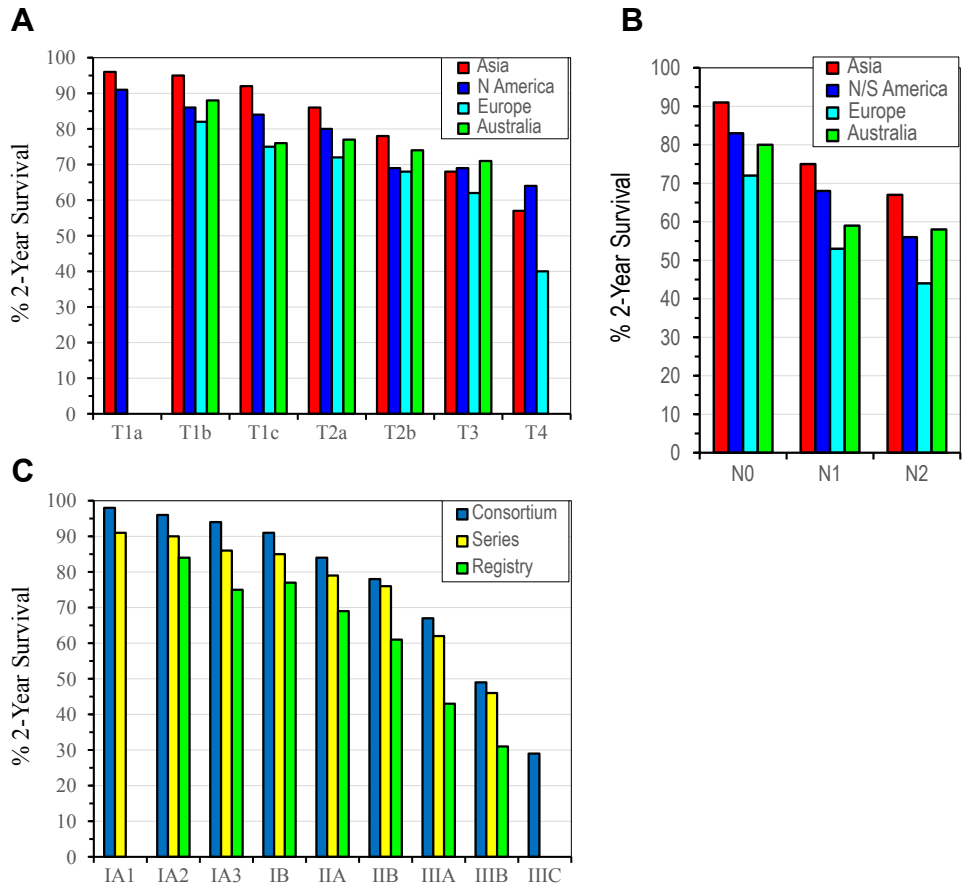

Figure 3. Comparison of survival by location of data contributors and type of source data. (*A*) Two-year overall survival by T category and location of data contributors (pathologic stage [p-stage] N-any M0 R-any tumors). (*B*) Two-year survival by N category and location of data contributors (p-stage T-any M0 R0 tumors). (*C*) Two-year survival by type of database and p-stage (p-stage T-any M0 R0 tumors). The graphs demonstrate that although discrimination is generally maintained, the actual prognosis (i.e., calibration) is markedly different. Results are for p-stage patients with tumors diagnosed in 1999–2010 using eighth-edition definitions. Categories with limited sample size (<50 patients) are omitted. N America, North America; N/S America, North and South America.

Table 1. Guiding Principles for the Development of Stage Classification Proposals

Descriptors should be applicable to both clinical and pathologic stage classification
Changes in T, N, and M categories should not compromise the use of data from the previous staging system whenever possible
Criteria for clinical T, N, and M categories should match those for pathologic T, N, and M categories
If boundaries between categories are to be changed, there should be overwhelming evidence to support such changes
Unproven descriptors should be "flagged" for further testing
Pathologic stage classification should include consideration of incompletely resected cases (R1, R2)
Prediction of prognosis associated with tumor stage should be supplemented by other prognostically important factors in a validated system (e.g., histologic diagnosis, age, sex, performance status, comorbidities)
Evidence from external sources should be taken into consideration.

T, tumor; N, node; M, metastasis.
Data from the principles used for development of the seventh-edition proposals for lung cancer.[12]

N-any, within a geographic region, histologic type, database type, etc.) along with clinical judgment to make decisions about stage classification changes.

The SPFC developed guiding principles for the development of proposals for the seventh edition of stage classification; these were carried forward for the eighth edition (Table 1) and used to guide decisions of the SPFC. The SPFC also considered additional aspects in developing classification proposals (e.g., statistical analysis of outcomes of patients in the IASLC database, degree of consistency across subgroups, practical aspects related to specific descriptors).

A goal of the SPFC was to conduct sufficient validation to be confident that the stage classification schema would function well. It should be emphasized that this refers to validation of the stage classification, meaning the ability to discriminate ordered groups of tumors according to anatomic extent of disease. Development of a prognostic model that forecasts outcomes (as opposed to sorting into ordered groups) is a different task, requiring inclusion of all relevant prognostic information and assessment of calibration (the accuracy of the prediction of the outcome in question). A framework for validation of stage classification is outlined in Table 2.[13]

Characteristics of the Database

The database used for analysis has been previously described.[14] This involved 94,708 patients whose disease was diagnosed between 1999 and 2010. After exclusions, 70,967 patients with NSCLC and 6189 with SCLC were available for analysis. This involved a worldwide collection of patients from 35 sources and 16 countries. Europe (49%) and Asia (44%) are strongly represented, whereas the proportions of the database coming from North America (5%), Australia (1.7%), and South America (0.3%) were much lower.[14] The yearly number of cases submitted was fairly even (~3000 per year) except for large contributions from Japan in 1999, 2002, and 2004, reflecting three initiatives to gather data from a national consortium of institutions recording surgical cases (38,928 cases) and a smaller bolus from Turkey in 2005 (3885 cases) (Supplementary Fig. 1). Most of the patients underwent an operation (with 57.7% receiving an operation alone and 27% receiving an operation in combination with chemotherapy and/or radiotherapy), 9.3% underwent chemotherapy alone, 4.7% chemotherapy and radiotherapy, and 1.5% radiotherapy alone.[14] The source of the cases was a consortium in 57% of cases (meaning a group of institutions that registered all cases) and a registry in 34%

Table 2. Criteria to Assess Stage Classification

Category	Characteristic	Explanation
Transportability (applicability)	Historical	Is discrimination maintained in populations from a different era than the one used to develop the system?
	Geographic	Is discrimination maintained in populations from a different geographic region than the ones used to develop the system?
	Methodologic	Is discrimination maintained in cohorts identified in different ways (e.g., CT vs. PET, c-stage vs. p-stage)?
	Spectrum	Is discrimination maintained in populations with a different spectrum of disease (e.g., different histotypes, smokers vs. nonsmokers, symptom detected vs. screen detected)?
	Follow-up interval	Is appropriate ordering of groups maintained across the follow-up spectrum (e.g., at 1, 2, 3, 5, and 10 years)?
Implementability	Simplicity	Is it simple, intuitive, easy, and practical to implement?
	Clarity	Is it defined in unambiguous terms?

CT, computed tomography; PET, positron emission tomography; c-stage, clinical stage; p-stage, pathologic stage.

(involving all cases within a region).[13] There were 40,263 clinical stage (c-stage) M0 cases and 36,830 pathologic stage (p-stage) M0 cases with complete T and N category information. Patients receiving preoperative therapy were excluded from p-stage analyses (i.e., yp-stage cases were excluded from p-stage cohorts).

An electronic data capture (EDC) system was established to collect detailed and complete data from participating institutions, thus avoiding the limitations of the retrospective data, which were not designed specifically for the purpose of defining stage classification and contained varying levels of detail and missing information. The EDC comprised 4667 cases from 23 institutions in 14 countries.[14] Although the EDC data set is much smaller, it has distinct advantages because it is much more detailed and complete. We used both the larger data set of available data from multiple sources and the EDC data to derive the proposed eighth edition of the classification.

Table 3 depicts the number of NSCLC cases available for analysis in various subsets (i.e., cases with sufficient detail in the relevant T, N, and M descriptors; R status; and other factors to allow performance of the major exploratory and validation analyses).

Statistical Analysis

We performed extensive survival analyses of patients in the IASLC database to inform the decisions regarding the best way to define the T, N, and M categories and stage groupings. Cases with missing data were excluded from the analysis. Because actual outcomes varied in association with many factors (database type, geographic region, time period, etc.) decisions were informed not by the prognosis per se but by differences in prognosis between categories and stage groupings that were consistent among many separate analyses of particular cohorts. Specifically, the SPFC considered the presence or absence of differences in prognosis between (heterogeneous) and within (homogeneous) categories and stage groupings and the consistency of those findings across separate analyses of particular cohorts. The SPFC also considered whether differences were clinically meaningful, because in a large data set statistical significance can sometimes be seen with trivial differences. Details of the cohorts considered and the analyses used to choose definitions of T, N, and M categories and stage groups varied by subcommittee, reflecting differences in the issues, the available data, and the value of particular assessments. These details are presented in the context of each subcommittee later in this article.

The statistical analysis was carried out by Cancer Research And Biostatistics, building on experience gained from development of the seventh edition of the stage classification. Cases used for the analyses summarized in this article were restricted to NSCLC; SCLC was evaluated separately as reported elsewhere.[15] Survival was measured from the date of diagnosis for c-stage tumors and the date of operation for p-stage tumors, and it was analyzed using Kaplan-Meier methods. Survival estimates were compared using the likelihood ratio test from Cox proportional hazards regression. The recursive partitioning and amalgamation (RPA) analysis was performed using the statistical package R, version 3.1.0 (R Foundation for Statistical Computing, Vienna, Austria). Cox regression analysis, adjusted for baseline factors (age, sex, region, and cell type) was performed using the proportional hazards regression procedure of SAS System for Windows, version 9.4 (SAS Institute Inc., Cary, NC). The discrimination of the classification was assessed primarily by using the R^2 measure. This measure was deemed to be the best metric to use, given the strengths and weaknesses of different measures and characteristics of the data (e.g., the number of categories, tied pair combinations of risk scores, censored cases).

Table 3. Non-Small Cell Cases Available for T, N, and M Category Analysis

Pathologic Stage	n	Clinical Stage	n
T component			
N0 M0 R0	21,133	N0 M0	10,230
N-any M0 R0	27,986	N-any M0	12,956
N0 M0 R-any	22,257		
N component			
T-any M0 R0	29,728	T-any M0	38,910
T1 M0 R0	12,824	T1 M0	15,681
T2 M0 R0	12,785	T2 M0	13,476
T3 M0 R0	3636	T3 M0	5004
T4 M0 R0	483	T4 M0	4729
T-any M0 R-any	31,426		
M component (nonsurgically managed)			
–	–	T-any N-any M1a	324
–	–	T-any N-any M1b	735
Stage grouping			
T-any N-any M0	31,936	T-any N-any M0	16595
T-any N-any M1	0	T-any N-any M1	882
Total	31,936	Total	17,477

Results

Evaluation of the T Component

T size cutpoints were chosen using a running log rank statistic,[16] initially in a p-stage N0 M0 R0 cohort, but then confirmed in a p-stage N0 M0 R-any, an N-any M0 R0, a c-stage N0 M0 and a c-stage N-any M0 cohort (NSCLC).[3] This confirmed previous size cutpoints identified in the seventh edition but also suggested cutpoints in 1-cm increments. Potential T-descriptor categories by size alone were then compared with those identified by

an isolated nonsize descriptor (e.g., visceral pleural invasion, pericardial invasion) using a multivariate Cox regression analysis that adjusted for age, sex, histologic type, and geographic region. Tumors with more than one positive descriptor within a T category were considered separately from those with only one positive descriptor. Descriptors that identified tumors with a prognosis similar to those of another group were combined, and those that demonstrated a significant difference were separated. This was done initially in a p-stage N0 M0 R0 cohort but confirmed in p-stage N0 M0 R-any and N-any M0 R0 cohorts and in c-stage N0 M0 and N-any M0 cohorts.[3]

Tumors with more than one positive T descriptor were compared with those with only 1 positive T descriptor in multiple ways, but the results were inconsistent (often differing between analyses done in different cohorts, especially p-stage and c-stage cohorts). Therefore, the number of positive T descriptors does not affect the T category—it is determined by the highest T as defined by any one descriptor. The final proposals for T descriptors are presented in the T-component article.[3]

Evaluation of the N Component

The discrimination of the N categories was first demonstrated in c-stage T-any M0 NSCLC cases, then confirmed in each T category.[4] The differences between the groups defined by each N category were highly statistically significant,[4] with the exception that significance was not demonstrated between the N0 and N1 categories in T3 and in T4 cases. The discriminatory ability of the N categories was also confirmed in p-stage cases, specifically in T-any M0 R-any and T-any M0 R0 cases (all differences being highly statistically significant).[4] Furthermore, within each p-stage T category (M0 R0), the N categories defined cohorts with progressively worse survival, although statistical significance was lacking between the N2 and N3 categories because of a limited number of p-stage N3 M0 R0 tumors.[4]

Patients with sufficient detail to be evaluable for the N-component analysis were largely contributed from Japan. Furthermore, there were differences in survival between different regions (Fig. 2).[4] However, geographic applicability of the N categories was demonstrated in separate comparisons within geographic regions (Asia, North and South America, Europe, and Australia).[4] Further details of the N-component analysis, including exploratory analyses related to the number of positive nodes and nodal stations, is described in the N-component article.[4]

Evaluation of the M Component

The analysis of the M component involved data on 1059 patients with c-stage M1 NSCLC that was not surgically resected; the data were collected using the EDC system in 2000–2012 (because only these cases had enough detail for analysis). There was no difference in survival among different M1a descriptors (i.e., pleural/pericardial nodules, pleural/pericardial effusion, contralateral/bilateral pulmonary nodules) or between single or multiple M1a descriptors (in c-stage T-any N-any M1a cases). Therefore, these descriptors were retained within the M1a category, just as in the seventh edition. Among patients with distant (extrathoracic) metastases, prognosis did not differ according to the site of metastasis in patients with a single metastasis or in patients with metastases in only a single organ site. However, there was a significant survival difference between patients with a single metastasis, with multiple metastases in a single organ, and with multiple metastases in multiple organs.[5] Confirmation of any of these findings within subsets (e.g., geographic regions, treatment groups) was not performed, and comparison with an M0 cohort was also not done.[5]

Evaluation of Stage Grouping

Candidate stage grouping schemes were developed using a recursive partitioning and amalgamation (RPA) algorithm[17] based on survival in a training set (a randomly selected sample of two-thirds of the cases stratified by type of data submission and time period of case entry [1999–2004 versus 2005–2010]). The RPA algorithm generates tree-based models using survival data and log-rank test statistics to determine best-split points on the basis of the newly proposed ordered T and N categories. M1 cases were also split into a training and validation set but were not the focus of the tree-based analysis.

Initial candidate stage grouping schemes were developed using the training set of M0 cases and best stage. Candidate groupings of the training set, with M1 cases added, were then evaluated using adjusted Cox regression analysis to assess whether there was an ordered, progressive decrease in survival and meaningful separation between the groups. The candidate staging schemes were also evaluated with respect to their clinical relevance and practical use. These groupings were then internally validated in the validation set as well as in larger cohorts involving all clinically staged or pathologically staged tumors (including M1 tumors and both the training and validation sets).

Survival was assessed for clinical, pathologic, and best stage (i.e., pathologic if available, otherwise clinical), using adjusted survival curves (adjusted using inverse probability weights reflecting the proportion of cases that were from registry databases [versus others] in each stage category).[18,19] This method was used to adjust for the different survival in registry databases in

general and the varying proportion of registry cases among the stage groups. Contrasts between adjacent stage groups were evaluated by Cox regression analysis, adjusted for baseline factors (age, sex, performance status, region, and cell type) and type of database submission. Assessment of candidate stage groupings was performed using the R^2 statistic, which can be viewed as the percent of variance of OS explained by the model.

The final stage grouping proposal was selected by consensus from among the candidate schemes on the basis of its statistical properties and its relevance to clinical practice. In the final eighth edition of the classification scheme the R^2 statistic (adjusted for database type) for c-stage is 65.5 and for p-stage is 41.6 (versus 64.6 and 39.8, respectively, for the seventh edition of the classification scheme).[6] With the inclusion of covariates of age, sex, and histologic type this improves slightly to 68.3 for c-stage and 46.9 for p-stage (versus 67.5 and 45.7, respectively, for the seventh edition of the classification scheme). The maximum possible R^2 score is 100 (i.e., a perfect model).

Validation

Internal validation of the proposed stage classification was conducted to assess the consistency of the proposed stage classification across meaningful subgroups. This validation was focused on the discriminatory ability of the stage classification, meaning the ability to discriminate ordered groups of tumors according to anatomic extent of disease. The type of validation and the degree to which it was done (e.g., the number of subgroup analyses) varied among the T, N, and M components and stage grouping. This was determined largely by what was possible from the available data but was also influenced by what was deemed necessary by the SPFC to establish reasonable confidence in the proposed classification scheme. An overview of what was done is provided in Table 4.

Geographic applicability is crucial for a worldwide classification system. This has been demonstrated for the T and N components and the stage grouping through internal validation for broad geographic regions in both the seventh- and eighth-edition classification proposals. Because of a limited number of patients with sufficient detail for the M component it was not possible to confirm geographic transportability of the M classification. Further analysis by more finely parsed regions, particular countries, or regions defined by the degree of economic or health care system development was considered but was not possible.

The stage classification system should be useful over a period of years until the next edition is developed. This underscores the importance of historical transportability, which is further highlighted by the fact that the current proposal is based on patients whose disease was diagnosed between 1999 and 2010. Historical applicability has been demonstrated for the N categories, and for most of the T categories and descriptors as well, as evidenced by similar discriminatory impact in the 1990–1999 and 1999–2010 data sets. It was not possible to assess this for the M component because of insufficient detail in prior data. We tested for historical transportability by applying the new (eighth-edition) stage classification to the 1990–1999 data set (Supplementary Fig. 2). In general, good discrimination was demonstrated for the new (eighth-edition) classification.

Table 4. Applicability Assessments of the Eighth Edition of the Stage Classification System

	T	N	M	Groups
Historical	Yes[a]	Yes	No	Yes
Geographic	Yes	Yes	No	Yes
Methodologic[b]	Yes	Yes	No	Yes
Spectrum[c]	Yes	Yes	Yes	Yes
Follow-up interval	Yes	Yes	Yes	Yes
Implementability	No	No	No	No

[a]Done for those descriptors that have not changed between the seventh and eighth editions.
[b]Done for c-stage versus p-stage.
[c]Accomplished for SCLC versus NSCLC, and for NSCLC histotypes.

We can be fairly confident that the stage classification system has methodologic transportability, meaning that it applies to patients evaluated in different ways. The analyses underlying the eighth-edition proposals have included validation that the proposals apply to tumors staged clinically as well as pathologically. It is not possible with the current IASLC data to evaluate methodologic transportability for further details such as the use or nonuse of positron emission tomography. However, the historic and geographic transportability surely also reflect differences in the use of various staging tools.

Spectrum transportability refers to applicability across a spectrum of tumor biology/aggressiveness. Internal validation has demonstrated that the system works well for SCLC and for NSCLC,[15] which can be expected to be most likely to show a discrepancy. Neuroendocrine (i.e., carcinoid) tumors are on the other end of the spectrum; testing of the eighth edition in these tumors is planned but not yet completed. However, the fact that it was demonstrated that the seventh-edition schema was applicable to neuroendocrine tumors[20] provides a level of confidence that this will again be demonstrated. Spectrum transportability for stage groups has also been demonstrated by an internal analysis comparing adenocarcinoma versus squamous carcinoma versus other NSCLC subtypes (Supplementary Fig. 3).

Qualitative demonstration of applicability across a range of follow-up intervals is demonstrated by the survival curves in the various analyses; these do not show areas of superiority of groups that subsequently cross over to being inferior. It was not deemed necessary to carry out separate statistical analyses at various time points.

Implementability, meaning the degree to which the system can be consistently used, cannot be evaluated until it is implemented. The seventh edition of the system was demonstrated to be implemented and function appropriately in general, as shown by many external validation studies[21–35] as well as by application of the seventh-edition scheme to the 1999–2010 data set.[3,4,6] However, increasing complexity could create confusion. Assessment of the consistency of classification would likely require an audit of cases that assesses how closely the eighth edition of the TNM classification recorded in actual practice matches what should have been done upon independent review; such an assessment is outside of the capability of the SPFC.

Discussion

This article summarizes the principles, considerations and statistical methods used by the SPFC in the development of proposals for the eighth edition of the TNM classification of lung cancer. This involved an international multispecialty committee of lung cancer experts and a sophisticated defined process. Multiple subcommittees considered many detailed aspects. A large worldwide database of almost 100,000 cases was analyzed to guide decisions in conjunction with practical, clinical, and therapeutic considerations. This analysis was extensive, multifaceted, and detailed, with multiple assessments for consistency of the discriminatory findings across different subgroups defined by clinical status, region, database type, and time—thus providing a measure of internal validity. This multiyear project built on knowledge and experience gained from the seventh edition of the lung cancer stage classification. The process involved multiple layers of critical review before being submitted to the UICC/AJCC for consideration and published for review by the entire medical community. These features represent strengths that are unparalleled in the development of stage classification schemes of malignant tumors.

Limitations

Despite the size and global nature of the stage classification database, it remains largely a collection of available retrospective data. The process of collecting data involved extensive review, clarifications, and cleaning of the data but was limited to institutions or organizations that were willing and able to participate and by what was available to contribute. Certain regions are either underrepresented (e.g., North America) or not represented (e.g., Africa). Conversely, some areas are overrepresented, most notably Japan. Cases managed nonsurgically are underrepresented. The amount of detail that was available from different sources varied. Despite the size of the database, certain potential tumor descriptors were collected too infrequently to permit a robust analysis of their impact.

External validation of the proposed eighth edition of the classification is desirable, ideally involving a large independent data set that has all of the detail addressed by the eighth-edition proposal. However, larger databases often lack the level of detail, and smaller databases have the limitations imposed by their size. Nevertheless, the thoracic community is encouraged to carry out external validation wherever possible in a scientifically sound manner. External validation using the Surveillance, Epidemiology, and End Results database is not possible owing to lack of sufficient detail; however, the SPFC is conducting external validation using the U.S. National Cancer Data Base; the results of this analysis will be discussed in a separate manuscript.

Criteria to Assess Stage Classification

Many people are stimulated to investigate particular factors, which often were not addressed in the stage classification proposal, in their own database. There is certainly value in analyses of institutional data, particularly when details that may not be generally available are accessible. The UICC specifically tracks and considers such publications.[36] However, (isolated) evaluation of a particular factor must be viewed as hypothesis generating. The discriminatory value of the characteristic is unclear if it has not been subjected to confirmation in multiple patient subgroups, institutions, and regions, as was done for the stage classification system. Furthermore, an exploratory evaluation of multiple factors in a data set is associated with a high rate of false-positive findings (i.e., a factor that will not hold up in independent assessment in another data set) if the number of events (i.e., patients experiencing the outcome of interest) per variables examined is less than 10 to 20.[37,38] Investigators should keep these points in mind and resist the temptation to suggest new classification schemes that overstep what is possible to conclude from exploratory analyses of institutional data sets.

The stage classification system has also motivated individuals to investigate how well the system works in independent databases. The SPFC welcomes well-done external validation studies. As with scientific hypothesis, the generalizability of a system is established by

being tested in diverse settings. The more numerous and diverse the settings in which the system is tested and found to discriminate well, the more likely it will generalize to an untested setting. To promote application of a high level of science to such initiatives, it is worth discussing concepts fundamental to external validation.

Requirements for Robust External Validation

An external validation study should assess the discriminatory ability of the stage classification system (not prediction of prognosis). Discrimination denotes the ability of the system to identify distinct groups with an ordered relationship to one another—this is the goal of the stage classification system. Prognosis depends on many factors beyond the anatomic extent of disease, and it varies by region, treatment approach, the extent of testing to define the stage, the time period, the type of database, and many other patient-related (e.g., age, comorbidities), tumor-related (e.g., histotype, molecular characterization), and environment-related factors. The variability in prognosis across time periods and among regions within the IASLC database underscores this (Figs. 2 and 3).

An appropriate validation study should involve an independent data set with sufficient size. This means that it should involve few patients (we suggest $<10\%$) who were also included in the IASLC database, and should involve enough patients and have sufficient follow-up to permit a robust assessment of relative outcomes. A rule of thumb for multivariate models is that there should be at least 10 to 20 events per variable.[37,39,40] Studies specifically evaluating external validation suggest that a minimum of 100 to 200 events are needed.[41–43] It is useful for the investigators to be clear about what is being tested (i.e., geographic, historic, methodologic, and spectrum transportability [Table 2]). Assessment of any type of transportability can be useful to define areas of strength or weakness of the stage classification system, provided the assessment is scientifically sound. Although formal standards have been established for reporting on prognostic risk prediction modeling studies,[43] such standards for validation of a classification scheme have not been developed. To better assess the scientific strength of an external validation study, we suggest that such studies explicitly report on the points outlined in Table 5.

Difference between Stage Classification and Prognosis

Stage classification is a fundamental tool in the management of patients with malignant tumors. Prediction of prognosis is also a fundamental desire of physicians and patients. However, it is important to recognize that stage classification and prediction of prognosis are two fundamentally different processes.[1]

Stage classification is a nomenclature describing the anatomic extent of a cancer. Stage classification is inherently concrete and is designed to apply to an individual and to consistently produce the same stage assignment given the same staging information. It must be applicable and consistent throughout the world, and it must remain relatively static (only periodic refinement as for the eighth edition is allowed). It should cluster similar patients and separate groups that are distinct from one another. Although prognosis (actually, it was consistent differences in prognosis among multiple analyses) was used as a tool by the SPFC when considering how to coalesce patients into groups, the goal of the classification scheme is to provide a simple, consistent, and stable language to describe similar groups of tumors.

Prognosis, on the other hand, is dependent on many factors (e.g., age, comorbidities, whether and how the patient is treated, genetic features of the tumor). Prognosis is inherently fluid: it is constantly changing as treatment advances are made; even for a specific patient it is constantly changing (i.e., it is different at the point of diagnosis, upon successful completion of treatment, a year after treatment, etc.). Thus, although the anatomic extent of disease is a very important aspect influencing prognosis, actual prognosis is dependent on many factors that are inherently fluid and constantly changing. An appropriate, clinically useful prognostic model must take all of the factors that have a major impact into account, must be sufficiently detailed to apply to a specific patient, and must have a flexible structure that provides the most up-to-date prediction possible.

Although the AJCC and UICC have long worked to have exactly matching stage classification, terms, and definitions in their manuals, there is a discrepancy of terms with the eighth edition. The AJCC eighth edition will use the singular term *prognostic stage groups* to describe the grouping system regardless of whether it is solely anatomic or also includes nonanatomic factors in the calculation. The UICC will use the term *stage groups* to refer to stage classification based strictly on anatomic factors and prognostic stage groups for a separate classification that includes nonanatomic factors for those disease sites in which prognostic groups have been defined by the inclusion of nonanatomic factors.

Appropriate Use of Prognosis from the IASLC Database

There is a great temptation to apply the prognosis shown by the survival graphs to patients with a particular stage. It is important to recognize the limitations inherent to this exercise. The IASLC survival curves represent a global average of available cases across the world that

Table 5. Suggested Factors to Report in an External Validation Study

Factor	Detail
Independent cases	Cases not already included in IASLC database
Sample size	No. cases and events (total and per analyzed subgroup)
Follow-up	Median f/u (till death or last known status); no. lost to f/u
Ordering	What is the order of subgroups examined?
Discrimination	Degree of statistical difference between cohorts
Consistency	Were findings consistent in different cohorts (e.g., c-stage, p-stage, N0, N-any, R0, R-any cohorts)?
Homogeneity	Degree of variability with an analyzed subgroup
Sample size limitation	In event of lack of a difference between subgroups, what is the power to detect a clinically meaningful difference?
Confounding	Was a multivariate analysis done to assess whether an observed difference was independently associated with the factor of interest?

IASLC, International Association for the Study of Lung Cancer; No., number; f/u, follow-up; c-stage, clinical stage; p-stage, pathologic stage.

were diagnosed between 1999 and 2010 and treated in a variety of ways. How well this average applies to a specific patient whose disease is diagnosed today within a particular region and treated in a specific way is questionable at best. Physicians (and patients) want a prediction that is applicable to a particular patient, not a general average from a decade ago. Furthermore, the prediction of prognosis is inherently associated with a degree of uncertainty due to the fact that we can never predict everything that may happen in the future.

A measure of the ability of the stage classification to predict prognosis is the R^2 statistic, which is also referred to as the percent of variance explained (PVE). This can be viewed as a measure of the degree to which the observed outcomes are predicted by the model (in this case the stage classification). For c-stage alone, the PVE for the proposed eighth edition is 65.5; for p-stage, the PVE is 41.6 (perfect prediction of prognosis would be 100).[6] This can be interpreted as demonstrating that anatomic stage explains approximately two-thirds of the variability of actual prognosis in c-stage and less than half in p-stage. If additional covariates of age, sex, and histologic type are added, the PVE increases only slightly to 68.3 and 46.9 in the IASLC data set for c-stage and p-stage, respectively. The higher PVE in c-stage may result from the fact that this cohort contains a much higher proportion of advanced-stage tumors and patients treated palliatively (i.e., groups in which the tumor is a strong driver of outcome). The p-stage cohort contains more early-stage tumors and curatively treated patients, in which prognosis is primarily determined by other factors such as age and comorbidities. This highlights that the extent to which the anatomic extent of disease can determine prognosis is limited and varies according to the tumor stage, the treatment given, and other factors.

The survival outcomes as observed in the SPFC analysis can therefore be appropriately viewed as only a general indicator of prognosis. Applying this to patients in the current clinical setting requires clinical judgment, factoring in general trends over time, by region, and by treatment. Making a specific estimate for an individual patient should further take into account patient-related factors (e.g., comorbidities) and, if possible, tumor-related factors and treatment-related factors (e.g., treatment response). At this point, this must be done qualitatively (i.e., by "clinical judgment," because the development of validated quantitative prognostic models is rudimentary).[44] Finally, it must be accepted that making a prediction into the future for a patient whose disease is diagnosed or treated today involves a substantial amount of inherent uncertainty, as opposed to a (retrospective) model that involves previous patients in whom outcomes have already occurred.

Conclusion

A universal language to describe the anatomic extent of disease is a critical cornerstone in managing patients. Refinements for the eighth edition of the TNM classification of lung cancer have been developed by using an international database of nearly 100,000 patients and an extensive multifaceted analysis. This analysis has demonstrated consistent discriminatory ability in multiple subset analyses, and it provides an assessment of applicability across time periods, geographic regions, histologic types, methods of stage evaluation, and follow-up. Nevertheless, scientifically sound external evaluation by the broader medical community is welcome.

Appendix

IASLC Staging and Prognostic Factors Committee

Peter Goldstraw, past chair, Royal Brompton Hospital and Imperial College, London, United Kingdom; Ramón Rami-Porta, chair, Hospital Universitari Mutua Terrassa, Terrassa, Spain; Hisao Asamura, chair-elect, Keio University, Tokyo, Japan; David Ball, Peter MacCallum Cancer Centre, Melbourne, Australia; David G. Beer, University of Michigan, Ann Arbor, Michigan; Ricardo

Beyruti, University of Sao Paulo, Sao Paulo, Brazil; Vanessa Bolejack, Cancer Research And Biostatistics, Seattle, Washington; Kari Chansky, Cancer Research And Biostatistics, Seattle, Washington; John Crowley, Cancer Research And Biostatistics, Seattle, Washington; Frank Detterbeck, Yale University, New Haven, Connecticut; Wilfried Ernst Erich Eberhardt, West German Cancer Centre, University Hospital, Ruhrlandklinik, University Duisburg-Essen, Essen, Germany; John Edwards, Northern General Hospital, Sheffield, United Kingdom; Françoise Galateau-Sallé, Centre Hospitalier Universitaire, Caen, France; Dorothy Giroux, Cancer Research And Biostatistics, Seattle, Washington; Fergus Gleeson, Churchill Hospital, Oxford, United Kingdom; Patti Groome, Queen's Cancer Research Institute, Kingston, Ontario, Canada; James Huang, Memorial Sloan-Kettering Cancer Center, New York, New York; Catherine Kennedy, University of Sydney, Strathfield Private Hospital Campus, Strathfield, New South Wales, Australia; Jhingook Kim, Samsung Medical Center, Seoul, Republic of Korea; Young Tae Kim, Seoul National University, Seoul, Republic of Korea; Laura Kingsbury, Cancer Research And Biostatistics, Seattle, Washington; Haruhiko Kondo, Kyorin University Hospital, Tokyo, Japan; Mark Krasnik, Gentofte Hospital, Copenhagen, Denmark; Kaoru Kubota, Nippon Medical School Hospital, Tokyo, Japan; Antoon Lerut, University Hospitals, Leuven, Belgium; Gustavo Lyons, British Hospital, Buenos Aires, Argentina; Mirella Marino, Regina Elena National Cancer Institute, Rome, Italy; Edith M. Marom, M. D. Anderson Cancer Center, Houston, Texas; Jan van Meerbeeck, Antwerp University Hospital, Edegem, Belgium; Alan Mitchell, Cancer Research And Biostatistics, Seattle, Washington; Takashi Nakano, Hyogo College of Medicine, Hyogo, Japan; Andrew G. Nicholson, Royal Brompton and Harefield National Health Service Foundation Trust and Imperial College, London, United Kingdom; Anna Nowak, University of Western Australia, Perth, Australia; Michael Peake, University of Leicester, United Kingdom; Thomas Rice, Cleveland Clinic, Cleveland, Ohio; Kenneth Rosenzweig, Mount Sinai Hospital, New York, New York; Enrico Ruffini, University of Torino, Torino, Italy; Valerie Rusch, Memorial Sloan-Kettering Cancer Center, New York, New York; Nagahiro Saijo, National Cancer Center Hospital East, Chiba, Japan; Paul Van Schil, Antwerp University Hospital, Edegem, Belgium; Jean-Paul Sculier, Institut Jules Bordet, Brussels, Belgium; Lynn Shemanski, Cancer Research And Biostatistics, Seattle, Washington; Kelly Stratton, Cancer Research And Biostatistics, Seattle, Washington; Kenji Suzuki, Juntendo University, Tokyo, Japan; Yuji Tachimori, National Cancer Center, Tokyo, Japan; Charles F. Thomas, Jr., Mayo Clinic, Rochester, Minnesota; William Travis, Memorial Sloan-Kettering Cancer Center, New York, New York; Ming S. Tsao, The Princess Margaret Cancer Centre, Toronto, Ontario, Canada; Andrew Turrisi, Sinai Grace Hospital, Detroit, Michigan; Johan Vansteenkiste, University Hospitals, Leuven, Belgium; Hirokazu Watanabe, National Cancer Center Hospital, Tokyo, Japan; and Yi-Long Wu, Guangdong Lung Cancer Institute, Guangdong General Hospital, Guangzhou, People's Republic of China.

Advisory Board of the IASLC Mesothelioma Domain

Paul Baas, The Netherlands Cancer Institute, Amsterdam, The Netherlands; Jeremy Erasmus, M. D. Anderson Cancer Center, Houston, Texas; Seiki Hasegawa, Hyogo College of Medicine, Hyogo, Japan; Kouki Inai, Hiroshima University Postgraduate School, Hiroshima, Japan; Kemp Kernstine, University of Texas Southwestern Medical Center, Dallas Texas; Hedy Kindler, The University of Chicago Medical Center, Chicago, Illinois; Lee Krug, Memorial Sloan-Kettering Cancer Center, New York, New York; Kristiaan Nackaerts, University Hospitals, Leuven, Belgium; Harvey Pass, New York University, New York, New York; and David Rice, M. D. Anderson Cancer Center, Houston, Texas.

Advisory Board of the IASLC Thymic Malignancies Domain

Conrad Falkson, Queen's University, Ontario, Canada; Pier Luigi Filosso, University of Torino, Italy; Giuseppe Giaccone, Georgetown University, Washington, District of Columbia; Kazuya Kondo, University of Tokushima, Tokushima, Japan; Marco Lucchi, University of Pisa, Pisa, Italy; and Meinoshin Okumura, Osaka University, Osaka, Japan.

Advisory Board of the IASLC Esophageal Cancer Domain

Eugene Blackstone, Cleveland Clinic, Cleveland, Ohio.

Participating Institutions in the New IASLC Lung Cancer Staging Project

F. Abad Cavaco and E. Ansótegui Barrera, Hospital La Fe, Valencia, Spain; J. Abal Arca and I. Parente Lamelas, Complejo Hospitalario de Ourense, Ourense, Spain; A. Arnau Obrer and R. Guijarro Jorge, Hospital General Universitario de Valencia, Valencia, Spain; D. Ball, Peter MacCallum Cancer Centre, Melbourne, Australia; G. K. Bascom, Good Samaritan Hospital, Kearney, Nebraska; A. I. Blanco Orozco and M. A. González Castro, Hospital Virgen del Rocío, Seville, Spain; M. G. Blum, Penrose Cancer Center, Colorado Springs, Colorado; D. Chimondeguy, Hospital Universitario Austral, Argentina; V. Cvijanovic, Military Medical

Academy, Belgrade, Serbia; S. Defranchi, Hospital Universitario-Fundacion Favaloro, Buenos Aires, Argentina; B. de Olaiz Navarro, Hospital de Getafe, Getafe, Spain; I. Escobar Campuzano and I. Macía Vidueira, Hospital de Bellvitge, L'Hospitalet de Llobregat, Spain; E. Fernández Araujo and F. Andreo García, Hospital Universitari Germans Trias i Pujol, Badalona, Spain; K. M. Fong, Prince Charles Hospital, Brisbane, Australia; G. Francisco Corral and S. Cerezo González, Hospital La Mancha Centro, Ciudad Real, Spain; J. Freixinet Gilart, Hospital Universitario 'Dr. Negrín', Las Palmas de Gran Canaria, Spain; L. García Arangüena, Hospital Sierrallana, Torrelavega, Spain; S. García Barajas, Hospital Infanta Cristina, Badajoz, Spain; P. Girard, L'Institut Mutualiste Montsouris, Paris, France; T. Goksel, Turkish Thoracic Society, Turkey; M. T. González Budiño, Hospital General Universitario de Oviedo, Oviedo, Spain; G. González Casaurrán, Hospital Gregorio Marañón, Madrid, Spain; J. A. Gullón Blanco, Hospital San Agustín, Avilés, Spain; J. Hernández Hernández, Hospital de Ávila, Avila, Spain; H. Hernández Rodríguez, Hospital Universitario de Tenerife, Santa Cruz de Tenerife, Spain; J. Herrero Collantes, Hospital Universitario Nuestra Señora de la Candelaria, Santa Cruz de Tenerife, Spain; M. Iglesias Heras, Hospital de Ávila, Ávila, Spain; J. M. Izquierdo Elena, Hospital Nuestra Señora de Aránzazu, Donostia, Spain; E. Jakobsen, Danish Lung Cancer Registry, Denmark; S. Kostas, Athens School of Medicine, Athens, Greece; P. León Atance and A. Núñez Ares, Complejo Hospitalario de Albacete, Albacete, Spain; M. Liao, Shanghai Lung Tumor Clinical Medical Center, Shanghai, People's Republic of China; M. Losanovscky, Clinica y Maternidad Suizo Argentina, Buenos Aires, Argentina; G. Lyons, Hospital Britanico de Buenos Aires, Buenos Aires, Argentina; R. Magaroles and L. De Esteban Júlvez, Hospital Joan XXIII, Tarragona. Spain; M. Mariñán Gorospe, Hospital de San Pedro de Logroño, Logroño, Spain; B. McCaughan and C. Kennedy, University of Sydney, Sydney, Australia; R. Melchor Íñiguez, Fundación Jiménez Díaz, Madrid, Spain; L. Miravet Sorribes, Hospital La Plana, Castellón, Spain; S. Naranjo Gozalo and C. Álvarez de Arriba, Hospital Universitario Marqués de Valdecilla, Santander, Spain; M. Núñez Delgado, Hospital de Meixoeiro, Vigo, Spain; J. Padilla Alarcón and J. C. Peñalver Cuesta, Instituto Valenciano de Oncología, Valencia, Spain; J. S. Park, Samsung Medical Center, Seoul, Republic of Korea; Harvey Pass, New York University Langone Medical Center and Cancer Center, New York, New York; M. J. Pavón Fernández, Hospital 'Severo Ochoa', Leganés, Spain; M. Rosenberg, Alexander Fleming Institute and Hospital de Rehabilitación Respiratoria, Buenos Aires, Argentina; E. Ruffini, University of Torino, Torino, Italy; V. Rusch, Memorial Sloan-Kettering Cancer Center, New York, New York; J. Sánchez de Cos Escuín, Hospital de Cáceres, Cáceres, Spain; A. Saura Vinuesa, Hospital de Sagunto, Sagunto, Spain; M. Serra Mitjans, Hospital Universitari Mutua Terrassa, Terrassa, Spain; T.E. Strand, Cancer Registry of Norway, Norway; D. Subotic, Clinical Centre of Serbia, Belgrade, Serbia; S. Swisher, M. D. Anderson Cancer Center, Houston, Texas; R. Terra, University of Sao Paulo Medical Center, Sao Paulo, Brazil; C. Thomas, Mayo Clinic Rochester, Rochester, Minnesota; K. Tournoy, Ghent University Hospital, Ghent, Belgium; P. Van Schil, Antwerp University Hospital, Edegem, Belgium; M. Velasquez, Fundacion Clinica Valle del Lili, Cali, Colombia; Y. L. Wu, Guangdong General Hospital, Guangzhou, People's Republic of China; and K. Yokoi, Japanese Joint Committee for Lung Cancer Registry, Osaka, Japan.

Lung Subcommittees

T Descriptors Subcommittee. Ramon Rami-Porta (chair), David Ball, Vanessa Bolejack, John Crowley, Dorothy J. Giroux, Jhingook Kim, Gustavo Lyons, Thomas Rice, Kenji Suzuki, Charles F Thomas, Jr., William D Travis, and Yi-Long Wu.

N Descriptors Subcommittee. Hisao Asamura (chair), David Ball, Kari Chansky, John Crowley, Peter Goldstraw, Valerie Rusch, Paul van Schil, Johan Vansteenkiste, Hirokazu Watanabe, Yi-Long Wu, and Marcin Zielinski.

M Descriptors Subcommittee. Wilfried Eberhardt (chair), Kari Chansky, John Crowley, Young Tae Kim, Haruhiko Kondo, Alan Mitchell, and Andrew Turrisi.

Validation and Methodology Subcommittee. Patti Groome (chair 2010–2015), Frank Detterbeck (chair 2015–2017), Vanessa Bolejack, John Crowley, Catherine Kennedy, Mark Krasnik, and Michael Peake.

Prognostic Factors Subcommittee. Jean-Paul Sculier (chair), Kari Chansky, John Crowley, Dorothy J. Giroux, Fergus Gleeson, and Jan van Meeerbeeck.

Neuroendocrine Tumors Subcommittee. William D. Travis (chair), Hisao Asamura, Kari Chansky, John Crowley, and Dorothy J. Giroux.

SCLC Subcommittee. Andrew Nicholson (chair), Ricardo Beyruti, Kari Chansky, John Crowley, Kouru Kubota, and Andrew Turrisi.

Biologic Factors Subcommittee. Ming S. Tsao (chair), David G. Beer, John Crowley, and Yi-Long Wu.

Ad Hoc Workgroups

T Coding and Size Measurement in Preinvasive and Lepidic Adenocarcinoma Workgroup. William D.

Travis (chair), Hisao Asamura, Alex Bankier, Mary Beth Beasley, Frank Detterbeck, Douglas B. Flieder, Jin Mo Goo, Heber MacMahon, David Naidich, Andrew Nicholson, Charles A. Powell, Mathias Prokop, Ramon Rami-Porta, Valerie Rusch, Paul van Schil, and Yasushi Yatabe.

Multiple Pulmonary Sites of Involvement Workgroup. Frank Detterbeck (chair), Douglas A. Arenberg, Hisao Asamura, Vanessa Bolejack, John Crowley, Jessica S. Donington, Wilbur A. Franklin, Nicolas Girard, Edith M. Marom, Peter J. Mazzone, Andrew G. Nicholson, Valerie W. Rusch, Lynn T. Tanoue, and William D. Travis.

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 http://dx.doi.org/10.1016/j.jtho.2016.06.028.

References

1. Detterbeck FC. Stage classification and prediction of prognosis: the difference between accountants and speculators. *J Thorac Oncol*. 2013;8:820–822.
2. Gospodarowicz M, O'Sullivan B. Prognostic factors in cancer patient care. In: Gospodarowicz MK, ed. *Prognostics Factors in Cancer*. 2nd ed. Hoboken, NJ: Wiley-Liss; 2001:95–104.
3. Rami-Porta R, Bolejack V, Crowley J, et al. The IASLC Lung Cancer Staging Project: proposals for the revisions of the T descriptors in the forthcoming eighth edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2015;10:990–1003.
4. Asamura H, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the N descriptors in the forthcoming eighth edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2015;10:1675–1684.
5. Eberhardt W, Mitchell J, Crowley J, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the M descriptors in the forthcoming eighth edition of the TNM classification of lung cancer. *J Thorac Oncol*. 2015;10:1515–1522.
6. Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2016;11:39–51.
7. Travis D, Asamura H, Bankier AA, et al. The IASLC Lung Cancer Staging Project: proposals for coding T categories for subsolid nodules and assessment of tumor size in part-solid tumors in the forthcoming eighth edition of the TNM classification of lung cancer [e-pub ahead of print]. *J Thorac Oncol*. 2016. http://dx.doi.org/10.1016/j.jtho.2016.03.025. accessed July 25, 2016.
8. Detterbeck F, Nicholson F, Franklin W, et al. The IASLC Lung Cancer Staging Project: summary of proposal revisions of the classification of lung cancers with multiple pulmonary sites of involvement in the forthcoming eighth ediiton of the TNM classification. *J Thorac Oncol*. 2016;11:639–650.
9. Detterbeck F, Bolejack V, Arenberg D, et al. The IASLC Lung Cancer Staging Project: background data and proposals for the classification of lung cancer with separate tumor nodules in the forthcoming eighth ediiton of the TNM classification for lung cancer. *J Thorac Oncol*. 2016;11:681–692.
10. Detterbeck F, Arenberg D, Asamura H, et al. The IASLC Lung Cancer Staging Project: background data and proposed criteria to distinguish separate primary lung cancers from metastatic foci in patients with two lung tumors in the forthcoming eighth edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2016;11:651–665.
11. Detterbeck F, Arenberg D, Asamura H, et al. The IASLC Lung Cancer Staging Project: background data and proposals for the application of TNM staging rules to lung cancer presenting as multiple nodules with ground glass or lepidic features or a pneumonic-type of involvement in the forthcoming eighth edition of the TNM classification. *J Thorac Oncol*. 2016;11:666–680.
12. Groome PA, Bolejack V, Crowley J, 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 tumours. *J Thorac Oncol*. Aug 2007;2:694–705.
13. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130:515–524.
14. Rami-Porta R, Bolejack V, Giroux DJ, et al. The IASLC Lung Cancer Staging Project: the new database to inform the eighth edition of the TNM classification of lung cancer. *J Thorac Oncol*. 2014;9:1618–1624.
15. Nicholson A, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the clinical and pathologic staging of small cell lung cancer in the forthcoming eighth edition of the TMN classification for lung cancer. *J Thorac Oncol*. 2016;11:300–311.
16. Crowley J, LeBlanc M, Jacobson J, Salmon S. Some exploratory tools for survival analysis. Paper presented at: First Seattle Symposium in Biostatistics: Survival Analysis. November 20–21, 1995; Seattle, WA.
17. LeBlanc M, Crowley J. Survival trees by goodness of split. *J Am Stat Assoc*. 1993;88:457–467.
18. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed*. 2004;75:45–49.
19. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168:656–664.
20. Travis WD, Giroux DJ, Chansky K, et al. The IASLC Lung Cancer Staging Project: proposals for the inclusion of broncho-pulmonary carcinoid tumors in the forthcoming (seventh) edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2008;3:1213–1223.
21. Marshall HM, Leong SC, Bowman RV, Yang IA, Fong KM. The science behind the 7th edition Tumour, Node,

Metastasis staging system for lung cancer. *Respirology*. 2012;17:247–260.
22. Rami-Porta R, Asamura H, Goldstraw P. Predicting the prognosis of lung cancer: the evolution of tumor, node and metastasis in the molecular age—challenges and opportunities. *Transl Lung Cancer Res*. 2015;4:415–423.
23. Kameyama K, Takahashi M, Ohata K, et al. Evaluation of the new TNM staging system proposed by the International Association for the Study of Lung Cancer at a single institution. *J Thorac Cardiovasc Surg*. 2009;137:1180–1184.
24. Suzuki M, Yoshida S, Tamura H, et al. Applicability of the revised International Association for the Study of Lung Cancer staging system to operable non-small-cell lung cancers. *Eur J Cardiothorac Surg*. 2009;36:1031–1036.
25. Yano T, Morodomi Y, Ito K, et al. Verification of the newly proposed T category (seventh edition of the tumor, node, and metastasis classification) from a clinicopathological viewpoint in non-small cell lung cancer—special reference to tumor size. *J Thorac Oncol*. 2010;5:45–48.
26. Dassanayake DLB, Muthunayake TM, Senevirathna KHMP, Siribaddana A. Staging of lung cancer in a tertiary care setting in Sri Lanka, using TNM 7th edition. A comparison against TNM6. *BMC Res Notes*. 2012;5:143–143.
27. León-Atance P, Moreno-Mata N, González-Aragoneses F, et al. Multicenter analysis of survival and prognostic factors in pathologic stage I non-small-cell lung cancer according to the new 2009 TNM classification. *Arch Bronconeumol*. 2011;47:441–446 [in English, Spanish].
28. Strand TE, Rostad H, Wentzel-Larsen T, Von Plessen C. A population-based evaluation of the seventh edition of the TNM system for lung cancer. *Eur Respir J*. 2010;36:401–407.
29. Chien C-R, Yang S-T, Chen C-Y, et al. Impact of the new lung cancer staging system for a predominantly advanced-disease patient population. *J Thorac Oncol*. 2010;5:340–343.
30. Jhun BW, Lee K-J, Jeon K, et al. Clinical applicability of staging small cell lung cancer according to the seventh edition of the TNM staging system. *Lung Cancer*. 2013;81:65–70.
31. Wang J, Wu N, Zheng Q, et al. Evaluation of the 7th edition of the TNM classification for lung cancer at a single institution. *J Cancer Res Clin Oncol*. 2014;140:1189–1195.
32. Ruffini E, Filosso PL, Bruna MC, et al. Recommended changes for T and N descriptors proposed by the International Association for the Study of Lung Cancer—Lung Cancer Staging Project: a validation study from a single-centre experience. *Eur J Cardiothorac Surg*. 2009;36:1037–1044.
33. Ou SHI, Zell JA. Validation study of the proposed IASLC staging revisions of the T4 and M non-small cell lung cancer descriptors using data from 23,583 patients in the California Cancer Registry. *J Thorac Oncol*. 2008;3:216–227.
34. Zell JA, Ou SHI, Ziogas A, Anton-Culver H. Validation of the proposed International Association for the Study of Lung Cancer non-small cell lung cancer staging system revisions for advanced bronchioloalveolar carcinoma using data from the California Cancer Registry. *J Thorac Oncol*. 2007;2:1078–1085.
35. Lee JG, Lee CY, Bae MK, et al. Validity of International Association for the Study of Lung Cancer proposals for the revision of N descriptors in lung cancer. *J Thorac Oncol*. 2008;3:1421–1426.
36. Webber C, Gospodarowicz M, Sobin L, et al. Improving the TNM classification: findings from a 10-year continous literature review. *Int J Cancer*. 2014;135:371–378.
37. Austin P, Steyereberg E. Events per variable (EPV) and the relative performance of different strategies for estimating the out of sample validity of logistic regression models [e-pub ahead of print]. *Stat Methods Med Res*. 2014. http://dx.doi.org/10.1177/0962280214558972, accessed February 17, 2016.
38. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49:1373–1379.
39. Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy. *J Clin Epidemiol*. 1995;48:1495–1501.
40. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *J Clin Epidemiol*. 1995;48:1503–1510.
41. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med*. 2016;35:214–226.
42. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol*. 2005;58:475–483.
43. Moons K, Altman D, Reitsma J, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162:W1–W73.
44. Mahar A, Compton C, McShane L, et al. Refining prognosis in lung cancer. A report on the quality and relevance of clinical prognostic tools. *J Thorac Oncol*. 2015;15:1576–1589.