

# The International Association for the Study of Lung Cancer Staging Project: Methods and Guiding Principles for the Development of the Ninth Edition TNM Classification



Frank C. Detterbeck, MD,<sup>a,\*</sup> Katherine K. Nishimura, PhD, MPH,<sup>b</sup>  
Vanessa J. Cilento, MPH,<sup>b</sup> Meredith Giuliani, MD,<sup>c</sup> Mirella Marino, MD,<sup>d</sup>  
Raymond U. Osarogiagbon, M.B.B.S.,<sup>e</sup> Ramon Rami-Porta, MD, FETCS,<sup>f,g</sup>  
Valerie W. Rusch, MD,<sup>h</sup> Hisao Asamura, MD,<sup>i</sup>  
the International Association for the Study of Lung Cancer (IASLC) Staging and  
Prognostic Factors Committee and Advisory Boards<sup>j</sup>

<sup>a</sup>Department of Surgery, Yale University, New Haven, Connecticut

<sup>b</sup>Cancer Research and Biostatistics, Seattle, Washington

<sup>c</sup>The Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada

<sup>d</sup>Department of Pathology, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Regina Elena National Cancer Institute, Rome, Italy

<sup>e</sup>Multidisciplinary Thoracic Oncology Program, Baptist Cancer Center, Memphis, Tennessee

<sup>f</sup>Thoracic Surgery Service, Hospital Universitari Mutua Terrassa, Terrassa, Spain

<sup>g</sup>Lung Cancer Group, Network of Centers for Biomedical Research in Respiratory Diseases (CIBERES), Terrassa, Spain

<sup>h</sup>Thoracic Surgery Service, Sloan Kettering Cancer Center, New York, New York

<sup>i</sup>Division of Thoracic Surgery, School of Medicine, Keio University, Tokyo, Japan

Received 30 December 2021; revised 11 February 2022; accepted 18 February 2022

Available online - 10 March 2022

## ABSTRACT

**Introduction:** Stage classification provides a consistent and concise nomenclature about the anatomic extent of the cancer. This is a fundamental cornerstone in the

management of patients; it enables reporting results and facilitates comparing one treatment to another and judging how closely clinical trial results apply to an individual patient. A nomenclature must be relatively static; however,

\*Corresponding author.

<sup>j</sup>See [Supplementary Appendices 1 and 2](#).

**Disclosure:** Dr. Nishimura reports statistical consulting by Cancer Research And Biostatistics (CRAB) employees for manuscripts related to the staging project is funded by the International Association for the Study of Lung Cancer (IASLC); statistical consulting by CRAB employees for data collection and analysis in the staging project is funded by IASLC; consulting fees—statistical consulting by CRAB employees for the staging project is funded by the IASLC; support for attending meeting and/or travel per IASLC contract with CRAB includes funds to attend the World Conference on Lung Cancer conference to participate in staging project meetings. Dr. Giuliani reports receiving personal payment for consulting fees from Bristol-Myers Squibb and honoraria fees from AstraZeneca and Bristol-Myers Squibb for lectures. Dr. Osarogiagbon reports receiving funding support for R01CA172253; consulting fees from the National Institutes of Health National Cancer Institute Division of Cancer Control and Population Sciences Health Systems and Interventions Research Branch, American Cancer Society, Association of Community Cancer Centers, AstraZeneca, Memorial Sloan Kettering Cancer Center, Druckenmiller Center for Lung Cancer Research, Genentech/Roche, and Tryptych Health Partners; honoraria fees from Biodesix and Medscape for lectures; patents planned, issued, or pending for three issued patents for a lymph node specimen collection kit: two United States, one People's

Republic of China; reports being on the following boards—Board Chair, Hope Foundation for Cancer Research, Member, LUNGEvity Health Equity Council, Steering Committee, National Lung Cancer Round Table, Co-Chair Ad hoc Committee on Diversity and Inclusion, Fleischner Society, Member, and GO2 Foundation Scientific Advisory Board; reports stock or stock options with Eli Lilly, Gilead Sciences, and Pfizer. Dr. Rusch reports receiving institutional grants to support clinical trials from Genelux, Inc., and Genentech; travel reimbursement only for robotic teaching session in 2019 from Intuitive Surgical; is a Member, Data Safety Monitoring Committee MARS II Trial, United Kingdom, but received no reimbursement; is a Co-Chair, Thoracic Malignancy Staging Committee, National Cancer Institute/National Institutes of Health and receives travel and meeting preparation reimbursement. Dr. Asamura reports receiving grants from the following: Taiho Pharmaceutical, Johnson and Johnson, Covidien, Astellas, Eli Lilly & Co, AstraZeneca, and Chugai. The remaining authors declare no conflict of interests.

Address for correspondence: Frank C. Detterbeck, MD, Department of Surgery, Yale University, 330 Cedar Street, New Haven, CT 06520-8062. E-mail: [frank.detterbeck@yale.edu](mailto:frank.detterbeck@yale.edu)

© 2022 Published by Elsevier Inc. on behalf of International Association for the Study of Lung Cancer.

ISSN: 1556-0864

<https://doi.org/10.1016/j.jtho.2022.02.008>

periodical refinement is needed to adjust to a changing landscape of clinical relevance. Changes must be well justified and thoughtfully developed to maintain the ability to communicate clearly and facilitate comparisons across time.

**Methods:** For thoracic malignancies (lung, pleura, thymus, and esophagus), the International Association for the Study of Lung Cancer (IASLC) has leveraged its worldwide multidisciplinary reach, permitting a sophisticated approach to this process. Refinement of stage classification for the ninth edition of TNM is underway; this article describes the approach adopted by the IASLC Staging and Prognostic Factors Committee.

**Results:** Key guiding principles include the ability to maintain communication over time, a classification that discriminates homogeneous cohorts of tumors consistently across the world in multiple settings, treatment approaches, and patient characteristics, including clinical relevance and practical applicability. The IASLC has again assembled a large international database to permit multifaceted analysis. Providing confidence that the classification performs consistently in multiple settings, treatments, and patients requires consistent discrimination in multiple subset analyses. Although observed outcomes of patients in the 2011 to 2019 database are essential, considerations about how the classification will be used are also important to ensure clinical relevance and applicability.

**Conclusions:** The strategy developed by the Staging and Prognostic Factors Committee is carefully designed to provide useful refinements to the stage classification of thoracic malignancies for the ninth edition of TNM classification of cancers.

© 2022 Published by Elsevier Inc. on behalf of International Association for the Study of Lung Cancer.

*Keywords:* Lung cancer; Non-small cell lung cancer; TNM classification; Lung cancer staging; Prognosis

---

## Introduction

Stage classification is a fundamental cornerstone in the management of patients with cancer. It is a nomenclature about the anatomical extent of the cancer that is essential in many ways; for example, a consistent and concise description of tumor extent in a cohort of patients enables reporting results, facilitates comparing one treatment to another, and judging how closely clinical trial results can be applied prospectively to a patient. Tumor burden has a strong impact on prognosis, in addition to other patient, tumor, and environmental factors and the selected treatment.<sup>1,2</sup> Thus, stage classification is a crucial tool in conducting and reporting research studies, selecting treatment, estimating prognosis, and guiding global cancer control activities.

A single, consistent stage classification system for use around the world is defined by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC). These separate organizations work collaboratively to define a single system. A nomenclature must be relatively static; constant change would create confusion and an inability to communicate. Nevertheless, medicine is constantly evolving; imaging and other assessments improve, new treatments are developed, new characteristics are recognized as relevant. Thus, the classification must be periodically refined—at protracted intervals with limited, well-justified changes, so that we adjust to a changing landscape of clinical relevance yet maintain the ability to communicate clearly and facilitate comparisons across time. The current eighth edition of the stage classification took effect worldwide on January 1, 2017 (in the United States, on January 1, 2018). The next update (the ninth edition) is scheduled to be introduced in 2024.

Classification of the anatomical extent of a tumor involves the following three components: T denotes the primary tumor extent, N the extent of regional lymph node involvement, and M the extent of distant metastasis. Each component is divided into several categories (e.g., T1, T2, T3, T4), which may sometimes be further subdivided (e.g., T1a, T1b, T1c). The categories are defined by specific anatomical characteristics, known as descriptors. Finally, to facilitate clinical use, the classification is simplified by coalescing various T, N, and M categories to form stage groups.

For lung cancer, the AJCC-UICC stage classification system has relied on an unparalleled scientific analysis done by the Staging and Prognostic Factors Committee (SPFC) of the International Association for the Study of Lung Cancer (IASLC) that began in 1996. The IASLC developed global databases (DBs) of approximately 100,000 patients each that provided the basis for the seventh and eighth editions of the stage classification system.<sup>3,4</sup> This initiative was expanded to include other thoracic malignancies, such as follows: mesothelioma (in collaboration with the International Mesothelioma Interest Group), thymic malignancies (in collaboration with the International Thymic Malignancies Interest Group), and esophageal cancer (in collaboration with the Worldwide Esophageal Cancer Collaboration). This article describes the principles and processes that are being used by the IASLC SPFC to provide a solid scientific basis for proposals for the ninth edition of stage classification of thoracic malignancies. The primary focus is lung cancer; the approach for other thoracic malignancies is similar albeit with limitations owing to the lower incidence of these cancers.

## Fundamental Principles

### Primary Objective

The goal of the SPFC is to refine the stage classification of thoracic malignancies for the ninth edition of the AJCC-UICC stage classification. Classification inherently amalgamates individual cases (consisting of a spectrum of tumors) into cohorts. It is essential that the categorization can be applied consistently and is clinically relevant.

Clinical relevance means that the stage classification must be useful in the context of how patients are (currently) managed. Nevertheless, a connection between stage classification, treatment, and outcomes should not be misunderstood to mean that stage classification alone defines treatment or outcome. Stage classification is only one factor, and it is only a nomenclature we use to communicate.

Stage classification applies to a tumor. The implications of a particular anatomic tumor extent may vary in different patients, who have many additional characteristics (e.g., age, comorbidities, preferences and values, suitability for treatment). Stage classification does not apply to patients, only to the tumor itself.

Although anatomic tumor extent has many clinical implications (e.g., treatment approach, prognosis), many other tumor characteristics have clinical implications, which can be described in additional classification schemes. Examples include histologic classification, genomic characteristics (e.g., targetable driver mutations, molecular profiling), and biological aggressiveness (e.g., positron emission tomography [PET] avidity, observed growth kinetics). Such additional tumor features may have minimal or major clinical implications depending on the circumstances—for example, the NSCLC histotype has little impact in a stage I lung cancer but is important in choosing chemotherapy in stages II to IV lung cancer. Such factors (including patient- and environment-related factors) are not part of stage classification, which explicitly addresses the anatomic extent of a tumor.

### Scope

Although the stage classification system is a universal global standard, discrepancies exist in economic and health care resources in different regions. Furthermore, although this undoubtedly affects outcomes, the effect on the discrimination between stage groups is likely less pronounced. Nevertheless, the availability of data on patients with lung cancer is limited in regions with constrained resources. Therefore, although the SPFC is committed to developing a classification that is useful worldwide, validation in low- and middle-income countries is problematic.

Stage classification is intertwined with clinical relevance, outcomes, and treatment. Nevertheless, stage classification cannot serve as a treatment guideline. Treatment must continually advance, whereas stage classification must remain consistent with only carefully considered periodic updates. Appropriate treatment depends on patient and tumor characteristics. Stage classification is only a nomenclature we use to communicate; it is not the name we put on a tumor's extent that defines treatment—treatment selection is based on the outcomes of relevant clinical trials and how well they apply to a particular patient. Furthermore, stage classification must apply regardless of the treatment given.

The stage classification system is not a prognostic prediction system.<sup>1</sup> It reflects only one aspect (tumor burden) that affects prognosis, and the observed outcomes apply to a global average of patients diagnosed in the past (1999–2010 for the eighth edition), who received various treatments. The SPFC has separately undertaken development of a prognostic prediction system.<sup>5</sup> Such a system, if robust, must reflect as many prognostic factors as possible (tumor-related, patient-related, environment-related, and treatment-related), must provide a contemporary estimate, and be flexible so that new factors can be included as they are identified. This is a complex undertaking, as the impact of many prognostic factors varies across the clinical setting, tumor stage, and treatment approach. Finally, appropriate validation of a model in a rapidly changing field is challenging.<sup>6</sup>

### Guiding Principles

As with the previous revisions (seventh and eighth editions),<sup>7,8</sup> the SPFC has defined general guiding principles (Table 1) for refinement of the stage classification system.

Criteria to define stage classification must reflect properties that are inherent to tumors with a particular anatomic extent, that manifest consistently in a broad variety of patients, settings, and treatments; and that are stable over time. It is also important to consider clinical relevance—for example, similarity to criteria currently used in treatment selection (although treatment paradigms evolve and partially depend on factors other than tumor extent). Practical issues also matter. Distinguishing a subgroup may simply contribute complexity if the incidence is rare or if management is not affected.

Prognosis is a major tool in developing a stage classification system. Anatomic extent of disease has a major impact on prognosis for most tumors broadly across regions, settings, and treatment approaches. Nevertheless, a sophisticated approach is needed, because prognosis also depends on patient characteristics (e.g., age,

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

1. Existing definitions of T, N, and M categories should be maintained unless there is compelling evidence to support a change. This will ensure backward compatibility with the previous staging system, whenever possible (this is less important for the stage grouping because compatibility can be achieved if consistent T, N, and M categories are available).
2. The T, N, and M categories and stage groups should group together tumors that have similar clinical implications and discriminate cohorts with different clinical implications:
  - a. This should be based on relevant data as much as possible. The key metric is consistent discrimination between categories and groups (across various settings, histotypes, global regions, clinical settings, c-stage, p-stage, R status) to ensure general applicability.
  - b. Practical considerations can influence clinical relevance (e.g., generalizability, familiarity, ease of use, prevalence of a characteristic, facility of identifying the characteristic, applicability to treatment considerations).
3. Lung cancer stage classification must be consistent with the general rules for TNM classification of malignant tumors:
  - a. The TNM system is based on the following three components: T for the extent of the primary tumor, N for the extent of regional lymph node involvement, and M for the extent of distant metastasis(es).
  - b. c-stage classification (cTNM) involves all available information that reflects the anatomic extent of a tumor before the initiation of treatment. This may involve evidence from symptoms, physical examination, imaging, endoscopy, biopsies, and surgical exploration. Thus, the c-stage and the confidence in the assessment may evolve as evaluation of a patient's tumor progresses during investigations to assess the extent of the tumor.
  - c. p-stage classification (pTNM) includes all information available from a surgical resection (or attempted resection), supplemented with all available information for c-stage classification.
  - d. Specific TNM categories may be combined into stage groups.
  - e. The definitions for T, N, and M categories and stage groups should be the same for c-stage and p-stage classification.
  - f. Definitions of TNM categories and stage groups may be telescoped or expanded for clinical or research purposes as long as the basic definitions of the TNM categories and stage groups are not changed.
  - g. If there is doubt about the correct T, N, or M category to which a particular case should be allotted, the lower (less advanced) category should be chosen.
  - h. In the case of multiple primary tumors in one organ, the tumor with the highest T category should be classified and the multiplicity or the number of tumors should be indicated in parentheses, for example, T2(m) or T2(5) (in lung cancer, this applies to multifocal ground-glass/lepidic adenocarcinomas but not to separate primary tumors).

c-stage, clinical stage; cTNM, clinical TNM; p-stage, pathologic stage; pTNM, pathologic TNM; R, residual tumor.

comorbidities, socioeconomic status, values), environmental factors (e.g., region, health care system, access, risk factors), time period (e.g., 1990s, 2000s, 2010s), other tumor characteristics (e.g., grade, metabolic activity, genomic characteristics), treatment approach, and the point in time when prognosis is assessed (e.g., at diagnosis, on [successful] completion of treatment, 1 year later). These additional characteristics primarily affect the actual prognosis (i.e., calibration) and have less impact on the discriminatory value of a classification. Therefore, the way prognosis is used as a tool to guide stage classification is to assess consistency in discrimination across multiple subgroup analyses—not prognosis by itself in one analysis.

The presence of many confounding factors creates a risk of potential false-positive (or negative) results when using prognosis to discriminate between cohorts. The use of prognosis demands that (1) it is applied in a large cohort that approximates the universe of contemporary patients; (2) major potential confounding factors are accounted for either by multivariate adjustment or consistent discrimination in subset analyses involving these factors; and (3) validation analyses reveal transportability (generalizability) across multiple domains (historical, geographic, methodologic [e.g., clinical-stage [c-stage], pathologic-stage [p-stage], thoroughness of

stage evaluation], disease spectrum [e.g., histotype] and follow-up intervals). A robust analysis requires a sufficient sample size (i.e., number of events in each [subgroup] analysis, taking into account the number of additional factors being assessed).

## Approach to Developing Stage Classification Refinements

### Structure of SPFC

The SPFC is divided into several domains (lung, thymic, mesothelioma, and esophageal). The lung cancer domain is divided into subcommittees (steering committee [SC], methods and validation, T descriptors, N descriptors, M descriptors, stage grouping, R factor, small cell lung cancer [SCLC], neuroendocrine tumors, multiple pulmonary sites of lung cancer, ground-glass/lepidic/in situ tumors, imaging, lymph node chart, molecular, prognostic prediction; see [Supplementary Appendix 2](#)).

The SPFC chair is appointed by the IASLC Board of Directors (BOD). The SPFC chair appoints a SC and chairs of the domains and subcommittees, with approval of the BOD. The term of these appointments lasts until the next edition of the stage classification takes effect. The domain and subcommittee chairs

**Table 2. Process of Developing Revisions to the TNM Classification**

1. Planning phase—identification of potential factors to consider, including the following:
  - o Weaknesses of eighth edition (areas of ambiguity, poor implementability)
  - o Potential heterogeneity within an eighth edition category or group
  - o Factors suggested from external analyses (literature)
  - o Factors previously suggested for further analysis (in seventh or eighth edition)
  - o Factors suggested by advances in stage evaluation (imaging, interventions)
  - o Factors suggested by emerging decision points for treatment interventions
  - o Factors suggested by the subcommittees of the SPFC
2. Implementation phase—development of a strategy
  - o Define details of potential analyses (see text)
  - o Assess ability of 2011–2019 IASLC-CRAB DB to address the questions (sample size, number of events, prevalence, robustness of internal assessment)
  - o Identify additional data sources that could contribute to analysis of particular factors
  - o Identify sources for external validation
  - o Select factors for which analysis is feasible
3. Exploratory analysis
  - o Clean and assess the IASLC-CRAB DB (eliminate errors, assess need for adjustment in combining data of different types/sources, assess patterns of missingness in data)
  - o Produce graphs of the relevant outcomes by potential factors (all cases and across subsets: e.g., clinical and pathologic setting, relevant T, N, M categories, and relevant treatment modalities)
  - o Assess acceptability of potential factors by adequate clinical relevance, consistent ordering
  - o Internal assessment of generalizability across lung cancer histotypes, continents
  - o Identify and analyze additional subsets to address potential confounders and explore inconsistencies
  - o Consider supplementary graphs of additional outcomes (e.g., recurrence, progression) and their interpretations
4. Confirmatory analysis and selection
  - o Assess discrimination of potential factors identified in the exploratory analysis—that is, assess statistical significance of differences between adjacent categories, groups (taking into account issues of clinical relevance and sample size—for example, ordering that is predominantly consistent and not statistically significant trends may be acceptable when sample size limits the power of detection)
  - o Assess homogeneity within groups (e.g., confidence interval for survival parameters; statistical significance analysis); note that demonstration of reasonable homogeneity does not preclude finding outliers or additional parameters that could be used to subdivide a group
  - o Selection of factors after detailed review of steps up to this point and discussion within subcommittees and before the entire SPFC
5. Internal and external validation (may be simultaneous with refinement, vetting, and publication)
  - o Internal validation of generalizability:
    - historical (i.e., backward application of the ninth edition system within 2010–2020 DB and 1999–2010 DB, where possible; forward application of the eighth edition system in the ninth edition DB)
    - geographic (i.e., regions, already done in confirmatory phase)
    - methodologic (e.g., different levels of health care sophistication; extensive vs. limited stage evaluation; academic vs. community, urban vs. rural)
    - spectrum (e.g., SCLC, NSCLC subtypes, carcinoid; screening vs. regular setting; persons who smoke vs. persons who never smoked)
    - follow-up interval (i.e., no subsequent crossover of curves; automatically confirmed)
  - o External validation in an appropriate DB
6. Refinement, vetting, and publication of proposals
  - o Proposals of each subcommittee presented and discussed among entire SPFC
  - o Draft papers circulated to all SPFC members for revision and approval
  - o External journal review process
  - o Publication in JTO with open dissemination, open ability to comment
7. Acceptance by UICC/AJCC

AJCC, American Joint Committee on Cancer; CRAB, Cancer Research And Biostatistics; DB, database; IASLC, International Association for the Study of Lung Cancer; JTO, Journal of Thoracic Oncology; SPFC, Staging and Prognostic Factors Committee; UICC, Union for International Cancer Control.

appoint members to their respective groups, with approval of the SC and BOD. The subcommittees are constituted to be international and multidisciplinary and include members with clinical and other expertise that contributes to the classification of thoracic malignancies. The subcommittees include biostatisticians of Cancer Research And Biostatistics (CRAB). Subcommittees may also appoint individuals as advisory board members on the basis of particular interests and subcommittee needs.

SPFC members are chosen on the basis of recognized interest and expertise in stage classification of thoracic malignancies, for example, as revealed by publications on this subject. Interested individuals may self-nominate or be nominated by any IASLC member. SPFC membership terms are 3 years, renewable twice (i.e., up to 9 years). Annually, the SC, in discussion with the domain and subcommittee chairs, reviews the membership of the subcommittees to address changes emerging from the needs of each subcommittee.

The SPFC subcommittees meet regularly by means of teleconference and in person at the annual IASLC World Conference on Lung Cancer. Proposals for classification revision are developed within the individual subcommittees and then presented and discussed at joint meetings. After revision and appropriate validation, the proposals are provided to the entire SPFC for further input and formal approval. These are then submitted to the AJCC and UICC—the bodies with the final authority to define the ninth edition stage classification of malignant tumors.

### Process

The process of developing revisions to the stage classification system can be divided into multiple phases (Table 2). It begins with assessment of shortcomings of the existing system and followed by a strategy to address these, for example, through collection of cases in the IASLC-CRAB DB. This may be supplemented by collaboration with other organizations for external validation. Potential classification revisions are based on demonstrated validity and generalizability and practical considerations, such as the amount of internal or external data, the extent of analysis and validation that is possible, and the clinical relevance (note that for most nonthoracic tumor sites and for early editions of lung cancer stage classification, the system was based largely on empirical considerations with little ability for confirmation).

### Database

The SPFC stage classification effort is a huge undertaking, made possible by the voluntary commitment of countless individuals in many institutions across the world, who have invested much effort to provide data to the SPFC. The world will be indebted to this army of selfless individuals.

The IASLC-CRAB DB is currently being assembled; a separate publication will address details of the DB. The size of the DB will be similar to that used for the seventh and eighth editions of lung cancer stage classification (81,495 and 77,156, respectively, in the final analysis). There is broad representation involving at least 25 countries. The DB consists of data entered prospectively into an electronic data capture system by approximately 50 participating sites and “batch” data sets from various countries and institutions. The data elements of the batch data sets must be mapped to those of the IASLC-CRAB DB to be included. Data have been contributed from countries in Africa, Asia, Australia, Europe, the Middle East, and North and South America.

The IASLC-CRAB DB has both strengths and weaknesses. Certain areas are overrepresented (Asia/Australia accounts for ~50% of the cases, Japan alone for ~25%), whereas other areas are underrepresented (Africa <1%). The completeness of the many data fields is variable. Nevertheless, the contribution of less-resourced countries and advanced stage cases is greater in the current DB than for the prior editions, reflecting a directed effort in these areas. To what degree skewedness of the data or missing data affects analysis of specific categories and components will become more apparent as these analyses mature.

Because this is a voluntary DB, the SPFC has no ability to control the size, regional distribution, or completeness of data elements beyond encouraging participation. The SPFC strongly encourages the use of the electronic data capture system because it leads to more complete data, but this represents only a minority of the cases.

### Analytical Approach and Statistical Methods

**Endpoints.** Multiple endpoints can be considered to assess the impact of the anatomic extent of cancer, but none is ideal (Table 3). All endpoints are associated with confounders, limitations, and practical constraints, with varying impact depending on the context. For example, tumor extent may be the primary driver of overall survival (OS) in an aggressive tumor (e.g., SCLC) or extensive disease (widespread metastases), especially if treatment has limited effectiveness. In contrast, in healthy patients with stage I NSCLC in a well-resourced setting, OS primarily reflects competing causes of death and the treatment given; tumor extent plays a lesser role. In this setting, recurrence may be a better endpoint. Furthermore, how patients present affects the spectrum of tumor aggressiveness. Screening (or frequent imaging resulting in incidental diagnosis) markedly increases the proportion of indolent tumors,<sup>9</sup> thereby affecting both prognosis and the apparent treatment effectiveness. Conversely, in settings with limited access to care and resources, presenting patients have a higher proportion of aggressive cancers.

At the origin of stage classification, OS was the only outcome considered—a logical choice at a time when treatment effectiveness, access to care, early detection, and availability of data were limited. For the ninth edition, the SPFC still deems OS to be the primary endpoint in general. Nevertheless, recurrence is deemed to be more representative of the impact of the tumor extent for early stage and indolent tumors.

The fact that all endpoints reflect disease extent and other factors creates challenges in assessment of the impact of the anatomic extent of a cancer. OS can be

**Table 3.** Potential Endpoints for Development of a Stage Classification System

Endpoint	Advantages	Disadvantages
Overall survival	Available, unambiguous	Multiple confounders: (patient-, tumor-, environment-, and treatment-related)
Natural history (not treated)	Eliminates treatment-related effects	Abstract, limited clinical relevance, confounded by competing causes of death
Recurrence	Eliminates competing causes of death, reflects nonlethal cancer outcomes	Rarely reported, includes treatment-related factors, dependent on follow-up interval and investigation intensity, potential ambiguity of recurrence vs. second primary (especially with multifocal ground-glass/lepidic lung cancer)
Disease-free survival	Often reported	Combines recurrence and death (including competing causes of death), includes treatment-related factors, dependent on follow-up interval and investigation intensity
Lung cancer-specific survival	Eliminates competing causes of death	Rarely reported, includes treatment-related factors, dependent on reliability of cause of death

viewed as the appropriate measure in a global, population-based contemporary DB because it represents a real-life assessment of tumor extent together with an overall average of treatment and other confounding factors. In that case, confidence in a potential classification is achieved if generalizability is found across multiple validation parameters, such as follows: historic (consistent discrimination over time), geographic (across regions), methodologic (e.g., computed tomography versus PET, c-stage versus p-stage), spectrum (e.g., different histotypes, smoking status, symptom detected versus screening detected), and follow-up intervals (e.g., at 1, 2, 3, and 5 years).<sup>8,10</sup>

A global population-based DB is not available, but a large DB might represent an adequate approximation. The lack of a population-based DB can be mitigated if the DB contains enough cases representing the full spectrum and subset analyses are conducted within the major components of this spectrum. The SPFC is evaluating potential deficiencies that are not adequately represented, and potential strategies to address these—for example, data for inclusion in the IASLC-CRAB DB or for external validation. Thus, by addressing how representative the DB is and demonstrating generalizability through subset and validation analyses, the SPFC has confidence that the IASLC-CRAB DB will permit a robust analysis.

**Confounding.** Anatomic disease extent has a major role in determining prognosis, but other tumor-, patient-, and environment-related factors also play a role. Defining to what degree an observed outcome is due to disease extent versus other factors is essentially impossible at this time. Nevertheless, it is unnecessary to disentangle this, provided anatomic disease extent is playing a role in a consistent manner. The SPFC has

therefore focused on how to be confident that this is the case (i.e., consistent discriminatory impact of anatomic extent of the tumor).

The SPFC has identified a first tier of subgroups in which it is essential that any proposed categories or groups on the basis of disease extent exhibit a consistent effect on ordering and discrimination. These include T, N, and M subgroups, pathologic and clinical-stage settings. A disease extent factor with discordant or variable ordering or lack of discrimination in any of these subgroup analyses is rejected as a useful stage classification parameter (with a possible exception if a factor occurs too rarely to permit a valid assessment in a particular subgroup analysis).

Treatment deserves special consideration. For all practical purposes, other prognostic factors cannot be changed either for a cohort or an individual patient (e.g., grade, genetic profile, age, performance status [PS], socioeconomic status, environment). Nevertheless, treatment can be changed easily and reflects available evidence, resources, and prevailing attitudes. Therefore, it is critical that subgroup analysis by commonly used treatment approaches demonstrates consistent ordering and discrimination (this analysis only addresses treatment modalities—not details, such as lobectomy versus segmentectomy, radiotherapy dose, systemic therapy agents).

A second-tier assessment to ensure generalizability is assessment of stage categories and groups among the major lung cancer histotypes and continental regions. If a stage classification parameter has consistent discrimination among these subgroup analyses (first and second tiers, major treatment approaches), we can be confident that regardless of confounding by other prognostic factors, the parameter is consistent, relevant, and applicable in clinical settings.

**Statistical Analysis.** Developing refinements to stage classification of lung cancer is not primarily a matter of statistical analysis of a large DB for multiple reasons. Observed outcomes reflect confounding factors in addition to the tumor extent; clinical judgment is needed to assess the impact of these. In a large DB, statistical significance may be present even for a small difference in outcomes that is clinically meaningless. There is a temptation to subdivide further and further with large data sets, but with an increasing risk of limited clinical relevance and spurious results (driven by confounders). Finally, characteristics or groupings suggested by rote statistical analysis may not be clinically relevant (i.e., not mesh with clinical care). Therefore, the SPFC approach is to start with an assessment of clinical relevance and evaluation of factors likely influencing observed outcomes and only then to add statistical assessment to supplement the interpretation and confidence in the conclusions.

Specifically, the SPFC process begins with visual assessment of outcomes with respect to potential descriptors, categories, and groupings. These are explored with subsets to assess consistency and ordering. Specific endpoints, potential confounders, and particular subsets are considered to judge to what extent the observed outcomes reflect the impact of the tumor extent. Factors that seem to be useful to discriminate cohorts on the basis of anatomic extent are then subjected to further statistical analysis. The key metric is the consistent presence of differences in prognosis between categories and stage groupings across multiple analyses of particular cohorts. Although statistical significance is sought, acceptance of a definition can be made on the basis of ordering and trends, because some subgroup analyses may be insufficiently powered owing to cohort size, the number of events, and the number of confirmatory analyses.

Survival is measured from the date of diagnosis for c-stage tumors and the date of surgery for p-stage tumors and analyzed using Kaplan-Meier methods. Cases with missing relevant data are excluded from the analyses. Survival estimates are compared using the likelihood ratio test from Cox proportional hazards regression. Cox regression analysis, adjusted for baseline factors (e.g., age, sex, region, cell type), is 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 is assessed primarily by using the  $R^2$  measure and the concordance index (C-statistic). These measures are deemed to be best, 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). Additional analytical methods, including recursive partitioning and supervised machine learning techniques, such as stepwise model selection, penalized Cox regression (e.g., Lasso, Elastic Net), and random forests, may be used when deemed appropriate.

**Constraints.** In general, the SPFC views differences in OS of less than 5% as a questionable basis on which to define stage categories or groups because of concern about clinical relevance and the possibility of spurious results (differences related to confounding factors rather than tumor extent). Furthermore, the benefit of splitting into such marginal subgroups may be outweighed by unwieldy complexity of the stage classification system. This 5% threshold is a suggestive yardstick; given the number of subgroup analyses, it cannot be a hard criterion.

Factors that are noted in less than 50 cases are also generally considered questionable to include in the definition of stage classification. Statistical analysis is based on the number of events (e.g., deaths); with less than 50 cases, there are generally too few events to permit robust subset analyses for consistency. Finally, the SPFC considers descriptors that are recorded in less than one of 1000 cases for potential omission from the stage classification. Such rare descriptors are of questionable clinical relevance, their prognostic impact is difficult to assess, and they may primarily overcomplicate the classification.

### *Validation and Generalizability*

A first tier (T, N, and M categories; pathologic and clinical-stage groups; common treatment modalities) and a second tier (major lung cancer histotypes and continental regions) of internal validation have already been mentioned. In addition, internal historic validation (consistent discrimination over time) will be done. This can involve an assessment of consistency throughout the accrual period of the ninth edition DB (factors occurring with sufficient frequency) or assessment in the eighth edition DB. Nevertheless, historic validation may not be possible for less frequent or newly recognized factors.

The SPFC intends to carry out internal validation for the major categories and stage groups across countries grouped as high income or middle income. It would also be desirable to validate spectrum generalizability by subset analyses involving smoking status and symptom-detected versus screen-detected cases, but this may not be possible. If possible, internal assessment of generalizability will be done across a range of the extent of stage evaluations performed (e.g., PET, invasive mediastinal assessment).

External validation in predominantly independent DB(s) is planned for the major categories and stage groups. In previous editions, this has involved Surveillance, Epidemiology, and End Results or National Cancer Database analyses.<sup>7,8</sup> The availability of independent DBs with sufficient detail to allow robust validation and the need for a specific population for validation (e.g., region) are key considerations in external validation. External validation in institutional DBs by independent investigators is encouraged; parameters to ensure that such analyses have a reasonable scientific foundation have been previously described.<sup>8</sup>

We anticipate that robust internal validation will not be possible for some potential descriptors. In this case, the SPFC seeks to conduct external validation. A data set with information on the specific factor would suffice, even if it is lacking information on other factors and possible confounders. Details of such an external analysis will depend on the factor in question and what is available for analysis.

## Discussion

Stage classification is intertwined with clinical relevance, outcomes, and treatment. One can argue that the process is circular: we use prognosis to help define the stage classification, the stage influences the treatment approach, and in turn the treatment has a major role in determining prognosis. Nevertheless, this is oversimplified, as there are other factors influencing prognosis and treatment. The SPFC is primarily focused on identifying characteristics of the anatomic extent of disease that are inherently relevant—while recognizing that this anatomical classification is used in a real-world context that includes a myriad of other factors.

Stage classification has served us well, facilitating communication that has been essential to progress in treatment approaches and outcomes. Clearly, there are weaknesses. Classification fundamentally defines specific cohorts of individual cases; however, any categorization of what is essentially a spectrum is inherently to some extent arbitrary. Although a global standard is needed, it is difficult to optimally meet the needs of all regions. Advanced tumors have been underrepresented, perhaps because anatomic tumor extent is less important in this setting. Although some attention is given to limited anatomical dissemination (“oligometastatic” tumors), it seems that classifications based on nonanatomical characteristics are more useful (e.g., PS, histotype, biomarker profile).

Molecular tumor profiling is not part of the TNM classification. The SPFC recognizes that anatomic extent of disease has diminished impact when the cancer is disseminated and other factors become much

more relevant; molecular profiling has particular value as a predictive factor for treatment response. This underscores that distinct classification schemes for various characteristics of tumors and patients are clinically useful (e.g., histotype, molecular profile, PS, chronic obstructive pulmonary disease class, and anatomic disease extent). The relative value of classifications of various characteristics of a tumor or a patient varies, depending on the situation at hand and the potential treatment. Expecting a single classification scheme to encompass all relevant aspects in all situations is unrealistic; focusing on those classifications that are relevant to the situation is more practical.

A prognostic model is needed, ideally tailored to provide individualized prediction. This is fundamentally different from stage classification,<sup>1</sup> despite the obvious overlap. Developing such a system is fraught with far greater complexity than the mere stage classification and will be addressed separately.

In conclusion, what is described in this article is a general strategic plan that seeks to provide a robust method of refining the stage classification and address weaknesses as much as possible. We recognize that intentions are not always matched by reality. Nevertheless, it is a framework that provides a plan toward a goal. This article is intended to articulate details of the strategy for the sake of alignment and transparency, but also with the hope for engagement; we welcome criticism and efforts to improve the process. The SPFC is confident that this course will move us forward. At any rate, the SPFC efforts to refine stage classification of thoracic malignancies are an advance over the initial proposals, which were largely empirical.

## CRedit Authorship Contribution Statement

**Frank C. Detterbeck:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision.

**Katherine K. Nishimura, Vanessa J. Cilento:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing.

**Meredith Giuliani:** Writing - review & editing.

**Mirella Marino, Ray U. Osarogiagbon:** Conceptualization, Methodology, Writing - review & editing.

**Ramon Rami-Porta, Valerie W. Rusch:** Conceptualization, Methodology, Writing - review & editing, Supervision.

**Hisao Asamura:** Conceptualization, Methodology, Writing - review & editing, Supervision.

**All other authors (listed in Appendices):** Writing - review & editing.

## Acknowledgments

The work of Dr. Valerie Rusch is supported in part by the National Institutes of Health/National Cancer Institute Cancer Center Support Grant P30 CA008748.

## 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/](http://www.jto.org/) and at <https://doi.org/10.1016/j.jtho.2022.02.008>

## References

1. Detterbeck F. 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, Inc.; 2001:95-104.
3. Goldstraw P, Crowley J. The International Association for the Study of Lung Cancer international staging project on lung cancer. *J Thorac Oncol.* 2006;1:281-286.
4. 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.
5. Detterbeck F. Developing a prognostic prediction model for lung cancer. In: Rami-Porta R, ed. *Staging Manual in Thoracic Oncology*. 2nd ed. North Fort Meyers, FL: Editorial Rx Press; 2016:237-258.
6. Kattan M, Hess K, Amin M, et al. American Joint Committee on Cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. *CA Cancer J Clin.* 2016;66:370-374.
7. 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.* 2007;2:694-705.
8. Detterbeck F, Chansky K, Groome P, et al. The IASLC Lung Cancer Staging Project: methodology and validation used in the development of proposals for revision of the stage classification of non-small cell lung cancer in the forthcoming (8th) edition of the TNM classification of lung cancer. *J Thor Oncol.* 2016;11:1433-1446.
9. Detterbeck FC, Gibson CJ. Turning gray: the natural history of lung cancer over time. *J Thorac Oncol.* 2008;3:781-792.
10. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 1999;130:515-524.