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

Patti A. Groome, PhD, Vanessa Bolejack, MPH,† John J. Crowley, PhD,† Catherine Kennedy, RMRA,‡ Mark Krasnik, MD,§ Leslie H. Sobin, MD,‖ and Peter Goldstraw, FRCS¶ on Behalf of the International Staging Committee, Cancer Research and Biostatistics, Observers to the Committee and Participating Institutions*

Introduction: In 1996, the International Association for the Study of Lung Cancer (IASLC) launched a worldwide TNM staging project to inform the next edition (seventh) of the TNM lung cancer staging system. In this article, we describe the methods and validation approaches used and discuss the internal and external validity of the recommended changes.

Methods: The International Staging Committee agreed on a number of general principles that guided the decision-making process. Internal validity was addressed by visually assessing the consistency of Kaplan-Meier curves across database types, geographic regions and addressing external validity, by assessing the similarity of curves generated using the population-based Surveillance Epidemiology and End Results cancer registry data to those generated using the project database. Cox proportional hazards regression was used to calculate hazard ratios between the proposed stage groupings with adjustment for cell type, sex, age, and region.

Results: Calls for data by the International Staging Committee resulted in the creation of an international database containing information on more than 100,000 cases. The present work is based on analyses of the 67,725 cases of non-small cell lung cancer. Validation checks were robust, demonstrating that the suggested staging changes are stable within the data sources used and externally. For example, suggested changes based on tumor size were well supported, with statistically significant hazard ratios ranging from 1.14 to 1.51 between adjacent pairs in the Surveillance Epidemiology and End Results data.

Conclusions: Lung cancer stage definitions have never been subjected to such an intense validation process. We do accept, however, that this work is limited in ways that can only be addressed by a prospective database, which we intend to develop. In the meantime, we think that this new system will greatly improve the usefulness of TNM lung staging across all of its purposes.

Key Words: TNM classification, Non-small cell lung cancer, Staging validity, International database.

(*J Thorac Oncol.* 2007;2: 694–705)

*Queen's Cancer Research Institute, Kingston, Ontario, Canada; †Cancer Research and Biostatistics, Seattle, Washington; ‡University of Sydney, Australia; §Gentofte University Hospital, Copenhagen, Denmark; ‖Armed Forces Institute of Pathology, Washington, DC; ¶Royal Brompton Hospital, Imperial College, London, United Kingdom.

See Appendix for participants.

Disclosure: This work was funded by a restricted educational grant from Eli Lilly and Company. No individual from the company had any role in evaluating the data or in preparing the manuscript. The project was also supported by the AJCC grant "Improving AJCC/UICC TNM Cancer Staging."

A supplementary Appendix of internal validation figures is available via the ArticlePlus feature at www.jto.org. Please go to the August issue and click on the ArticlePlus link posted with the article in the Table of Contents to view this material.

Address for correspondence: Patti A. Groome, PhD, Division of Cancer Care and Epidemiology, Queen's Cancer Research Institute, 10 Stuart Street, Level 2, Kingston, Ontario, Canada K7L 3N7. E-mail: patti.groome@krcc.on.ca

ISSN: 1556-0864/07/0208-0694

In 1996, the International Association for the Study of Lung Cancer (IASLC) launched a worldwide TNM staging project to create international databases that would be used to continue the excellent efforts of Dr. Cliff Mountain, who pioneered this approach to lung cancer staging in 1973.[1–3]

Successive iterations of TNM staging for lung cancer have addressed shortcomings identified by the oncology community. Similarly, the IASLC recognized that it is important that further revisions continue to be made to ensure that the international staging system for lung cancer remains fit for its purpose. The work of the International Staging Committee (ISC) that oversaw the conduct of this study will inform the seventh edition of the international staging system for lung cancer.

The ISC identified the following issues that needed to be rectified:

1. There was a lack of validation for individual T, N, and M descriptors in previous iterations.
2. The relatively small database on which previous revisions have been based made it unlikely that the individual descriptors had been adequately assessed.
3. This previously used database was recruited from a limited geographic area and was predominately composed of surgical cases.
4. Criticisms in the literature needed to be addressed and reconciled.

This article, written by the Validation and Methodology Subcommittee of the ISC, describes the methods and validation approach used in this retrospective phase of the ongoing IASLC staging initiative and discusses the internal and external validity of the T, N, and M category changes and the stage groupings that are being recommended by the ISC.

METHODS

The IASLC is the only global organization dedicated to the study of lung cancer, with members involved in all aspects of lung cancer diagnosis, imaging, management, and research. Therefore, the IASLC thought that it was ideally placed to organize a large, international database of lung cancer cases collected from diverse geographic areas and managed by all modalities of care. The ISC that oversaw the conduct of this study was established in 1998. Further details regarding the project development have been described elsewhere.[3] At a meeting held in London in 2001, representatives from 23 institutions in 12 countries presented data from their individual databases representing in excess of 80,000 cases. Calls for data eventually increased this number to more than 100,000.

As the project progressed, the ISC created subcommittees to manage the various elements of the project[3]:

1. T Descriptors Subcommittee
2. N Descriptors Subcommittee
3. M Descriptors Subcommittee
4. Small-Cell Lung Cancer
5. Nodal Chart Subcommittee
6. Prognostic Factors Subcommittee
7. Validation and Methodology Subcommittee

Cancer Research and Biostatistics, a nonprofit organization based in Seattle, WA, with extensive experience in the conduct and analysis of multicenter studies in North America, is managing and analyzing the project database. Cancer Research and Biostatistics has been responsible for collecting, translating, and compiling the data, creating a data dictionary, and providing the subcommittees with analyses that have informed our recommendations.

Guiding Principles

The ISC agreed on a number of general principles that guided the decision-making process throughout this project:

1. Recommendations should be based on cTNM (clinical TNM) and pTNM (pathologic TNM) to ensure relevancy to all who are diagnosed with lung cancer.
2. Changes should not compromise the use of data from the previous staging system whenever possible.
3. Evidence from external sources should be included in our deliberations.
4. Unproven descriptors should be "flagged" for a later prospective phase.

In a subsequent breakout session, the Validation and Methodology Subcommittee recommended the following additional principles:

5. Design a system that is identical for both pathologic and clinical staging.
6. If boundaries between categories are to be changed, there should be overwhelming evidence to support such changes. This criterion is in recognition of TNM General Rule 6.[4]
7. Stage grouping boundaries are not constrained by condition 2 listed above.
8. Surgical cases where there was residual disease at the resection margins, being either microscopic (R1) or macroscopic (R2), should be included in the analyses, provided there were no significant changes to the pathologic staging conclusions
9. The prognostic ability of the revised staging system should be verified by a multivariate analysis that considers other prognostic factors, i.e., histology, age, sex.

Study Population

Collaborating institutions were recruited through announcements in the journal *Lung Cancer*, by presentations at the IASLC World Conferences and at other conferences and workshops. Institutions known to have data were contacted directly by members of the ISC. Data were accepted from all parts of the globe for all modalities of care, including best supportive care. The time frame was defined as all patients treated between the beginning of 1990 and the end of 2000. Cases were screened for adequate follow-up, histology, and baseline TNM staging. Methods used to compile the database are described in detail elsewhere.[3]

The total number of patients submitted to Cancer Research and Biostatistics was 100,869. Of these, 81,015 passed the initial screening requirements of having a new diagnosis rather than a diagnosis of recurrent disease, of either small-cell lung cancer (SCLC) or non-small cell lung cancer (NSCLC), adequate follow-up for survival calculations, and a complete set of either cTNM or pTNM at baseline. Results of the initial screening requirements are summarized in Table 1, updating our previous report.[3]

Of the 81,015 patients passing the initial screen, 36% were treated with surgery only, 11% with radiotherapy alone, 21% with chemotherapy alone, 9% with best supportive care or no treatment, and the remaining patients were managed with combined treatment modalities. The distributions by stage and geographic region for included cases of SCLC and NSCLC are given in Figure 1.

A small number of cases had been submitted from series or registries and also from clinical trial groups, which may have caused duplication of cases. Where identifiers

TABLE 1. Initial Screening of Submitted Cases

Total cases submitted	100,869
Passed initial screen for SCLC and NSCLC analyses	81,015
Excluded	19,854
Carcinoid	546
Other tumors (sarcoma, other)	569
Other reasons	
Outside 1990–2000 time frame	5443
Incomplete survival data	1505
Unknown histology or occult	2468
Incomplete stage information	7720
Recurrent case (or unknown status)	1536
Duplicate cases removed	67

SCLC, small-cell lung cancer; NSCLC, non-small cell lung cancer.

permitted or in regions where specific clinical trial participation was known, double entries were excluded. However, it was not possible to quantify accurately all such cases.

The study population that passed the initial screening criteria contained 67,725 cases of NSCLC and 13,290 cases of SCLC. The two groups were analyzed separately, and as the focus of this article is the NSCLC cases, the SCLC cases have been excluded from further discussion.

The NSCLC study population included 53,640 clinically staged cases and 33,933 pathologically staged cases, with 20,006 cases both clinically and pathologically staged.

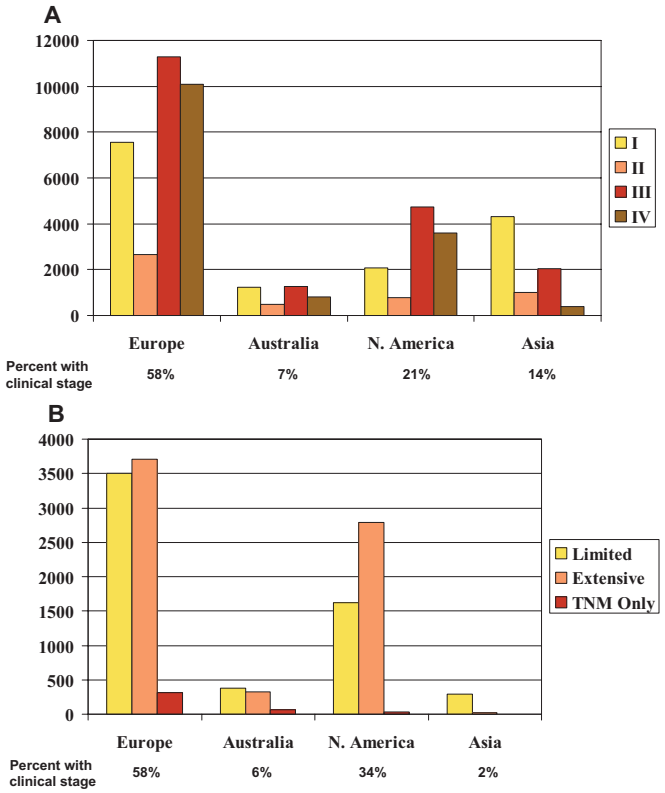


FIGURE 1. Stage distribution by continent non-small cell lung cancer (*A*) and small-cell lung cancer (*B*).

The groups who contributed these NSCLC cases to the project are summarized in Table 2, classified according to the definitions listed.

1. Clinical trial: participants in studies designed to investigate alternate treatments
2. Population-based registry: all individuals diagnosed with lung cancer in a defined region, including those diagnosed at death
3. Institutional registry: all individuals diagnosed with lung cancer and admitted to a particular institution
4. Consortium: same as institutional registry but spanning multiple institutions
5. Series: all individuals diagnosed with lung cancer and treated by a particular doctor or service
6. Surgical series: all individuals diagnosed with lung cancer and treated by a particular surgeon or surgical unit

The geographic source of the 67,725 NSCLC patients by continent was as follows: 40,059 from Europe, 12,178 from North America, 10,216 from Asia, and 5,272 from Australia. Ninety-five percent were followed until death or at least 2 years; and 88%, until death or 5 years. Median follow-up for the 17,754 patients alive at last contact was 5.3 years.

Endpoints and Statistical Methods

Survival was measured from the date of entry (date of diagnosis for registries, date of registration for protocols) for clinically staged data and the date of surgery for pathologically staged data to the date of death or last contact and was calculated by the Kaplan-Meier method.[5] For the external validation analyses, Cox proportional hazards regression was used to calculate hazard ratios (HRs) between adjacent groups and to assess the prognostic value of the proposed stage groupings in both the test set and the external validation. In that instance, given the large number of cases available, adjustment for cell type, sex, age, and region was conducted. All analyses were conducted using the SAS System for Windows Version 9.0.

Validation Methodology

Because the data collected were from diverse sources and are not, strictly speaking, population based, we were concerned that conclusions reached may be biased. The Validation and Methodology Subcommittee was formed to address this concern by making recommendations about a validation approach and interpreting the validation analyses. Specifically, the mandate of this committee was to ensure that the recommendations for change were internally and externally valid and that the project would meet Union Internationale Contre le Cancer methodological expectations.

The internal validation approach was to compare results of interest among types of databases (consortium/surgical series versus clinical trials versus series/registries) and among geographic regions (North America versus Australasia versus Europe). If the direction and magnitude of effects were relatively consistent within these subgroups, that would be considered evidence validating the results. Survival curves were visually compared using all cases with the relevant

TABLE 2. Screened NSCLC Cases by Type of Contributing Group

Clinical Trial Groups		Registries		Consortia		Surgical Series	
MacCallum	183	Amsterdam	8897	Japan	6931	China	1732
MRC	1659	Flemish	3590	IFCT	2539	Korea	832
IFCT	920	Rotterdam	1133	GCCB-S	2894	Sydney	1572
ELCWP	1385	**Institutional Registry**		**Series**		Prince Charles	773
IALT	1867	Heidelberg	4455	Taiwan	721	St. Vincent's	17
SLCG	438	**Surgical Registry**		QRI	2452	Gdansk	1231
EORTC	1123	Norway	2112	Western	275	Torino	1137
CALGB	1830			Faculty Hospital, Plzen	1486	Grenoble	677
NCCTG	1111			Leuven	770	Ankara	543
ECOG	1737			Jules-Bordet	547	Belgrade	344
SWOG	1859			MDACC-RT	840	Warsaw	213
RTOG	1768			Johns Hopkins University	851	Perugia	99
NCIC	550					MSKCC	880
						MDACC-TCVS	489
						Prince Margaret	191
						Wayne State University	72

MRC, Medical Research Council; IFCT, Intergroupe Francophone deCancerologie Thoracique; GCCB-S, Bronchogenic Carcinoma Co-operative Group of the Spanish Society of Pneumology and Thoracic Surgery; ELCWP, European Lung Cancer Working Party; IALT, International Adjuvant Lung Cancer Trial; SLCG, Spanish Lung Cancer Group; EORTC, European Organization for Research and Treatment of Cancer; CALGB, Cancer and Leukemia Group B; HSP, NCCTG, North Central Cancer Treatment Group; ECOG, Eastern Cooperative Oncology Group; SWOG, Southwest Oncology Group; RTOG, Radiation Therapy Oncology Group; NCIC, National Cancer Institute of Canada; QRI, Queensland Radium Institute; MDACC-RT, M.D. Anderson Cancer Center–Radiation Therapy; MDACC-TCVS, M.D. Anderson Cancer Center–Thoracic and Cardiovascular Surgery; MSKCC, Memorial Sloan-Kettering Cancer Center.

information and also comparing pTNM curves separately from cTNM curves. We did not conduct statistical tests during the validation stage as the numbers often varied widely between data sources, thereby compromising the usefulness of statistical comparisons. For internal validation of the T component analysis by size, a training set of two thirds of the cases was used for testing and validated in the other one third of cases. Similarly, internal validation of proposed stage groupings was conducted by holding out a validation set of one third of the cases.

For external validation, cases of NSCLC diagnosed from 1998 to the end of 2000 from the Surveillance, Epidemiology, and End Results Program (SEER) database were chosen.[6] This time frame offered the most detailed extent of disease data. Given the nature of these data, we assigned stage using pTNM when available and cTNM otherwise, that is, using best stage. Hazard ratios and significance levels were calculated for pairwise comparisons of adjacent categories.

In general, our conclusions discounted deviant results when they were based on unstable curves that were the result of small numbers. Our overarching goal was to ensure that no large, deviant subset drove our final recommendations and that whenever it was possible (i.e., the data were available), our results were reflected in our population-based external data set.

RESULTS

Internal and External Validity of Suggested Changes to the T Categories

The following are observations of the Validation and Methodology Subcommittee, in support of the recommendations of the T Subcommittee. Table 3 contains a summary of the Kaplan-Meier survival results and Figures 2 and 3 show the results of our external validation analyses. A supplemental appendix of the Kaplan-Meier survival curves for the internal validity comparisons is available on this journal's Web site. Changes referred to below are described in the paper on the ISC T-category recommendations.[7]

Change 1: Subclassify T1 as T1a and T1b by Size

On balance, this finding is both internally and externally valid. Referring to Table 3, we found that the pT finding was clearly driven by the Australasian data, which made up 62% (2322 of 3579) of cases, with a median survival of 124 months for T1a and 103 months for T1b. The European cT data also showed distinct differences in survival with median survival 64 and 46 months, respectively. When we examined this split by database type, the series/registry subset (19% of pN0 cases used) was not as distinct as the consortia/surgical subset with differences in median survival of 6% and 16%, respectively. Clinical staging was available on only 3283 subjects, whereas pathologic stage was available for 9007. Referring to the clinically staged results, the North American clinical trial and registries/series subsets, which made up 23% of the whole, were not able to distinguish between these groups. As shown in Figure 2, the SEER external validation in the N0 population supports the split with a 4% difference at 5 years and an HR of 1.27 ($p = 0.04$) and in the surgical subset of the SEER, whose staging would have been pathologic, the HR was 1.16 ($p = 0.02$) (data not shown).

TABLE 3. Median Overall Survival in Months among Subgroups Compared in the Conduct of Internal Validity Checks for Proposed T Category Changes

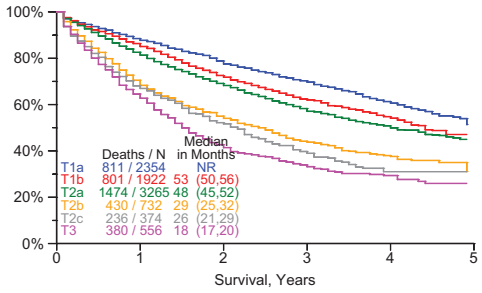
T Category Changes 1–3

	Region						Database Type					
	Clinical Stage			Pathological Stage			Clinical Stage			Pathological Stage		
cN0 and pN0 Only	NA (n = 473)	A/Au (n = 284)	Eur (n = 2526)	NA (n = 1073)	A/Au (n = 4737)	Eur (n = 3197)	C/S (n = 2537)	R/S (n = 465)	CTs (n = 281)	C/S (n = 7125)	R/S (n = 1697)	CTs (n = 185)
T1a, ≤2 cm	59	74	64	112	124	92	69	66	81	119	95	—
T1b, −3 cm	74	67	46	95	103	84	46	69	NR	103	89	—
T2a, −5 cm	77	38	38	93	89	64	38	40	77	87	57	NR
T2b, −7 cm	60	41	27	66	68	43	28	30	64	62	39	NR
X, >−7 cm	47	20	15	52	34	25	15	20	47	31	25	77
T3	35	18	16	39	38	25	17	16	33	33	25	—

T Category Changes 4 and 5

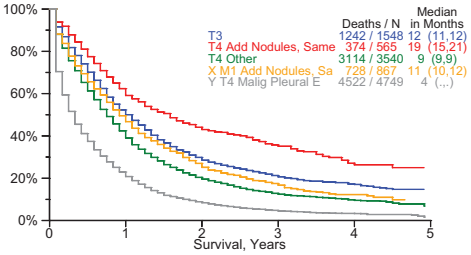
	Region Pathologic Stage			Database Type Pathologic Stage	
Any pN	NA (n = 81)	A/Au (n = 1403)	Eur (n = 868)	C/S (n = 2151)	R/S (n = 201)
T3	25	28	20	24	27
T4 additional nodules only	21	21	21	21	21
T4 by other factor	43	14	16	15	12
X additional nodules same side	NR	20	11	19	18
Y pleural dissemination	—	19	13	18	11

NA, North America; A/Au, Asia and Australia; Eur, Europe; C/S, Consortia and Surgical Series; R/S = Registries and Series; CTs, clinical trials; NR, median survival not reached.



	1 Yr	5 Yrs	Comparison	HR	P
T1a	88%	51%			
T1b	85%	47%	vs T1a:	1.27	<.0001
T2a	81%	45%	vs T1b:	1.14	0.0039
T2b	68%	31%	vs T2a:	1.51	<.0001
T2c	67%	31%	vs T2b:	1.15	0.0924
T3	63%	26%	vs T2c:	1.18	0.0464

FIGURE 2. Overall survival of T1–T3 N0 by size of tumor in the Surveillance, Epidemiology, and End Results Program and results of pairwise comparisons. HR, hazard ratio.



	1 Yr	5 Yrs	Comparison	HR	P
T3	50%	15%			
T4 Add Nodules, Same Lobe	59%	25%	vs T3:	0.70	<.0001
T4 by Other Factor	39%	7%	vs T4 Same Lobe:	1.88	<.0001
M1 Add Nodules, Same Side	47%	10%	vs Other T4:	0.86	0.0002
T4 Pleural Dissemination	21%	2%	vs Other T4:	1.72	<.0001

FIGURE 3. Overall survival of T3–T4 or M1 by same-side nodules in the Surveillance, Epidemiology, and End Results Program and results of pairwise comparisons.

Change 2: Subclassify T2 as T2a and T2b by Size

This finding was consistent across regions with the exception of the Australasian clinical results, where the numbers were small ($n = 145$) and the North American pathologic results, where the curves do not split until 3 years. Both the pathologically staged results and the consortia/surgical clinically staged results are consistent and stable, but the number of clinically staged T2b cases was small in the clinical trial and registry/series sources. Overall, for those instances in which the curves were stable, the difference in median survival ranged from 10% to 27% (Table 3). The SEER external validation in the N0 population strongly supports the split both overall (Figure 2, 14% difference at 5 years, HR = 1.51, $p < 0.0001$) and in the surgical subset (15% difference at 5 years, HR = 1.45, $p < 0.0001$, data not shown).

Change 3: Reclassify T2 Tumors Larger than 7 cm as T3

Referring to Table 3, median survival differences between those subjects with previously staged T2 tumors larger than 7 cm and those with T3 disease were generally small, ranging from −2% to a 14% difference. Where the differences were greatest, the numbers were small: 21 clinically staged and 46 pathologically staged T2 >7 cm cases in the North American subset, and 17 clinically staged T2 >7 cm cases in the clinical trial subset. The population-based SEER data indicate some similarity between the T2b group and the T2 >7 cm group (HR = 1.15, $p = 0.09$), but the T2 >7 cm and T3 curves are statistically significantly different with an HR of 1.18 ($p = 0.05$) observed, and this finding is more stable when all SEER cases (including N+) are considered (HR = 1.27, $p < 0.0001$, data not shown).

Change 4: Reclassify T4 Tumors by Additional Nodules in the Primary Lobe as T3

Referring to Table 3, with the exception of the North American subset, which was small, cases with additional nodules in the primary lobe experienced better survival than other sixth edition T4 cases, with differences in median survival of 7% in the Australasian group and 5% in the European group. Median survival differences were similar by database type: 6% in the consortia/surgical group and 9% in the registry/series group. We were not able to look at this issue using cTNM because this data element was only available from the surgical data. Referring to the SEER data presented in Figure 3, we can see an extremely large difference in survival between those with T4 disease due to additional nodules in the same lobe compared with the rest (HR = 1.9, $p < 0.0001$). When we restricted to those treated with surgery in the U.S. SEER data, the additional nodule group's survival was better than those even with sixth edition T3 disease (5-year survival rate, 40% and 27%, respectively, data not shown). Note that the SEER Program T3 group excluded those with disease in the adjacent rib.

Change 5: Reclassify M1 by Additional Nodules in the Ipsilateral Lung (different lobe) as T4

This finding was driven by the consortia/surgical pathologic data source ($n = 168/180$) that was also almost all from the Australasian region ($n = 136/180$) in which median survival between the additional nodules group was 4% and 6% higher than the "Other T4" patient group. The European pathologic data support the change ($n = 43$) in that the curve for these patients is quite similar to the "Other T4" group (data not shown), although median survival is 5% lower. Referring to Figure 3, the SEER data indicate that the 11-month median survival of this group falls above the other T4 cases (9 months) with a HR of 0.86 ($p = 0.0002$).

Validation of the Decision to Make No Changes to the N Categories

The following are observations of the Validation and Methodology Subcommittee in support of the recommendation of the N Subcommittee to retain the current N category definitions. Table 4 contains a summary of the Kaplan-Meier

TABLE 4. Median Overall Survival in Months among Subgroups Compared in the Conduct of Internal Validity Checks for Investigation of TNM Sixth Edition N Categories, cN0–cN3, cM0 Only

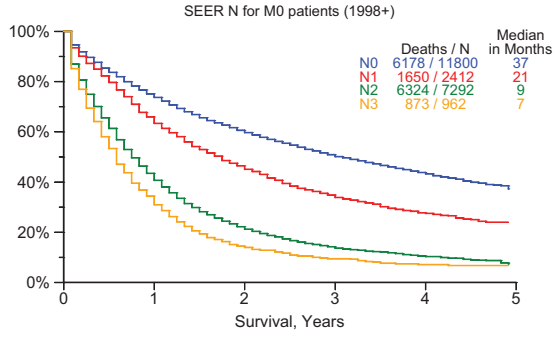
	Region			Database Type		
	NA (n = 6942)	A/Au (n = 10,220)	Eur (n = 21,100)	C/S (n = 14,190)	R/S (n = 16,445)	CTs (n = 7626)
cN0	43	70	29	62	27	23
cN1	24	32	18	38	17	18
cN2	17	20	12	22	11	14
cN3	12	11	8	12	8	10

NA, North America; A/Au, Asia and Australia; Eur, Europe; C/S, Consortia and Surgical Series; R/S, Registries and Series; CTs, clinical trials.

survival results comparing the existing TNM N categories by region and database type, and Figure 4 shows the results of our external validation analyses of those categories. A supplemental appendix of the Kaplan-Meier survival curves for the internal validity comparisons is available on this journal's Web site. Further details regarding this recommendation are provided in the paper on the ISC N-category recommendations.[8]

The prognostic value of the TNM sixth edition N categories were validated internally by comparing the survival curves of cM0 cases across geographic regions and database types in the IASLC database. All subsets of the IASLC data set generated N category survival curves that were prognostically distinct. However, the degree of spread among the curves varied, likely reflecting variability in case mix within N categories among the geographic regions and data types. Referring to Table 4, the cN0 to cN3 median survival difference ranged from 21 months in Europe to 59 months in Australasia. The difference in median survival between cN0 and cN3 was smaller in the clinical trials data (13 months) compared with registries (19 months) and consortia/series (50 months). The position of the curves also varied. For example, the median survival of cN0 patients was 29 months in Europe, 43 months in North America, and 70 months in Asia.

External validation was assessed using the SEER data presented in Figure 4. Each curve is prognostically distinct, although the difference between the N2 and N3 curves is smaller than what we observed in the IASLC data set overall[8] (median survival difference of 2 months in SEER and 5 months in IASLC). The paper presenting the ISC N-category recommendations proposes areas for further investigation when more detailed data become available.[8]



	1 Yr	5 Yrs	Comparison	HR	P
N0	74%	37%			
N1	63%	24%	vs N0:	1.52	<.0001
N2	41%	7%	vs N1:	1.89	<.0001
N3	31%	7%	vs N2:	1.26	<.0001

FIGURE 4. Overall survival by TNM sixth edition N category in Surveillance, Epidemiology, and End Results Program (SEER) and results of pairwise comparisons.

Internal and External Validity of Suggested Changes to the M Categories

The following are observations of the Validation and Methodology Subcommittee in support of the recommendations of the M Subcommittee. Table 5 contains a summary of the Kaplan-Meier survival results comparing the existing TNM N categories by region and database type and Figure 5 shows the results of our external validation analyses of those categories. A supplemental appendix of the Kaplan-Meier survival curves for the internal validity comparisons is available on this journal's Web site. Changes referred to below are described in the paper on the ISC M-category recommendations.[9]

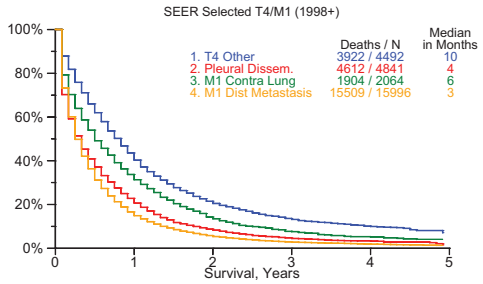
Change 1: Reclassify Pleural Dissemination (malignant pleural or pericardial effusions, pleural nodules) from T4 to M1

The distinction between this group and the other T4 patients was driven by the European and North American data in the clinically staged analyses. In these regions, the median survival difference between those with pleural dissemination and those with any T4 M0 disease was 5% and 6%, respectively (Table 5), and the curves are divergent throughout follow-up (see online supplemental appendix). The finding was consistent in two of the three data types, and the curves diverge after the first year in the registry/series group (see online supplemental appendix). Referring to Figure 5, the SEER data support the external validity of this finding with 4-month median survival for those patients with malignant pleural effusions compared with 10 months for all other T4s.

TABLE 5. Median Overall Survival in Months among Subgroups Compared in the Conduct of Internal Validity Checks for Proposed M Category Changes

	Region			Database Type		
	NA (n = 2212)	A/Au (n = 317)	Eur (n = 3063)	C/S (n = 908)	R/S (n = 1973)	CTs (n = 2711)
T4 M0 any N	16	10	12	21	7	16
Pleural dissemination	10	8	7	7	7	10
Contralateral lung nodules	11	9	10	—	10	10
M1 distant	7	5	5	6	4	7

NA, North America; A/Au, Asia and Australia; Eur, Europe; C/S, Consortia and Surgical Series; R/S, Registries and Series; CTs, clinical trials.



	1 Yr	5 Yrs	Comparison	HR	P
T4 Other	40%	7%			
T4 Pleural Dissem.	21%	2%	vs T4:	2.81	<.0001
M1 Contra Lung	31%	4%	vs Pleural Dissem:	0.75	<0.001
M1 Dist Metastasis	15%	1%	vs Contra Lung:	1.52	<.0001
M1 Distant vs Pleural Dissem.				*1.14*	*<0.001*

FIGURE 5. Overall survival of T4 and M1 minus same-side nodules in Surveillance, Epidemiology, and End Results Program (SEER) and results of pairwise comparisons.

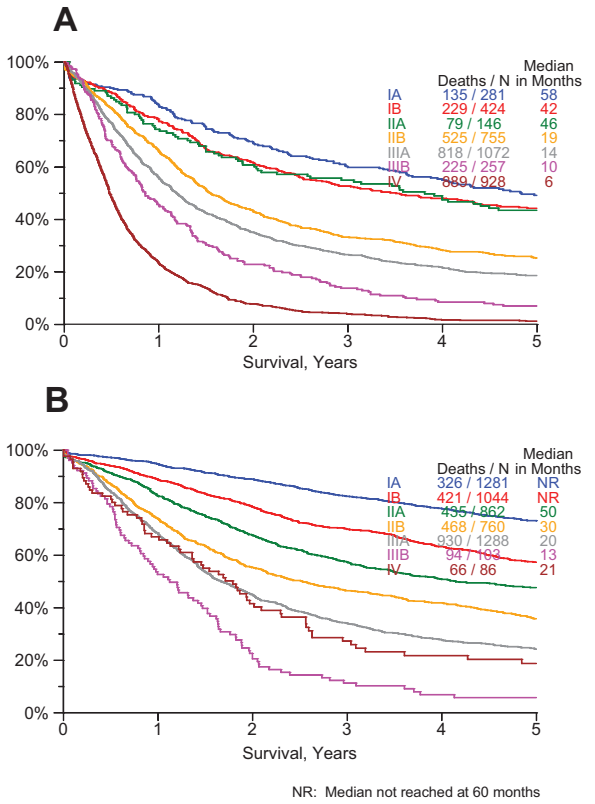


FIGURE 6. Overall survival by proposed TNM stage groupings, validation data set, clinical stage (*A*), and pathologic stage (*B*).

Change 2: Subclassify M1 into M1a and M1b

Referring to Table 5, those patients with metastatic disease attributable to distant metastases had a worse median survival (4–7 months) than those with pleural dissemination (7–10 months) or those with disease spread to the contralateral lung (9–11 months). The ordering of these findings was consistent across regions and database types (see online supplemental appendix). Referring to Figure 5, the SEER data show less separation between patients with pleural dissemination versus patients with distant metastases; however, the ordering matches that of the IASLC data that are the basis of this recommendation.[9]

Internal and External Validity of the Stage Grouping System

The following are observations of the Validation and Methodology Subcommittee, in support of the recommendations of the IASLC International Staging Committee. Figures 6 and 7 and Tables 6 through 9 present the results of the validation analyses of these proposed stage groupings. Changes referred to below are described in the paper on the ISC TNM stage grouping recommendations.[10] Note that these changes are proposed partly to improve prognostication but also to ensure that the groups are clinically sensible with regard to treatment considerations. This aspect of the decision-making process is described in the companion paper.[10]

The internal validity of the proposed TNM grouping scheme was assessed via the subset of data reserved as the validation data set. Tables 6 through 9 summarize the results of Cox proportional hazards regression applied to the validation set data, modeling the proposed grouping scheme with adjustment for cell type, sex, age, and region. A parallel analysis of the TNM, sixth edition (TNM 6) applied to the same data set is included for reference. The proposed group-

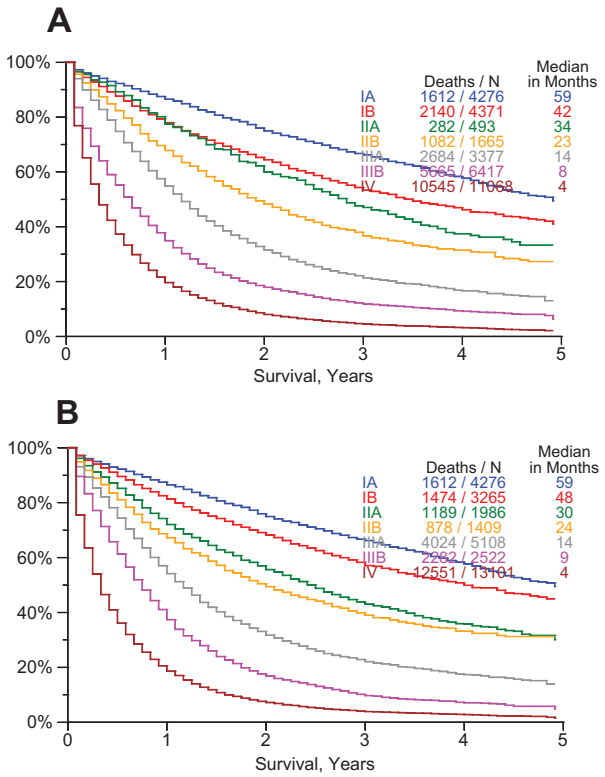


FIGURE 7. Overall survival by best stage in SEER: TNM sixth edition (*A*) and IASLC proposed (*B*).

TABLE 6. Cox Proportional Hazards Regression Models for TNM 6 and Proposed Clinical Stage Groupings (IASLC), Validation Data Set: Clinical Stage (TNM 6 and IASLC Proposed) as Indicator Variables

	HR for Comparison		*p*	
Comparisons	**TNM 6**	**IASLC**	**TNM 6**	**IASLC**
IB vs. IA	1.14	1.07	0.1886	0.5330
IIA vs. IB	1.12	1.03	0.7965	0.8178
IIB vs. IIA	1.46	1.72	0.3977	<0.0001
IIIA vs. IIB	1.36	1.34	<0.0001	<0.0001
IIIB vs. IIIA	1.42	1.45	<0.0001	<0.0001
IV vs. IIIB	1.87	1.83	<0.0001	<.0001
R^2	26.35	27.15		

Adjusted for cell type, sex, age, and region; $n = 3863$ (2999 events). TNM 6, TNM sixth edition; HR, hazard ratio; IASLC, International Association for the Study of Lung Cancer.

TABLE 7. Cox Proportional Hazards Regression Models for TNM 6 and Proposed Clinical Stage Groupings (IASLC), Validation Data Set: Clinical Stage (TNM 6 and IASLC Proposed) as an Ordered Variable

	HR		*p*	
Variable	**TNM 6**	**IASLC**	**TNM 6**	**IASLC**
Stage	1.40	1.43	<0.0001	<0.0001

Adjusted for cell type, sex, age, and region; $n = 3863$ (2999 events). TNM 6, TNM sixth edition; HR, hazard ratio; IASLC, International Association for the Study of Lung Cancer.

TABLE 8. Cox Proportional Hazards Regression Models for TNM 6 and Proposed Pathological Stage Groupings (IASLC), Validation Data Set: Pathologic Stage (TNM 6 and IASLC Proposed) as Indicator Variables

	HR for Comparison		*p*	
Comparisons	**TNM 6**	**IASLC**	**TNM 6**	**IASLC**
IB vs. IA	1.75	1.57	<0.0001	<0.0001
IIA vs. IB	1.17	1.33	0.1498	<0.0001
IIB vs. IIA	1.30	1.38	0.0170	<0.0001
IIIA vs. IIB	1.54	1.47	<0.0001	<0.0001
IIIB vs. IIIA	1.28	2.00	0.0004	<0.0001
IV vs. IIIB	0.79	0.65	0.1269	0.0063
R^2	29.32	30.76		

TNM 6, TNM sixth edition; HR, hazard ratio; IASLC, International Association for the Study of Lung Cancer. Adjusted for cell type, sex, age, and region; $n = 5424$ (3016 events).

TABLE 9. Cox Proportional Hazards Regression Models for TNM 6 and Proposed Pathologic Stage Groupings (IASLC), Validation Data Set: Pathologic Stage (TNM 6 and IASLC Proposed) as an Ordered Variable

	HR		*p*	
Variable	**TNM 6**	**IASLC**	**TNM 6**	**IASLC**
Stage	1.35	1.41	<0.0001	<0.0001

TNM 6, TNM sixth edition; HR, hazard ratio; IASLC, International Association for the Study of Lung Cancer. Adjusted for cell type, sex, age, and region; $n = 5424$ (3016 events).

ings were prognostically distinct using clinical stage for pairwise comparisons from stage IIA on, but not for comparisons between the early staged groups (Tables 6 and 7). Of note, the same failure occurred for TNM 6 and may be due to a peculiarity of the smaller subset of early stage cases in the clinically staged validation data set. The validation exercise did reproduce the improvement achieved by the proposed subgroupings in the separation of stages IIA and IIB for clinical stage.

For pathologic stage (Tables 8 and 9), application of the proposed subgroupings to the validation data set achieved the same general results as when applied to the training set and to all available cases as reported previously.[10] Improvements in the R^2 value for the proposed grouping scheme are small but persistent in the validation set analyses, both for clinical and pathologic staging. Survival curves for the proposed TNM subgrouping generated on the validation set data are shown for clinical and pathologic stages in Figure 6.

External validation was assessed via SEER best stage as shown in Figure 7. Both TNM 6 and the newly proposed system perform as expected but with some deficiencies.

For TNM 6, the separation between stages IB and IIA is weak out to 2 years. For the proposed system, stages IIA and IIB nearly converge at 5 years. For either system, stages IIIB and IV are ordered appropriately and separate well. The proposed system results in a more even distribution of cases among the stage groupings in this population-based registry, primarily due to a larger proportion of cases assigned to stage IIA.

DISCUSSION

The currently accepted staging system for NSCLC was adopted in 1997 by the American Joint Committee on Cancer and the Union Internationale Contre le Cancer,[4] as a response to the need for more specific patient groupings. The intent of this staging system was to provide a consistent and reproducible classification for describing the extent of disease. At the same time, it was intended to provide a prognostic tool to guide clinicians in their treatment choices, as well as being a common language for lung cancer clinicians and researchers throughout the world.

Numerous articles have suggested deficiencies in the current international staging system, and none of these were addressed by TNM 6.[4] For instance, concerns have been expressed in regard to a number of descriptors; most notably, the possibility that size criteria other than the 3-cm cutoff that separates T1 and T2 tumors may be significant.[11–14] Two recent articles using a Spanish database suggest that this cutoff be replaced with cuts at 2, 4, and 7 cm and that tumors >7 cm be considered T3.[12,13] Other studies support the 3-cm cutoff and add an additional cut at 5 cm.[13,14] This example underscores the controversies that have been discussed in the literature and it also supports the recommended changes to T stage by size made by the ISC.[7] In addition, the appropriateness of the designation given to additional pulmonary nodules in certain locations has been questioned,[15,16] and there have been suggestions for change of the stage groupings.[17,18] However, the validity of all the studies, individually and collectively, is limited by one or more shortcomings, which cause them to fall below the standards set by the Union Internationale Contre le Cancer Process for Change Subcommittee.[19]

The purpose of the ISC over the past 5 years has been to address these deficiencies using the power of numbers accrued in our international database, together with sophisticated statistical methods and expert advice from members of the global oncology community. This has been a remarkable initiative, but we have some concerns about the limitations of the data. First, several significant geographic areas were not represented. These include the old USSR, South America, Africa, and China. Second, because this was an opportunistic use of existing data sets, we had to make use of data that had been collected for other purposes. So some data fields were missing from each of the data sets. Because we needed extent of disease information beyond TNM stage, the availability of these data in particular varied across the various data sets. As much as possible, the ISC considered the degree of coverage of each variable in its deliberations. Third, the precision of clinical staging, and even pathologic staging, may be subject to variability. This variability will almost always linked to local practices, which determine what level of evidence is required to assign a stage grouping. Although it is not possible to measure the impact of inconsistent staging practices, we do want to point out that such variation is actually reflective of what goes on in practice. That said, some prognostic differences may have been masked by staging inconsistencies in our data. In this context, we think that both random and systematic errors would have attenuated our ability to see prognostic differences and that our examination of the data by type and region has helped to ensure that no errors have occurred in the other direction, that is, declaring prognostic differences when they do not exist. This study is a classic example of the trade-off that often occurs between control over data quality and applicability of research findings to the widest group. Last, given the large number of data sources that we used and their wide geographic coverage, we expect that treatment practices and treatment quality would have varied. It was not possible for us to control for the prognostic impact of treatment for two reasons: (1) controlling for treatment in a multivariate analysis would have interfered with our ability to see extent of disease effects due to the problem of confounding by indication and (2) we had no information about treatment quality per se. However, our data did include patients treated with the full spectrum of treatment options, including no treatment and best supportive care. If there is a treatment-related bias in our findings, it is likely in the direction of reducing our ability to see prognostic differences.

We used the SEER registries' data as our external validation source. These registries are the reference standard by which other registries in the world are measured, and they contain detailed extent of disease information that allowed us to assess key aspects of TNM staging. Because SEER is population based, the entire spectrum of lung cancer presentation is represented. We are not aware of any published findings on the accuracy of the SEER extent of disease data. However, referring to our Figures 4 and 7A, which report the SEER data split by the current TNM N categories and the current TNM stage groups respectively, we find that the wide separation of these curves is evidence of the prognostic validity of the SEER extent of disease assignment.

In the next phase of its work, the ISC is planning to widen its geographic coverage and move into primary data collection that will include purposeful retrospective data capture up to 2007 and prospective data capture starting in 2008. With more control over data collection and better geographic coverage, we will be able to further inform lung cancer staging. Specifically, the goal of this next phase will be to make proposals for consideration in the eighth edition of the TNM classification.

CONCLUSION

We have, with successive revisions of the TNM staging system, become comfortable with established treatment algorithms that have become linked to designated stages of disease. Some of our recommendations will force us all to readdress these links and perhaps change some of these algorithms. However, we should remember that previous

revisions have been informed by a smaller database that was never subjected to the intense validation process that we have used. In this respect, our recommendations are more robust than any previous submission. We accept that it is still limited in ways that can only be addressed by the next phase of our work, but, in the meantime, we also believe that this new system for lung staging will greatly improve the usefulness of TNM lung staging across all its purposes.

ACKNOWLEDGMENTS

Eli Lilly and Company provided funding to support the work of the International Association for the Study of Lung Cancer Staging Committee to establish a database and to suggest revisions to the sixth edition of the TNM classification for Lung Cancer (staging) through a restricted grant. Lilly had no input into the Committee's analysis of the data or suggestions for revisions to the staging system.

REFERENCES

1. Mountain CF, Carr DT, Anderson WA. A system for the clinical staging of lung cancer. *Am J Roentgenol Radium Ther Nucl Med* 1974;120:130–138.
2. Mountain CF, Staging classification of lung cancer. A critical evaluation. *Clin Chest Med* 2002;23:103–121.
3. Goldstraw P, Crowley JJ, on behalf of the IASLC International Staging Project. The International Association for the Study of Lung Cancer International Staging Project on Lung Cancer. *J Thorac Oncol* 2006;1: 281–286.
4. Sobin LH, Wittekind Ch, eds. TNM Classification of Malignant Tumours, Sixth Edition . New York: John Wiley & Sons, 2002.
5. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457–81.
6. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence–SEER 13 Regs Public-Use, November 2004 Sub (1973–2002 varying), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2005, based on the November 2004 submission. Accessed September 21, 2006.
7. Rami-Porta R, Ball D, Crowley JJ, et al. The IASLC Lung Cancer Staging Project: Proposals for revision of the T descriptors in the forthcoming seventh edition of the TNM classification for lung cancer. *J Thorac Oncol* 2007; 7:593–602.
8. Rusch VW, Crowley JJ, Giroux DJ, et al. The IASLC Cancer Staging Project: Proposals for revision of the N descriptors in the forthcoming seventh edition of the TNM classification for lung cancer. *J Thorac Oncol* 2007; 7:603–612.
9. Postmus PE, Chansky K, Crowley JJ, et al. The IASLC Lung Cancer Staging Project: Proposals for revision of the M descriptors in the forthcoming (7th) edition of the TNM classification for lung Cancer. *J Thorac Oncol* 2007;8:686–693.
10. Goldstraw P, Crowley J, Chansky K, et al. The IASLC lung cancer staging project: proposals for the revision of the TNM stage groupings in the forthcoming (7^{th}) edition of the TNM classification for Lung Cancer. *J Thorac Oncol* 2007;8:706–714.
11. Lopez-Encuentra A, Duque-Medina JL, Rami-Porta R, et al. Staging in lung cancer: is 3 cm a prognostic threshold in pathologic stage I non-small cell lung cancer? A multicenter study of 1,020 patients. *Chest* 2002;121:1515–1520.
12. Bronchogenic Carcinoma Cooperative Group of the Spanish Society of Pneumology and Thoracic Surgery (GCCB-S). Clinical tumour size and prognosis in lung cancer. *Eur Respir J* 1999;14:812–816.
13. Carbone E, Asamura H, Takei H, et al. T2 tumors larger than five centimeters in diameter can be upgraded to T3 in non-small cell lung cancer. *J Thorac Cardiovasc Surg* 2001;122:907–912.
14. Cangir AK, Kutlay H, Akai M, et al. Prognostic value of tumor size in non-small cell lung cancer larger than five centimeters in diameter. *Lung Cancer* 2004;46:325–331.
15. Vansteenkiste JF, De Belie B, Deneffe GJ, et al. Practical approach to patients presenting with multiple synchronous suspect lung lesions: a reflection on the current TNM classification based on 54 cases with complete follow-up. *Lung Cancer* 2001;34:169–175.
16. Horinouchi H, Kobayashi K. Surgical indications for lung cancer: influence of the M factor. *J Jpn Surg Soc* 2001;102:517–520.
17. Naruke T, Tshuchiya R, Kondo H, et al. Prognosis and survival after resection for bronchogenic carcinoma based on the 1997 TNM-staging classification: the Japanese experience. *Ann Thorac Surg* 2001;71:1759–1764.
18. Berghmans T, Lafitte JJ, Thiriaux J, et al. Survival is better predicted with a new classification of stage III unresectable non-small cell lung carcinoma treated by chemotherapy and radiotherapy. *Lung Cancer* 2004;45:339–348.
19. Gospodarowicz MK, Miller D, Groome PA, et al. The process for continuous improvement of the TNM classification. *Cancer* 2004;100: 1–5.

APPENDIX

IASLC International Staging Committee

P. Goldstraw (Chairperson), Royal Brompton Hospital, London, UK; H. Asamura, National Cancer Centre Hospital, Tokyo, Japan; D. Ball, Peter MacCallum Cancer Centre, Melbourne, Australia; E. Brambilla, Laboratoire de Pathologie Cellulaire, Grenoble Cedex, France; P.A. Bunn, University of Colorado Health Sciences, Denver, CO, USA; D. Carney, Mater Misericordiae Hospital, Dublin, Ireland; T. Le Chevalier, Institute Gustave Roussy, Villejuif, France; J. Crowley, Cancer Research and Biostatistics, Seattle, WA, USA; R. Ginsberg (deceased), Memorial Sloan-Kettering Cancer Center, New York, NY, USA; P. Groome, Queen's Cancer Research Institute, Kingston, Ontario, Canada; H. H. Hansen (retired), National University Hospital, Copenhagen, Denmark; P. Van Houtte, Institute Jules Bordet, Bruxelles, Belgium; J.-G. Im, Seoul National University Hospital, Seoul, South Korea; J. R. Jett, Mayo Clinic, Rochester, MN; H. Kato (retired), Tokyo Medical University, Tokyo, Japan; C. Kennedy, University of Sydney, Sydney, Australia; M. Krasnik, Gentofte Hospital, Copenhagen, Denmark; J. van Meerbeeck, University Hospital, Ghent, Belgium; T. Naruke (deceased), Saiseikai Central Hospital, Tokyo, Japan; E. F. Patz, Duke University Medical Center, Durham, NC, USA; P. E. Postmus, Free University Hospital, Amsterdam, The Netherlands; R. Rami-Porta, Hospital Mutua de Terrassa, Terrassa, Spain; V. Rusch, Memorial Sloan-Kettering Cancer Center, New York, NY; J. P. Sculier, Institute Jules Bordet, Brussels, Belgium; F. A. Shepherd, University of Toronto, Toronto, Ontario, Canada; Y. Shimosato (retired), National Cancer Centre, Tokyo, Japan; L. Sobin, Armed Forces Institute of Pathology, Washington, DC, USA; W. Travis, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; M. Tsuboi, Tokyo Medical University, Tokyo, Japan; R. Tsuchiya, National Cancer Center, Tokyo, Japan; E. Vallieres, Swedish Cancer Institute, Seattle, WA; J. Vansteenkiste, Leuven Lung Cancer Group, Leuven, Belgium; Yoh Watanabe (deceased), Kanazawa Medical University, Uchinada, Japan; and H. Yokomise, Kagawa University, Kagawa, Japan.

Cancer Research and Biostatistics

J. J. Crowley, K. Chansky, D. Giroux, and V. Bolejack, Seattle, WA.

Participating Institutions

O. Visser, Amsterdam Cancer Registry, Amsterdam, The Netherlands; R. Tsuchiya and T. Naruke (deceased), Japanese Joint Committee of Lung Cancer Registry, Japan; J. P. Van Meerbeeck, Flemish Lung Cancer Registry-VRGT, Brussels, Belgium; H. Bülzebruck, Thorax-Klinik am Universitatsklinikum, Heidelberg, Germany; R. Allison and L. Tripcony, Queensland Radium Institute, Herston, Australia; X. Wang, D. Watson, and J. Herndon, Cancer and Leukemia Group B (CALGB), USA; R. J. Stevens, Medical Research Council Clinical Trials Unit, London, UK; A. Depierre, E. Quoix, and Q. Tran, Intergroupe Francophone de Cancerologie Thoracique (IFCT), France; J. R. Jett and S. Mandrekar, North Central Cancer Treatment Group (NCCTG), USA; J. H. Schiller and R. J. Gray, Eastern Cooperative Oncology Group (ECOG), USA; J. L. Duque-Medina and A. Lopez-Encuentra, Bronchogenic Carcinoma Co-operative Group of the Spanish Society of Pneumology and Thoracic Surgery (GCCB-S), Spain; J. J. Crowley, Southwest Oncology Group (SWOG), USA; J. J. Crowley and K. M. W. Pisters, Bimodality Lung Oncology Team (BLOT), USA; T. E. Strand, Cancer Registry of Norway, Norway; S. Swann and H. Choy, Radiation Therapy Oncology Group (RTOG), USA; R. Damhuis, Rotterdam Cancer Registry, Rotterdam, The Netherlands; R. Komaki and P. K. Allen, M.D. Anderson Cancer Center–Radiation Therapy (MDACC-RT), Houston, TX; J. P. Sculier and M. Paesmans, European Lung Cancer Working Party (ELCWP), Europe; Y. L. Wu, Guangdong Provincial People's Hospital, P. R. China; M. Pesek and H. Krosnarova, Faculty Hospital Plzen, Czech Republic; T. Le Chevalier and A. Dunant, International Adjuvant Lung Cancer Trial (IALT), France; B. McCaughan and C. Kennedy, University of Sydney, Sydney, Australia; F. Shepherd and M. Whitehead, National Cancer Institute of Canada (NCIC), Canada; J. Jassem and W. Ryzman, Medical University of Gdansk, Gdansk, Poland; G. V. Scagliotti and P. Borasio, Università Degli Studi di Torino, S. Luigi Hospital, Orbassano, Italy; K. M. Fong and L. Passmore, Prince Charles Hospital, Brisbane, Australia; V. W. Rusch and B. J. Park, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; H. J. Baek, Korea Cancer Center Hospital, Seoul, South Korea; R. P. Perng, Taiwan Lung Cancer Society, Taiwan; R.C. Yung, A. Gramatikova, Johns Hopkins University, Baltimore, MD, USA; J. Vansteenkiste, Leuven Lung Cancer Group (LLCG), Leuven, Belgium; C. Brambilla and M. Colonna, Grenoble University Hospital–Isere Cancer Registry, France; J. Hunt and A. Park, Western Hospital, Melbourne, Australia; J. P. Sculier and T. Berghmans, Institute of Jules Bordet, Brussels, Belgium; A. K. Cangir, Ankara University School of Medicine, Ankara, Turkey; D. Subotic, Clinical Centre of Serbia, Belgrade, Serbia; R. Rosell and V. Aberola, Spanish Lung Cancer Group (SLCG), Spain; A. A. Vaporciyan and A. M. Correa, M.D. Anderson Cancer Center–Thoracic and Cardiovascular Surgery (MDACC-TCVS), Houston, TX; J. P. Pignon, T. Le Chevalier, and R. Komaki, Institut Gustave Roussy (IGR), Paris, France; T. Orlowski, Institute of Lung Diseases, Warsaw, Poland; D. Ball and J. Matthews, Peter MacCallum Cancer Institute, East Melbourne, Australia; M. Tsao, Princess Margaret Hospital, Toronto, Ontario, Canada; S. Darwish, Policlinic of Perugia, Perugia, Italy; H. I. Pass and T. Stevens, Karmanos Cancer Institute, Wayne State University, Detroit, MI, USA; G. Wright, St. Vincent's Hospital, Victoria, Australia; and C. Legrand and J. P. van Meerbeeck, European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium.

