



# Validation and comparison of several published prognostic systems for patients with small cell lung cancer

M. Paesmans\*, J.J. Lafitte<sup>#</sup>, J. Lecomte<sup>†</sup>, T. Berghmans<sup>+</sup>, A. Efremidis<sup>§</sup>, V. Giner<sup>f</sup>, O. Van Cutsem<sup>\*\*</sup>, A. Scherpereel<sup>###</sup>, A.P. Meert<sup>+</sup>, N. Leclercq<sup>+</sup>, P. Van Houtte<sup>††</sup> and J.P. Sculier<sup>+</sup>

**ABSTRACT:** The aim of the present study was to validate and compare published prognostic classifications for predicting the survival of patients with small cell lung cancer.

We pooled data from phase III randomised clinical trials, and used Cox models for validation purposes and concordance probability estimates for assessing predictive ability.

We included 693 patients. All the classifications impacted significantly on survival, with hazard ratios (HRs) in the range 1.57–1.68 (all  $p < 0.0001$ ). Median survival times were 16–19 months for the best predicted groups, while they were 6–7 months for the most poorly predicted groups. Most of the paired comparisons were statistically significant. We obtained similar results when restricting the analysis to patients with extensive disease. Multivariate Cox models for fitting survival data were also performed. The HRs for a single covariate were 8.23 (95% CI 5.88–11.69), and 9.46 (6.67–13.50), and for extensive disease were 5.60 (3.13–9.93), 12.49 (5.57–28.01) and 8.83 (4.66–16.64). Concordance probability estimates ranged 0.55–0.65 (overlapping confidence intervals).

Published classifications were validated and suitable for use at a population level. As expected, prediction at an individual level remains problematic. A specific model designed for extensive-disease patients did not appear to perform better.

**KEYWORDS:** Prognostic classifications, small cell lung cancer, survival, validation

Despite the sensitivity of small cell lung cancer to radiotherapy and chemotherapy, the prognosis of patients with this disease is poor, with the 5-yr survival rate being <10% [1]. The most reproducible prognostic factor is disease extent, which is also the main factor guiding therapy. Very few other prognostic factors have been clearly established: performance status, sex and some routine biological parameters, such as neutrophil or leukocyte counts, or albuminaemia [2, 3]. Our group, the European Lung Cancer Working Party (ELCWP), has long been interested in the identification of prognostic factors, and their integration in a classification system that could be used in care for providing information to patients and in clinical research for stratifying randomisation procedures or adjusting treatment comparisons. In 2000, we published such a classification, which we constructed using two different statistical strategies: recursive

partitioning and amalgamation (RECPAM) algorithms, which lead naturally to patient classification, and Cox regression modelling [4]. The RECPAM technique consists of starting with the whole population of patients and then dividing it into groups according to the prognostic covariate with the highest significance level on the outcome (overall survival in the present case). Once the population has been divided into two groups, the search for a new split is performed separately on the two obtained nodes. The partitioning process proceeds iteratively until no further variable is identified as a significant prognostic factor or until the number of patients becomes too small. A second iterative algorithm is then applied, as the terminal nodes obtained from different branches of the tree do not necessarily correspond to patients with significantly different survival distributions. After application of the amalgamation algorithm, the classification is immediate. From

## AFFILIATIONS

\*Data Centre,  
†Dept of Intensive Care Unit and Thoracic Oncology, and  
††Radiotherapy Dept, Institut Jules Bordet, Centre des Tumeurs de l'Université Libre de Bruxelles, Brussels,  
‡Dept of Pneumology, CHU de Charleroi, Charleroi, and  
\*\*Dept of Pneumology, Clinique Saint-Luc, Bouge, Belgium.  
#Dept of Pulmonary and Thoracic Oncology, CHRU de Lille, University of Lille II, and  
##INSERM U1019, Centre d'Infection et d'Immunité de Lille, Institut Pasteur de Lille, Lille, France.  
§Medicine Dept, Hellenic Cancer Institute – St Savas Oncology Hospital, Athens, Greece.  
fMedical Dept, Hospital de Sagunto, Valencia, Spain.

## CORRESPONDENCE

M. Paesmans  
Data Centre  
Institut Jules Bordet  
Rue Héger-Bordet 1  
B 1000 Brussels  
Belgium  
E-mail: marianne.paesmans@bordet.be

Received:  
July 14 2010  
Accepted after revision:  
Dec 15 2010  
First published online:  
May 26 2011

This article has supplementary material accessible from [www.erj.ersjournals.com](http://www.erj.ersjournals.com)

data on 763 patients registered in four clinical trials (one phase II and three phase III trials), we proposed a classification based on disease extent, Karnofsky performance index, age, sex and relative neutrophil count. Our system included four groups: one with the best prognosis, including only limited-disease patients; one with the worst prognosis, including only extensive-disease patients; and two intermediate groups composed of patients of any disease extent. To date, the RECPAM classification and final Cox model have not been further tested with the intent to validate them. Moreover, therapeutic standards have changed with the introduction of combined thoracic radiation and chemotherapy for patients with limited disease. We conducted the present study on a new series of patients with small cell lung

cancer with two objectives: 1) to validate our published classification and Cox model; and 2) to validate other previously published classifications/models and compare the different prognostic systems.

**MATERIAL AND METHODS**

We searched the literature for other published classifications based on series of patients with a sample size >500 subjects and including only purely prognostic factors (excluding therapeutic covariates). The limit of 500 patients was purely arbitrary but considered necessary to provide estimates of survival distributions with low variances. We identified four other studies with a published classification; all were based on

**TABLE 1** Patient characteristics

Characteristic	Trial					
	ELCWP 1922 [9]	ELCWP 1923 [8]	ELCWP 1994	All ELCWP trials	PAESMANS <i>et al.</i> [4]	SCULIER <i>et al.</i> [2]
<b>Subjects</b>	204	233	256	693	763	6609
<b>Sex</b>						
Male	163 (80)	196 (84)	211 (82)	570 (82)	689 (90)	4368 (66)
Female	41 (20)	37 (16)	45 (18)	123 (18)	74 (10)	2241 (34)
<b>Age yrs</b>	59 (32–75)	61 (34–74)	61 (40–83)	60 (32–84)	NR	NR
<b>Karnofsky performance status</b>						
60–70	14 (7)	57 (24)	64 (25)	135 (19)	250 (33)	1409 (21)
80–90	190 (93)	176 (76)	192 (75)	558 (81)	513 (67)	5200 (79)
100	88 (43)	41 (18)	31 (12)	160 (23)	NR	2039 (31)
<b>Disease extent</b>						
Limited	204 (100)	18 (8)	16 (6)	238 (34)	365 (48)	2870 (43)
Very limited	204	1	0	205	NR	NR
Extensive		215 (92)	240 (94)	455 (66)	398 (52)	3739 (57)
>2 metastatic sites		30	53	83	NR	NR
<b>Pleural effusion</b>						
Not documented	204 (100)	200 (86)	219 (86)	623 (90)	NR	NR
Yes		33 (14)	37 (14)	70 (10)		
<b>White blood cell count</b>						
<10 IU	152 (75)	155 (67)	153 (60)	460 (66)	516	NR
≥10 IU	52 (25)	78 (33)	100 (39)	230 (33)	241	
Not reported			3 (1)	3 (<1)	6	
<b>Relative neutrophil count</b>						
≤75%	166 (81)	154 (66)	152 (59)	472 (68)	528 (69)	NR
>75%	35 (17)	74 (32)	68 (27)	177 (26)	212 (28)	
NR	3 (1)	5 (2)	36 (14)	44 (6)	23 (3)	
<b>Alkaline phosphatases</b>						
Normal	101 (50)	69 (30)	107 (42)	277 (40)	297	NR
Abnormal	77 (38)	158 (68)	144 (56)	379 (55)	431	
NR	26 (13)	6 (3)	5 (2)	37 (5)	35	
<b>Creatinine</b>						
Normal	199 (98)	225 (97)	249 (97)	673 (97)	NR	NR
Abnormal	3 (1)	6 (3)	4 (2)	13 (2)		
NR	2 (1)	2 (1)	3 (1)	7 (1)		
<b>LDH</b>						
Normal	144 (71)	123 (53)	127 (50)	394 (57)	NR	NR
Abnormal	30 (15)	88 (38)	98 (38)	216 (31)		
NR	30 (15)	22 (9)	31 (12)	83 (12)		

Data are presented as n, n (%) or median (range). ELCWP: European Lung Cancer Working Party; NR: not reported; LDH: lactate dehydrogenase.

**TABLE 2** Distributions of the classifications

Classification	Group I	Group II	Group III	Group IV	Group V	Missing
<b>All patients<sup>#</sup></b>						
ALBAIN <i>et al.</i> [5]	145 (21)	60 (9)	227 (33)	178 (26)		83
SAGMAN <i>et al.</i> [6]	212 (31)	37 (5)	364 (53)	73 (11)		7
PAESMANS <i>et al.</i> [4]	113 (17)	352 (53)	103 (15)	98 (15)		27
SCULIER <i>et al.</i> [2]	195 (28)	53 (8)	279 (40)	166 (24)		
<b>Patients with extensive disease<sup>†</sup></b>						
ALBAIN <i>et al.</i> [5]			227 (56)	178 (44)		50
SAGMAN <i>et al.</i> [6]		23 (5)	356 (79)	73 (16)		3
PAESMANS <i>et al.</i> [4]		227 (53)	103 (24)	98 (23)		27
SCULIER <i>et al.</i> [2]		10 (2)	279 (61)	166 (37)		
FOSTER <i>et al.</i> [7]	232 (51)	34 (7)	131 (29)	4 (1)	52 (12)	2

Data are presented as n (%) or n. #: n=693; †: n=455.

RECPAM [2, 5–7]. When the publications also proposed a model based on regression analysis [2, 4, 7], we considered it in our comparisons too.

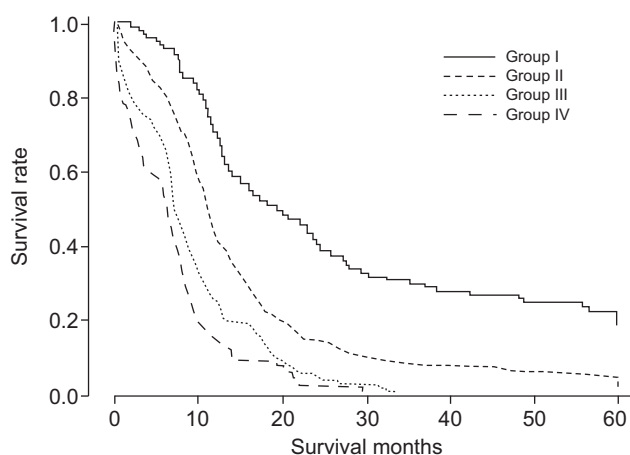
All the selected classifications and models were based on clinical, demographic or routine biological variables. All the RECPAM classifications lead to four prognostic groups, with the exception of the classification proposed by FOSTER *et al.* [7], which focuses on patients with extensive disease and has five different prognostic levels. The Cox regression models had four to five factors to explain the distribution of overall survival. The most frequently used covariates were disease extent, sex and age. Other covariates were lactate dehydrogenase level, white blood cell or neutrophil counts, creatinine level, alkaline phosphatases, or variables linked to disease stage. The full classifications and models are presented in the online supplementary material.

As validation series, we used a pooled database constituted from the databases of three further clinical trials that the ELCWP conducted on small cell lung cancer, from 1992 to 2008. Two of these studies are closed and the results have been published (ELCWP 1923 [8] and ELCWP 1922 [9]), while one is still ongoing (ELCWP 1994). The data on the patients registered in these studies were not included in the database used for derivation of our prognostic classification [4]. These are all phase III trials and their characteristics are described in the online supplementary material. Two trials included patients with extensive disease and treatment consisted of chemotherapy alone. The third trial was dedicated to patients with limited disease who were treated with combined chemoradiation and addressed the issue of the possible radiosensitising effect of cisplatin. For the ongoing trial, only patients randomised before October 31, 2007 were included in the database constructed for the present analysis, in order to achieve a theoretical follow-up duration of >2 yrs for all patients.

Some eligibility criteria were common to the three trials and similar to the eligibility criteria used for the trials considered in our previous study [4]. Small cell lung cancer had to be histologically proven and untreated; patients had to have normal haematological, hepatic and renal functions, have a Karnofsky performance status of  $\geq 60$ , have provided informed consent and be accessible for follow-up. The definition of limited disease was a disease confined to primary site, mediastinum and homolateral subclavicular lymph nodes without malignant pleural effusion in one trial [8]. In the other two studies, it was defined as a disease that could be treated in one radiotherapy field. Evaluation criteria in the trials included response to treatment after three and six courses of chemotherapy, and progression-free and overall survival.

### Statistical methodology

The objectives of the present study were, primarily, the validation of the results the ELCWP published in 2000 and, secondarily, the assessment and comparison of the prognostic values of the selected previously published RECPAM classifications and Cox models, using the new series of patients. The evaluation criterion was overall survival, measured from registration in the trials. All deaths were taken into account.



Group I	113	108	91	64	53	43	36	34	29	28	25	23	21
Group II	352	295	201	111	65	46	33	28	24	24	19	17	14
Group III	103	73	32	20	9	4	3	0					
Group IV	98	58	19	8	7	2	0						

**FIGURE 1.** Survival curves according to the European Lung Cancer Working Party prognostic classification.

All the data required for the assessment of the five published classifications were prospectively collected during the conduct of each of the trials.

We used nonparametric estimation of the survival distributions and comparison by log-rank tests as well as hazard ratio (HR) estimates based on Cox regression models. For assessing the prognostic value of the previously published Cox models, we constructed an overall score based on the published regression coefficients. The original scores were transformed to have the same range of theoretical values in order to allow comparison of HR estimates between the models.

As we had one model specific for patients with extensive disease, we tested the general classifications and models both on all the patients and on the subgroup of patients with extensive disease.

In the first two published trials that we used for constructing our validation series, we failed to identify any survival benefit between the arms [8, 9]. For the third, ongoing trial, we did not look at survival comparison between arms. We did not use any stratification by trial for our analyses, as one trial was dedicated to limited disease patients only and the others were limited exclusively to patients with extensive disease.

To assess the predictive ability of the prognostic covariates, we used the concordance probability estimate [10].

All the significance probabilities were two-tailed; a p-value <0.05 was considered statistically significant.

**RESULTS**

We collected data on a total of 693 patients, 204 from ELCWP 1922 [9], 233 from ELCWP 1923 [8] and 256 from the ongoing trial (ELCWP 1994). There was an intersection between institutions having recruited patients for the derivation series and the institutions contributing to the trials used for the present validation series. However, the overlap was not total and our validation project was intermediate between internal validation and external validation [11]. Patient characteristics are presented in table 1. Compared with our derivation series, there was an increase in the proportions of both females and patients with a Karnofsky index ≥80. The proportion of patients with limited disease was lower than in our previous study. There were some missing data for biological parameters preventing us from assessing the five classifications in all the patients. Depending on the classifications, rates of missing data ranged <1–12%.

Median length of follow-up was 119 months; death was observed in 646 (93%) patients. Theoretical follow-up was >2 yrs in 97% of the patients and >5 yrs in 85% of the patients.

**Distributions of classifications**

Table 2 presents the distribution of the classifications on the validation series, for all patients and for the subgroup of patients with extensive disease. Frequencies for each category of the classifications are presented after exclusion of patients with missing data, preventing us from allocating them to one of the categories. For each of the classifications, the lower the category is, the better the predicted survival time. Depending on the classification, the proportion of patients selected as having the best prognosis was 17–31% and the proportion of patients with the worst prognosis was 11–26%.

**Validation study**

Figure 1 shows survival curves according to the ELCWP prognostic classification. Group I had an estimated median survival time of 90 weeks, compared with 48, 34 and 28 weeks for groups II, III and IV, respectively. The overall comparison was highly significant (p<0.001), using the prognostic classification as a continuous covariate, as well as all the paired comparisons. Our previous Cox model identified four independent prognostic covariates. Table 3 provides the newly estimated regression coefficients for the validation series if all four covariates were entered into the model together with the same categorisation than for the derivation model. Three of them retained a statistically significant p-value (p<0.0001) while the last one, sex, was not significant, with a p-value of 0.08. The goodness-of-fit of a model including disease extent alone was improved if the RECPAM classification was used and further improved if the Cox model was used, confirming the usefulness of the integrated models.

**Prognostic models comparison**

In table 4, we present the estimates of the median survival times in each prognostic category, and an overall comparison reflecting the global prognostic value of each RECPAM classification using a nonparametric estimation method and the log-rank test. When all the patients were analysed, the overall comparison in each situation was highly significant. The HR estimates range 1.57–1.87 (change when comparing one category to the adjacent higher one). This means that the hazards for the group with the worst prognosis were four- to five-fold higher compared with the group with the best

**TABLE 3** Prognostic value of the four prognostic independent covariates from the study by PAESMANS *et al.* [4]

Characteristic	Derivation series <sup>#</sup>		Validation series <sup>†</sup>	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Karnofsky performance status ≥80	0.73 (0.60–0.86)	<0.001	0.56 (0.46–0.70)	<0.0001
Female sex	0.71 (0.54–0.94)	0.02	0.83 (0.66–1.02)	0.08
Limited disease	0.66 (0.57–0.78)	<0.001	0.45 (0.38–0.51)	<0.0001
Neutrophils ≤75%	0.80 (0.67–0.96)	0.05	0.68 (0.57–0.82)	<0.0001

HR: hazard ratio. <sup>#</sup>: n=763, 45 censored observations; <sup>†</sup>: n=693, 47 censored observations.

**TABLE 4** Estimates of median survival times in each category and overall comparison

Classification	Group I	Group II	Group III	Group IV	Group V	Overall	p-value
<b>All patients</b>							
ALBAIN <i>et al.</i> [5]	17.0 (14.4–20.8)	10.6 (7.5–12.5)	10.5 (9.4–11.5)	7.3 (6.4–11.5)		1.58 (1.45–1.72)	<0.0001
SAGMAN <i>et al.</i> [6]	15.7 (13.4–19.1)	13.3 (10.6–16.7)	9.5 (8.8–10.0)	6.4 (4.7–7.1)		1.60 (1.47–1.75)	<0.0001
PAESMANS <i>et al.</i> [4]	19.4 (13.9–23.9)	11.1 (10.6–11.8)	7.1 (6.6–8.7)	6.4 (4.0–7.3)		1.68 (1.53–1.84)	<0.0001
SCULIER <i>et al.</i> [2]	15.9 (13.5–19.9)	10.4 (7.6–13.0)	10.5 (9.6–11.2)	6.8 (6.1–7.8)		1.57 (1.45–1.71)	<0.0001
<b>Patients with extensive disease</b>							
ALBAIN <i>et al.</i> [5]			10.5 (9.4–11.5)	7.3 (6.4–8.3)		1.87 (1.51–2.31)	<0.001
SAGMAN <i>et al.</i> [6]		15.9 (11.0–21.3)	9.6 (9.0–10.3)	6.4 (4.7–7.1)		1.70 (1.35–2.13)	<0.001
PAESMANS <i>et al.</i> [4]		10.9 (10.0–11.6)	7.1 (6.6–8.7)	6.3 (4.0–7.3)		1.44 (1.27–1.64)	<0.001
SCULIER <i>et al.</i> [2]			10.5 (9.6–11.2)	6.8 (6.1–7.8)		1.72 (1.41–2.10)	<0.001
FOSTER <i>et al.</i> [7]	9.7 (8.8–10.7)	9.5 (7.0–11.1)	9.4 (8.6–10.4)		6.4 (3.3–7.8)	1.23 <sup>#</sup> (1.12–1.35)	<0.001

Data are presented as hazard ratios (95% CI), unless otherwise stated. For the calculation of the overall hazard ratio, only patients with all classifications available were considered (n=588 for all patients; n=383 for patients with extensive disease). The calculation of the hazard ratio was made using the classification as a continuous covariate. <sup>#</sup>: groups IV and V pooled.

prognosis. No heterogeneity was found between these HRs. As the overall comparisons were all significant, we performed paired comparisons (group *n* compared with group *n*+1), the results of which are presented in table 5. For the classification of ALBAIN *et al.* [5], there was no statistical evidence that groups II and III had different survival distributions. For the classification of FOSTER *et al.* [7], which focused on patients with extensive disease, group I could not be shown to be different from group II, and the comparison between groups II and III was not significant either. Group IV was not analysed because very few patients in our series belonged to that group.

The regression coefficients for the three Cox models were used to calculate an overall prognostic score. The scores were standardised in order to vary on a scale between 0 and 2 (on a patient population 18–85 yrs of age). The HRs associated with these scores are shown in table 6 and are all significantly different from 1. The more recent models [2, 7] appear to perform better than the third [4].

### Concordance probability estimates

Concordance probability estimates were calculated for each classification and model; predicted survival times were obtained from a Cox model using a single covariate that was a RECPAM classification or a covariate calculated from the regression coefficients of the published Cox models. They are presented in table 7 with 95% CIs. The Cox models had slightly higher coefficients than the RECPAM classifications. The coefficients were worse when the analysis was restricted to the population of patients with extensive disease, even for the models proposed by FOSTER *et al.* [7], which were specifically constructed on patients with extensive disease.

### DISCUSSION

Prognostic factor studies are numerous in the literature. However, reproducible and well-established independent prognostic factors are lacking [3]. Indeed, most often, identification of prognostic factors is performed retrospectively and there are few, if any, for patients with small cell lung

**TABLE 5** Paired comparisons

Classification	Group I versus II	Group II versus III	Group III versus IV	Group IV versus V
<b>All patients</b>				
ALBAIN <i>et al.</i> [5]	1.93 (1.39–2.64)	1.21 (0.90–1.66)	1.88 (1.53–2.31)	
SAGMAN <i>et al.</i> [6]	1.47 (1.00–2.11)	1.73 (1.23–2.51)	1.67 (1.29–2.15)	
PAESMANS <i>et al.</i> [4]	2.24 (1.77–2.87)	1.74 (1.39–2.18)	1.33 (1.00–1.77)	
SCULIER <i>et al.</i> [2]	2.17 (1.57–2.96)	1.13 (0.84–1.53)	1.77 (1.45–2.15)	
<b>Patients with extensive disease</b>				
ALBAIN <i>et al.</i> [5]			1.88 (1.53–2.31)	
SAGMAN <i>et al.</i> [6]			1.70 (1.30–2.18)	
PAESMANS <i>et al.</i> [4]		1.52 (1.20–1.92)	1.33 (1.00–1.77)	
SCULIER <i>et al.</i> [2]			1.77 (1.45–2.15)	
FOSTER <i>et al.</i> [7]	1.18 (0.79–1.69)	1.04 (0.71–1.57)		1.59 (1.15–2.17)

Data are presented as hazard ratios (95% CI).



**TABLE 6** Predictive value of the previously published Cox models

Classification	HR (95% CI)	p-value
<b>All patients</b>		
PAESMANS <i>et al.</i> [4]	8.23 (5.88–11.69)	<0.001
SCULIER <i>et al.</i> [2]	9.46 (6.67–13.50)	<0.001
<b>Patients with extensive disease</b>		
PAESMANS <i>et al.</i> [4]	5.60 (3.13–9.93)	<0.001
SCULIER <i>et al.</i> [2]	12.49 (5.57–28.01)	<0.001
FOSTER <i>et al.</i> [7]	8.83 (4.66–16.74)	<0.001

Hazard ratios (HRs) are for a continuous covariate calculated as a linear combination of the independent prognostic covariates weighted by their regression coefficients.

cancer. Phase III prognostic factor studies, as described by SIMON and ALTMAN [11], are prospectively conducted with *a priori* hypotheses and evaluation of the required sample size according to those hypotheses. Therefore, it is of crucial importance to validate results before making use of them. Furthermore, the independent prognostic value of a factor needs to be validated. In that context, we do not need to validate a single factor but rather a set of prognostic factors and, therefore, a classification or a model. There are several possibilities for validation exercises: internal validation with cross-validation or bootstrapping techniques, or external validation on new series of patients, which is the more convincing way to proceed whenever possible [12]. Several factors may be considered for external validation too: using patients recruited at the same institutions and during the same time period is less general than carrying out a validation study on patients of different origins and developing the disease later than the patients included in the derivation sets. We performed an external validation of our previous study [4] with an assessment of its historical transportability, as defined by JUSTICE *et al.* [13]; however, some institutions had contributed to both series of patients. It should be noted that the classification constructed by the International Association for the Study of Lung Cancer (IASLC) staging project is not entirely independent, as the ELCWP contributed to the constitution of the worldwide database and it is possible that data from the other series were also integrated into that database. This later classification was already validated as, in the IASLC study, two-thirds of the patients were used for model derivation and one-third was used for validation. Beyond the fact that the validation is not fully external, we should stress that our study was retrospective, leading to some missing data in the assessment of the classifications/models and to some heterogeneity in the way the required covariates were assessed. Furthermore, our patient population was restricted to a population of patients registered in clinical trials, limiting the generalisability of the conclusions.

Our validation was successful for both our RECPAM classification and our Cox model in a population of patients included in a clinical trial, despite the fact that treatment strategies have evolved with the introduction of combined chemoradiation modalities for limited disease. The prognosis of patients with

**TABLE 7** Concordance probability estimates (CPEs)

	CPE (95% CI)
<b>RECPAM classifications</b>	
All patients	
ALBAIN <i>et al.</i> [5]	0.63 (0.61–0.65)
SAGMAN <i>et al.</i> [6]	0.62 (0.61–0.65)
PAESMANS <i>et al.</i> [4]	0.62 (0.60–0.64)
SCULIER <i>et al.</i> [2]	0.63 (0.61–0.65)
Patients with extensive disease	
ALBAIN <i>et al.</i> [5]	0.58 (0.55–0.60)
SAGMAN <i>et al.</i> [6]	0.55 (0.53–0.57)
PAESMANS <i>et al.</i> [4]	0.58 (0.55–0.60)
SCULIER <i>et al.</i> [2]	0.57 (0.54–0.59)
FOSTER <i>et al.</i> [7]	0.56 (0.54–0.59)
<b>Cox models</b>	
All patients	
PAESMANS <i>et al.</i> [4]	0.64 (0.62–0.66)
SCULIER <i>et al.</i> [2]	0.65 (0.63–0.67)
Patients with extensive disease	
PAESMANS <i>et al.</i> [4]	0.58 (0.55–0.60)
SCULIER <i>et al.</i> [2]	0.59 (0.56–0.62)
FOSTER <i>et al.</i> [7]	0.60 (0.57–0.63)

The CPE represents the proportion of patients in whom the predicted and observed survival times are ordered the same way. RECPAM: recursive partitioning and amalgamation.

limited disease has improved and we may hypothesise that this is the joint result of a higher accuracy of staging techniques and a more effective treatment. Furthermore, we validated the four groups of our classification, as the ordered paired comparisons were all significant. Group I, which included only patients with limited disease, might be the target for developing new treatment modalities. A second conclusion for the validation of our models is that the concordance between predicted and true survival times in patients belonging to different risk groups is insufficient for making use of the models at the level of the individual patients. The prognostic covariate built on the Cox model performed slightly better than the RECPAM classification, which might be a disappointing result, as the RECPAM classification did not seem to benefit from the interaction effects that were naturally integrated during the building of the classification. Finally, the performance of our models decreased when applied to patients with extensive disease.

Most of these conclusions remain true when we are looking at the other proposed prognostic classifications published before the present study or developed recently, with a global validation and the most ordered paired comparisons being significant. Among the exceptions is the comparison between groups I and II in the classification of SAGMAN *et al.* [6]. This may be due to a small proportion of patients in group I and to a lack of power in that setting. Also, recently, the concept of very limited disease may have come to correspond to a highly selected patients population. The other exception is between groups II and III in the IASLC classification [2]; the proportion

of patients belonging to group II was also rather small and there might again be an issue of power in the confirmation of the separation between groups II and III. The most relevant difference between groups for the classifications of SCULIER *et al.* [2], PAESMANS *et al.* [4] and ALBAIN *et al.* [5] appeared to be the identification of the group with the best predicted survival distribution (group I), *i.e.* limited-disease patients with other favourable features. Depending on the classification, this group was more or less strictly defined, with logical consequences on the size of the group and its median survival duration.

One of the prognostic classifications [7] we compared was specifically developed for patients with extensive disease with the expectation of being more specific and, therefore, more accurate. Looking at our results, this is not the case, suggesting that other covariates than the one tested should be of importance or that the impact of the other covariates is associated with higher variability. Our sample size might also be insufficient, as this classification was based on five groups in a restricted validation series. The predictive ability, assessed by the concordance probability estimates of all the models, is decreased in the restricted populations of patients with extensive disease.

All the models studied here were based on easy-to-assess variables (age, sex and routine laboratory parameters), or on variables that would be required to determine tumour stage and therapeutic strategy anyway. Therefore, it should be recommended to use at least one classification for stratifying patients in a clinical trial or allow comparisons of patient populations. Indeed, all are of value, as there is no relevant difference between the models. To our knowledge, this is the first work validating several classification systems. All of them have been shown to be successful in a further series, showing once again that it is less relevant to identify isolated prognostic factors than to integrate them into a prognostic system and that several prognostic systems, although based on different covariates, may have comparable discriminant and predictive values. Disease extent, age, sex and performance status are, however, the cornerstones of the classifications.

None of these classifications takes into account molecular biological factors or gene signatures, and there is indeed room for improvement, as concordance probability estimates are clearly unsatisfactory. However, as they are very simple to assess, any study of these new possible prognostic factors should compare the predictive ability of more costly prognostic tools to more simple tests.

## STATEMENT OF INTEREST

None declared.

## REFERENCES

- 1 Shepherd FA, Crowley J, Van Houtte P, *et al.* The International Association for the Study of Lung Cancer lung cancer staging project: proposals regarding the clinical staging of small cell lung cancer in the forthcoming (seventh) edition of the tumor, node, metastasis classification for lung cancer. *J Thorac Oncol* 2007; 2: 1067–1077.
- 2 Sculier JP, Chansky K, Crowley JJ, *et al.* The impact of additional prognostic factors on survival and their relationship with the anatomical extent of disease expressed by the 6th Edition of the TNM Classification of Malignant Tumors and the proposals for the 7th Edition. *J Thorac Oncol* 2008; 3: 457–466.
- 3 Paesmans M. Prognostic factors in lung cancer. *Rev Mal Respir* 2005; 22: 8S76–8S80.
- 4 Paesmans M, Sculier JP, Lecomte J, *et al.* Prognostic factors for patients with small cell lung carcinoma: analysis of a series of 763 patients included in 4 consecutive prospective trials with a minimum follow-up of 5 years. *Cancer* 2000; 89: 523–533.
- 5 Albain KS, Crowley JJ, Leblanc M, *et al.* Determinants of improved outcome in small-cell lung cancer: an analysis of the 2,580-patient Southwest Oncology Group data base. *J Clin Oncol* 1990; 8: 1563–1574.
- 6 Sagman U, Maki E, Evans WK, *et al.* Small-cell carcinoma of the lung: derivation of a prognostic staging system. *J Clin Oncol* 1991; 9: 1639–1649.
- 7 Foster NR, Mandrekar SJ, Schild SE, *et al.* Prognostic factors differ by tumor stage for small cell lung cancer: a pooled analysis of North Central Cancer Treatment Group trials. *Cancer* 2009; 115: 2721–2731.
- 8 Sculier JP, Paesmans M, Lecomte J, *et al.* A three-arm phase III randomised trial assessing, in patients with extensive-disease small-cell lung cancer, accelerated chemotherapy with support of haematological growth factor or oral antibiotics. *Br J Cancer* 2001; 85: 1444–1451.
- 9 Sculier JP, Lafitte JJ, Efremidis A, *et al.* A phase III randomised study of concomitant induction radiochemotherapy testing two modalities of radiosensitisation by cisplatin (standard *versus* daily) for limited small-cell lung cancer. *Ann Oncol* 2008; 19: 1691–1697.
- 10 Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15: 361–387.
- 11 Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994; 69: 979–985.
- 12 Altman DG, Vergouwe Y, Royston P, *et al.* Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; 338: 1432–1435.
- 13 Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999; 130: 515–524.