Grading the neuroendocrine tumors of the lung: an evidence-based proposal

G Rindi, C Klersy[1], F Inzani, G Fellegara[2], L Ampollini[3], A Ardizzoni[4], N Campanini[5], P Carbognani[3], T M De Pas[6], D Galetta[7], P L Granone[8], L Righi[9], M Rusca[3], L Spaggiari[10], M Tiseo[4], G Viale[11], M Volante[9], M Papotti[9] and G Pelosi[11,12,†]

Institute of Anatomic Pathology, Università Cattolica del Sacro Cuore – Policlinico A. Gemelli, Rome, Italy
[1]Service of Biometry and Clinical Epidemiology, Research Department, IRCCS Fondazione Policlinico San Matteo, Pavia, Italy
[2]Service of Pathology, Centro Diagnostico Italiano, Milan, Italy
[3]Thoracic Unit, Department of Surgery, University of Parma, Parma, Italy
[4]Medical Oncology Unit, University Hospital of Parma, Parma, Italy
[5]Unit of Pathological Anatomy, Centre for Molecular and Translational Oncology, University Hospital, University of Parma, Parma, Italy
[6]Medical Oncology Unit of Respiratory Tract and Sarcomas, Department of Medical Oncology, European Institute of Oncology, Milan, Italy
[7]Division of Thoracic Surgery, European Institute of Oncology, Milan, Italy
[8]Department of Thoracic Surgery, Università Cattolica del Sacro Cuore – Policlinico A. Gemelli, Rome, Italy
[9]Division of Pathology, University of Turin at San Luigi Hospital, Orbassano, Torino, Italy
[10]Division of Thoracic Surgery, European Institute of Oncology, University of Milan School of Medicine, Milan, Italy
[11]Division of Pathology and Laboratory Medicine, European Institute of Oncology, Milan, Italy
[12]Department of Biomedical and Clinical Sciences 'Luigi Sacco', Università degli Studi, Milan, Italy
[†]G Pelosi is now at Department of Pathology and Laboratory Medicine, Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy

Correspondence should be addressed to G Rindi
Email
guido.rindi@rm.unicatt.it

Endocrine-Related Cancer

Abstract

Lung neuroendocrine tumors are catalogued in four categories by the World Health Organization (WHO 2004) classification. Its reproducibility and prognostic efficacy was disputed. The WHO 2010 classification of digestive neuroendocrine neoplasms is based on Ki67 proliferation assessment and proved prognostically effective. This study aims at comparing these two classifications and at defining a prognostic grading system for lung neuroendocrine tumors. The study included 399 patients who underwent surgery and with at least 1 year follow-up between 1989 and 2011. Data on 21 variables were collected, and performance of grading systems and their components was compared by Cox regression and multivariable analyses. All statistical tests were two-sided. At Cox analysis, WHO 2004 stratified patients into three major groups with statistically significant survival difference (typical carcinoid vs atypical carcinoid (AC), $P=0.021$; AC vs large-cell/small-cell lung neuroendocrine carcinomas, $P<0.001$). Optimal discrimination in three groups was observed by Ki67% (Ki67% cutoffs: G1 <4, G2 4–<25, G3 ≥ 25; G1 vs G2, $P=0.021$; and G2 vs G3, $P\leq 0.001$), mitotic count (G1 ≤ 2, G2 >2–47, G3 >47; G1 vs G2, $P\leq 0.001$; and G2 vs G3, $P\leq 0.001$), and presence of necrosis (G1 absent, G2 $<10\%$ of sample, G3 $>10\%$ of sample; G1 vs G2, $P\leq 0.001$; and G2 vs G3, $P\leq 0.001$) at uni and

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

multivariable analyses. The combination of these three variables resulted in a simple and effective grading system. A three-tiers grading system based on Ki67 index, mitotic count, and necrosis with cutoffs specifically generated for lung neuroendocrine tumors is prognostically effective and accurate.

Endocrine-Related Cancer
(2014) **21**, 1–16

Introduction

Lung neuroendocrine tumors include four different histologic subtypes, namely carcinoids (typical (TC) and atypical (AC)), large-cell neuroendocrine carcinoma (LCNEC), and small-cell lung carcinoma (SCLC), as defined by the World Health Organization (WHO) (Travis *et al*. 2004). This classification mainly relies on morphology, with the help of mitotic count and necrosis assessment. Its prognostic value was validated by several studies (Asamura *et al*. 2006, Righi *et al*. 2010). Difficulty in its reproducibility was, however, denoted, with high inter-observer variability (e.g. cytological features may be shared by AC and LCNEC and by LCNEC and SCLC making it difficult for diagnosis) and partial clinical efficacy (Travis *et al*. 1998, Marchevsky *et al*. 2001, Iyoda *et al*. 2007, den Bakker *et al*. 2010, Righi *et al*. 2010, Ha *et al*. 2012). The clinical significance of distinguishing between high-grade LCNEC and SCLC is not clear due to the overlapping survival curves usually observed (den Bakker *et al*. 2010, Righi *et al*. 2010, Ha *et al*. 2012).

An effective grading system for digestive neuroendocrine neoplasms was recently introduced by the European Neuroendocrine Tumor Society (ENETS) and endorsed by the WHO and the American Joint Cancer Committee (Rindi *et al*. 2007, Bosman *et al*. 2010, Edge *et al*. 2010). This three-tiers system largely relies on the assessment of the proliferation marker, Ki67, and proved accurate and predictive in large tumor series (Scarpa *et al*. 2010, Rindi *et al*. 2012).

Here, we applied the digestive neuroendocrine neoplasms grading principles to lung neuroendocrine tumors, aiming that they may improve the current WHO 2004 classification. Accordingly, we competitively tested a large multicenter cohort of lung neuroendocrine tumors with the WHO 2004 classification of lung neuroendocrine cancer (from now on WHO 2004) vs the current WHO classification of digestive neuroendocrine neoplasms (from now on ENETS/WHO 2010 grading) and also with other relevant grading variables.

Subjects and methods

Type of study

This was an observational retrospective longitudinal study, with an embedded cross-sectional diagnostic validation study.

Data collection

The data collection included 399 lung neuroendocrine tumors from the following centers: Milano (MI, $n=159$ neoplasms), Turin (TO, $n=90$ neoplasms), Parma (PR, $n=73$ neoplasms), and Rome (RM, $n=77$ neoplasms). The study was approved by the Ethic Committee of the Università Cattolica – Policlinico A. Gemelli, as coordinating Center. The study enrollment criteria included the patient who underwent surgery for cancer and the patient who was under observation for at least 1 year during 1989–2011. Data on cases were consecutively collected in each center, with no significant change in the surgical approach and standard therapy approach during the observation time. We analyzed 21 variables: sex, age at diagnosis in years, diagnosis according to WHO 2004 (Travis *et al*. 2004), status at follow-up, duration of follow-up in months, site, size, T (tumor), N (node), number of lymph nodes, M (metastases), the Union for International Cancer Control/the American Joint Committee on Cancer (UICC/AJCC) 2010 TNM stage (Travis *et al*. 2008, Edge *et al*. 2010), site of metastasis, mitosis/2 mm^2, Ki67 index defined according to the ENETS/WHO 2010 grading definition (Rindi *et al*. 2010), Ki67 index defined according to the Aperio automated computer-assisted quantitative method, Ki67 index defined according to the computer-assisted manual count method (see below), necrosis, chemotherapy, radiotherapy, somatostatin analog (SSA) therapy, and Octreoscan somatostatin receptor (SSR) imaging. For the variable necrosis, as the WHO 2004 lacks a quantitative definition for focal or diffuse, we arbitrarily and tentatively defined as focal (also meant

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

as punctate or spotty) when <10% of the sample or diffuse when equal or more than 10% of the sample. All deaths were recorded and were further classified as related to the underlying cancer by each hospital center, based on each patient's medical chart. Tumor-related death was defined as death directly or indirectly (e.g. therapy-related mortality) associated with the lung cancer. All data were cross-checked for inconsistencies by C Klersy and G Rindi. Every effort was made to minimize missing or incomplete data.

Ki67 assessment

Particular attention was paid to homogenous pathology assessment according to WHO 2004 and ENETS/WHO 2010 grading. All lung neuroendocrine tumors were reevaluated for consistency with the WHO 2004 classification (G Rindi, F Inzani, and G Fellegara) and discrepant cases, if present, discussed to reach consensus. The same investigators (F Inzani and G Rindi) performed the ENETS/WHO 2010 grading in the whole cohort. With the aim of obtaining the most homogenous data, Ki67 immunohistochemistry was centrally performed in the Immunohistochemistry Laboratory of the University of Parma by N Campanini with the MIB1 antibody. The Ki67 preparations of the entire cohort were assessed by the same investigators (G Rindi and F Inzani). With the aim of defining the most effective method for Ki67 determination in routine practice, a limited group of cases ($n=73$) were tested using the following three methods: i) the expert eye count method; ii) the Aperio automated computer-assisted quantitative method, and iii) the computer-assisted manual count method. The expert eye count method consisted of counting under a microscope 500–2000 cells at 40× in areas of highest nuclear labeling as required by ENETS/WHO 2010 (Rindi *et al.* 2010). The Aperio automated computer-assisted quantitative method consisted of scanning slides in the same areas at 40× using ScanScope XT (Aperio, Vista, CA, USA) and analyzing the digitized images with the Aperio Imagescope Software according to the manufacturer's indication. The computer-assisted manual count method consisted of taking a digital picture of the same areas and by manually counting on screen the Ki67 positive and negative nuclei using ImageJ Software (NIH, Bethesda, MA, USA; http://rsbweb.nih.gov/ij/) (Schneider *et al.* 2012). The expert eye and the Aperio-automated computer-assisted methods are commonly used in routine practice in pathology. The computer-assisted manual count method was considered as the reference (gold standard) method for quantitative Ki67% determination. This method adds the computer analysis power to the pathologists' cell discrimination efficacy.

Statistical analysis

A statistical analysis plan was defined and followed. Continuous data were described as the mean and S.D. or median and 25th–75th percentiles, and were compared by Kruskal–Wallis tests. Categorical data were described as counts and percentages and were compared by Fisher exact tests. For the Ki67 validation study, concordance between methods was evaluated with the Bland and Altman graphical method and the Lin's concordance correlation coefficient for Ki67 on a continuous scale and with the Kappa statistic after categorization according to tertiles of the distribution. The association of Ki67 eye and mitoses count was evaluated with the Spearman R and 95% CI. The optimal cutoffs for Ki67, maximizing both sensitivity and specificity to separate adjacent ENETS/WHO 2010 grading categories, were identified with the receiver operating curve (ROC) analysis.

The median follow-up was calculated by the inverse Kaplan–Meier method. Follow-up time was determined from the date of diagnosis to the date of death or the last follow-up for survivors. Overall survival and tumor-related death-free survival were estimated with the Kaplan–Meier method. For this latter analysis, patients dying from causes other than cancer were censored at their date of death. Death rates per 100 person-years and 95% CIs were reported. The Cox model was used to assess the prognostic value of a series of patient and tumor characteristics. Hazard ratios (HRs) and 95% CIs were also calculated. The proportional hazard assumption (Schoenfeld residuals) was always satisfied and model fit was assessed graphically with Cox–Snell residuals.

The performance of the grading systems was informally compared through Royston explained variation and the Harrell C discrimination statistics, in which the higher value was representative of better system performance. For this purpose, the model was fitted on a training sample and validated in a testing sample, after a random 2:1 split of the case series (Harrell *et al.* 1996, Newson 2010). Only cancers with data for all grading systems were used for comparative tests. All models were adjusted for stage. Given the collinearity of WHO 2004 and the ENETS/WHO 2010 grading, different multivariable models were fitted including either one of the two grading systems or the variables mitotic count, Ki67, and necrosis, while also controlling for other noncollinear predictors

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

(age, sex, site, and stage). Model validation and informal comparisons were done as referred earlier.

All analyses were performed with Stata 12 (Stata Corporation, College Station, TX, USA) and Medcalc 12, for the ROC curve analysis, (MedCalc Software, Mariakerke, Belgium). A two-sided P value <0.05 was considered statistically significant.

Results

Clinical pathological findings

Three centers (TO, PR, and RM) contributed about 60% of this cohort ($n=240$, 60.1%) with comparable numbers (PR=73, TO=90, and RM=77). The higher contribution by the Milan center ($n=159$, 39.8%) was likely due to its oncology referral practice. Patients were more often male ($n=245$, 61%) with a median age of 63.2 years (Table 1; range 11–86 years). The four WHO 2004 classes were balanced, with similar frequencies for TC and SCLC ($n=113$ (28%) and $n=108$ (27%) respectively) (Table 1). No discrepant diagnosis was observed between centers. The distribution of sex per WHO 2004 classes was substantially similar for TC and AC (TC female 68, male 45; AC 44 and 40), with a net male predominance for high-grade cancers (LCNEC female 17, male 76; SCLC 24 and 83). Neoplasms were more often located in the right lung ($n=217$, 60%; Table 1), and most in the upper lobe ($n=100$, not shown), similar to the left lung (upper lobe, $n=70$, not shown). The median tumor size was 2.8 cm (Table 1; mean 3.2 ± 2 cm, range 0.5–13 cm). This was reflected by prevalent T1 and T2 status (175 and 150 cases respectively; Table 1). More than half of cases were void of lymph node deposit (N0=215), with N1 status observed in about twice the cases than N2 (103 vs 55 respectively; Table 1), and only two cases with N3 status. Distant metastases (M1) were observed in 17 cases only, for 14 with available exact site definition (four in the liver, three in bones, two in contralateral lung, and one in the brain, pituitary, diaphragm, ovary, and thyroid). About 75% of cases were low stage (Stage I, $n=183$, 50% and Stage II, $n=90$, 24%) and 25% only were high stage (Stage III, $n=76$, 20.8% and Stage IV, $n=17$, 4.6%; Table 1) with similar distribution between WHO 2004 classes (not shown). Necrosis, either focal or diffuse, was observed in about 60% of tumors ($n=71$ and $n=166$ respectively; Table 1). When analyzed by the ENETS/WHO 2010 grading (Bosman *et al.* 2010, Rindi *et al.* 2010), about half tumors were G3 ($n=188$, 49.5%; Table 1), about one-third G2 ($n=123$, 32.4%; Table 1), and the remaining few were G1 ($n=69$, 18.2%; Table 1). Very limited information was available for standard chemotherapy, performed in 62 patients (three TC, ten AC, 18 LCNEC, and 31 SCLC), eight of which combined with radiotherapy and four with SSAs. Four further cases underwent also SSA therapy. Finally, information on SSR imaging with Octreoscan was available for 14 cases (three TC, five AC, two LCNEC, and one SCLC), of which three resulted negative (one TC, one AC, and one LCNEC).

Ki67 method validation

To validate the Ki67 method to be used for the entire cohort, we tested a fraction of present cases ($n=73$) for agreement analysis of data obtained by the expert eye count, the Aperio automated and the computer-assisted manual count methods, the latter considered as the gold standard. The expert eye count method showed a concordance correlation coefficient vs the computer-assisted manual method ($\rho\, c=0.98$, 95% CI=0.97–0.99; Fig. 1A) better, but similar to the value observed with the Aperio automated method ($\rho\, c=0.95$, 95% CI=0.94–0.97; Fig. 1B) and comparable to the value observed when comparing the expert eye count vs the Aperio automated method ($\rho\, c=0.94$, 95% CI=0.92–0.96; Fig. 1C). Both the expert eye and the Aperio methods were thus comparably effective as the gold standard. The expert eye count method, as the most practical in our hands, was adopted for the entire cohort.

Ki67-based grading system validation

The Ki67% distribution among WHO 2004 categories (TC, AC, LCNEC, and SCLC) was investigated. The Kruskal–Wallis equality of population rank test showed a statistically significant difference between the four WHO 2004 classes ($P<0.001$), with significant differences (after Bonferroni's correction) for pairwise comparisons between classes (TC vs AC, AC vs LCNC, and LCNC vs SCLC). The variable Ki67% was thus capable of distinguishing the four WHO 2004 classes in a statistically significant manner.

The specific Ki67% cutoff values discriminating the WHO 2004 classes were identified by ROC analysis. Between TC and AC classes, the Ki67% cutoff of >3% was recognized as discriminant (area under the ROC curve 0.712, moderate discrimination capacity; 95% CI=0.64–0.78), with 60% sensitivity (95% CI=48.4–70.8) and 78.70% specificity (95% CI=69.8–86.0). Between AC and LCNEC, the cutoff of >20% demonstrated optimal discrimination power (area under ROC curve 0.99, optimal discrimination capacity; 95% CI=0.98–1.00), with

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

Table 1 Clinical–pathological features

Variable	Patients	
	Valid cases	All
Age (years), median (25th–75th percentiles)	398	63.26 (53.17–70.79)
Sex (%)	399	
M		245 (61.40)
F		154 (38.60)
WHO 2004 class (%)	399	
TC		113 (28.32)
AC		84 (21.05)
LCNEC[a]		94 (23.56)
SCLC[b]		108 (27.05)
Site (%)	361	
RL		217 (60)
LL		144 (40)
Size (cm), median (25th–75th percentiles)	369	2.80 (1.70–4.00)
T (AJCC 2010) (%)	379	
T1		175 (46.1)
T2		150 (39.5)
T3		41 (10.8)
T4		13 (3.4)
N (AJCC 2010) (%)	372	
N0		212 (56.9)
N1		103 (27.6)
N2		55 (14.7)
N3		2 (0.5)
M (AJCC 2010) (%)	377	
M0		360 (94.5)
M1		17 (4.5)
Stage (AJCC 2010) (%)	366	
IA		116 (31.7)
IB		67 (18.3)
IIA		72 (19.7)
IIB		18 (4.9)
IIIA		69 (18.8)
IIIB		7 (1.9)
IV		17 (4.6)
Four stage (AJCC 2010) (%)	366	
I		183 (50)
II		90 (24.6)
III		76 (20.8)
IV		17 (4.6)
Necrosis (%)[c]	397	
Absent		160 (40.3)
Focal		71 (17.9)
Diffuse		166 (41.8)
Mitosis (three quantiles) (%)	370	
2		142 (38.4)
>2–47		107 (28.9)
48–125		121 (32.7)
Ki67 (three quantiles) (%)	380	
<5		128 (33.7)
5–<65		128 (33.7)
≥65–95		124 (32.6)
ENETS/WHO 2010 grading (%)	380	
G1		69 (18.2)
G2		123 (32.4)
G3		188 (49.5)

WHO, World Health Organization; TC, typical carcinoid; AC, atypical carcinoid; LCNEC, large-cell neuroendocrine carcinoma; SCLC, small-cell lung carcinoma; RL, right lung; LL, left lung; T, tumor definition according to the American Joint Cancer Committee 2010 (Edge *et al*. 2010); N, lymph nodes metastasis definition according to AJCC 2010; M, distant metastasis definition according to American Joint Cancer Committee 2010; Stage, stage definition according to AJCC 2010; ENETS, European Neuroendocrine Tumor Society; Ki67, Ki67 index defined according to ENETS/WHO 2010 (Bosman *et al*. 2010); Mitosis, mitotic index defined according to WHO 2004 (Travis *et al*. 2004).
[a]One combined LCNEC/adenocarcinoma.
[b]One combined SCLC/adenocarcinoma and one combined SCLC/LCNEC.
[c]Necrosis: focal, presence of <10% of sample, diffuse, more than 10% of sample.

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

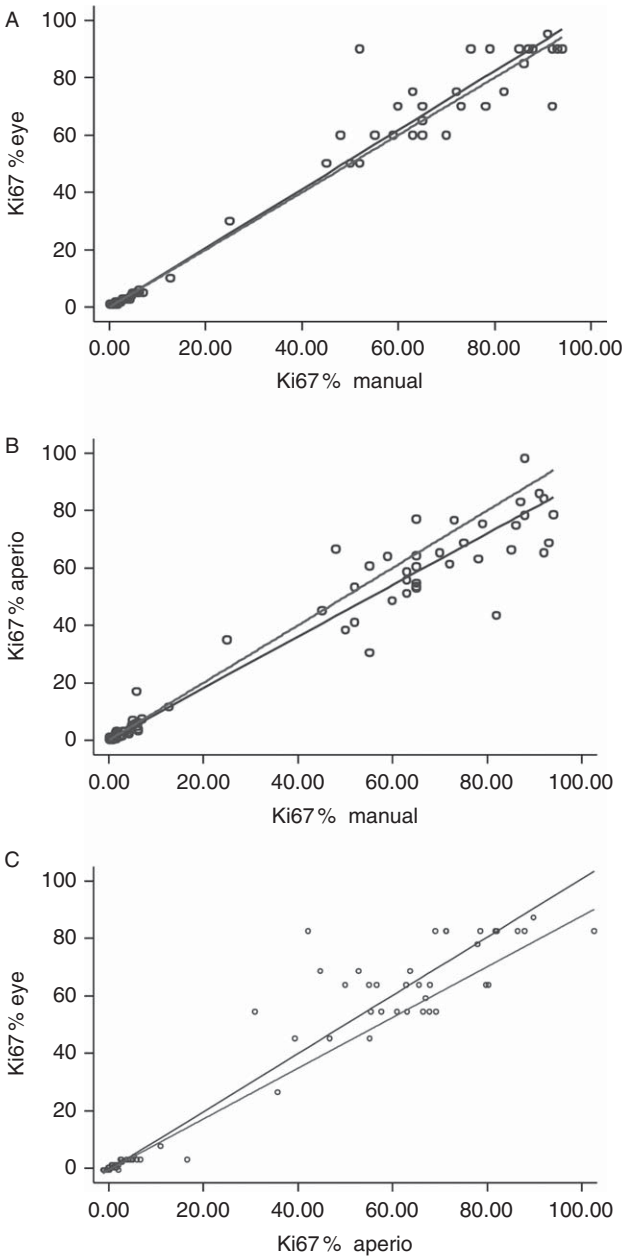

Figure 1
Graphical representation of the agreement data analysis for Ki67% by different methods on neuroendocrine neoplasm samples ($n=73$). The green line indicates the perfect agreement and the blue line indicates the mean distribution of values (single dots). (A) Agreement data between the expert eye count method vs the computer-assisted manual count method. (B) Agreement data between by the Aperio automated count method vs the computer-assisted manual count method. (C) Agreement data between the expert eye count method vs the Aperio automated count method.

96.67% sensitivity (95% CI=90.60–99.38) and 98.75% specificity (95% CI=93.20–100.00). Finally, the cutoff of >60% was discriminant between LCNEC and SCLC (area under the ROC curve 0.61, weak discrimination capacity; 95% CI=0.50–0.70), with 74% sensitivity (95% CI=64.30–82.30) and 45.56% specificity (95% CI=35.00–56.40). Overall, the Ki67% cutoffs discriminated with moderate to optimal discrimination power between the first three WHO 2004 classes, and with weak power between LCNEC and SCLC.

To assess whether the current Ki67-based ENETS/WHO 2010 grading for digestive neuroendocrine neoplasms (Bosman *et al*. 2010) could effectively discriminate lung neuroendocrine tumors, a Kruskal–Wallis equality-of-population rank test was performed. A statistically significant difference between lung neuroendocrine tumors was observed ($P<0.001$), with significant differences of all pairwise comparisons (after Bonferroni's correction). The association of ENETS/WHO 2010 grading and the WHO 2004 classes was demonstrated by Fisher exact test <0.001 (with significant *post-hoc* pairwise comparisons after Bonferroni's correction). This finding was strongly supported by an optimal Spearman's correlation value (0.858 on 380 observations, 95% CI=0.83–0.88). On the same line, a statistically significant distribution of mitoses was observed between the three ENETS/WHO 2010 classes (Kruskal–Wallis equality of population rank test, $P<0.001$, with significant *post-hoc* pairwise comparisons after Bonferroni's correction). The search for mitotic count cutoffs to discriminate different ENETS/WHO 2010 grades in lung neuroendocrine tumors was defined by ROC analysis. A cutoff of >0.4 was discriminant between G1 and G2 (area under the ROC curve 0.74, moderate discrimination capacity; 95% CI=0.67–0.80), with 76.86% sensitivity (95% CI=68.30–84.00) and 61.76% specificity (95% CI=49.20–73.30). A cutoff of >10 proved discriminant between G2 and G3 (area under the ROC curve 0.98, optimal discrimination capacity; 95% CI=0.95–0.99), with 98.83% sensitivity (95% CI=95.80–99.90) and 95.04% specificity (95% CI=98.50–98.20). Overall any grade increase of the ENETS/WHO 2010 grading corresponded to a change in WHO 2004 class and to a parallel change in mitotic count, with specific cutoffs for any grade change. These findings support the feasibility of a Ki67-based grading for lung neuroendocrine tumor classification.

Survival and grading systems: univariable analysis

Follow-up information for overall survival was available in 384 cases. The observed median time of follow-up was 70.72 months (25th–75th percentiles=38.49–104.60 months). A total of 148 patients died, which corresponded to a death rate of 9.24 deaths per 100 persons per year

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

(95% CI=7.87–10.86). The cumulative survival at 5 and 10 years was 62% (95% CI=56–67) and 51% (95% CI=44–58) respectively (Fig. 2A).

With the exception of stage, all grading variables tested were statistically significant predictors of survival. All predictors performed remarkably well and yielded comparable values, with relatively lower explained variation for necrosis (Royston explained variation 0.43, 95% CI=0.25–0.65) and relatively higher value for mitosis in four quantiles (Royston explained variation 0.499, 95% CI=0.307–0.738) (Table 2). The discrimination ability was substantially similar, spanning from the relatively lower values of Ki67 in four ROC values or KI67 in three quantiles (Harrel C statistics 0.71, 95% CI=0.63–0.78 for both), to the relatively higher value of mitosis in four quantiles (Harrel C statistics 0.78, 95% CI=0.71–0.85) (Table 2).

Both the WHO 2004 and the ENETS/WHO 2010 grading failed to provide clear discrimination between categories (Table 2 and Fig. 2B and C). Similarly, when variables were tested in four tiers, all failed to display discrimination. In detail, mitosis in four quantiles failed to separate the last two classes ($P=0.54$; Table 2 and Fig. 3A), while Ki67, either in four quantiles or in four cutoffs identified by the ROC curve, failed to separate the first two classes ($P=0.15$ and $P=0.26$ respectively; Table 2 and Fig. 3B and C), the latter one was also not discriminating between the last two ($P=0.35$; Table 2 and Fig. 3C). On the contrary, when tested in three tiers, either in tertiles (not shown for Ki67) or in three cutoffs identified by the ROC curve (Ki67 only), all variables displayed clear discrimination between categories (Table 2 and Fig. 4A, B and C).

As for the Stage, grouping stages (Stages IA–IB grouped in Stage I; Stages IIA–IIB grouped in Stage II; and Stages IIIA–IIIB grouped in Stage III) yielded a low explained variation (Royston explained variation 0.08, 95% CI=0.01–0.30), with relatively fair performance (Harrel C statistics 0.65, 95% CI=0.57–0.73) (Table 2). Statistically significant differences between stages were only observed for grouped stages, but the extremely wide 95% CI of Stage 4 graphically obscured the statistically significant differences (Table 2 and Supplementary Figure 1, see section on supplementary data given at the end of this article).

A sensitivity analysis on the role of grading systems and grading variables on tumor-related death gave results essentially comparable with those of Table 2, with similar, though slightly lower, Royston explained variation for survival prediction and discrimination ability by the Harrell C (Supplementary Table 1, see section on supplementary data given at the end of this article).

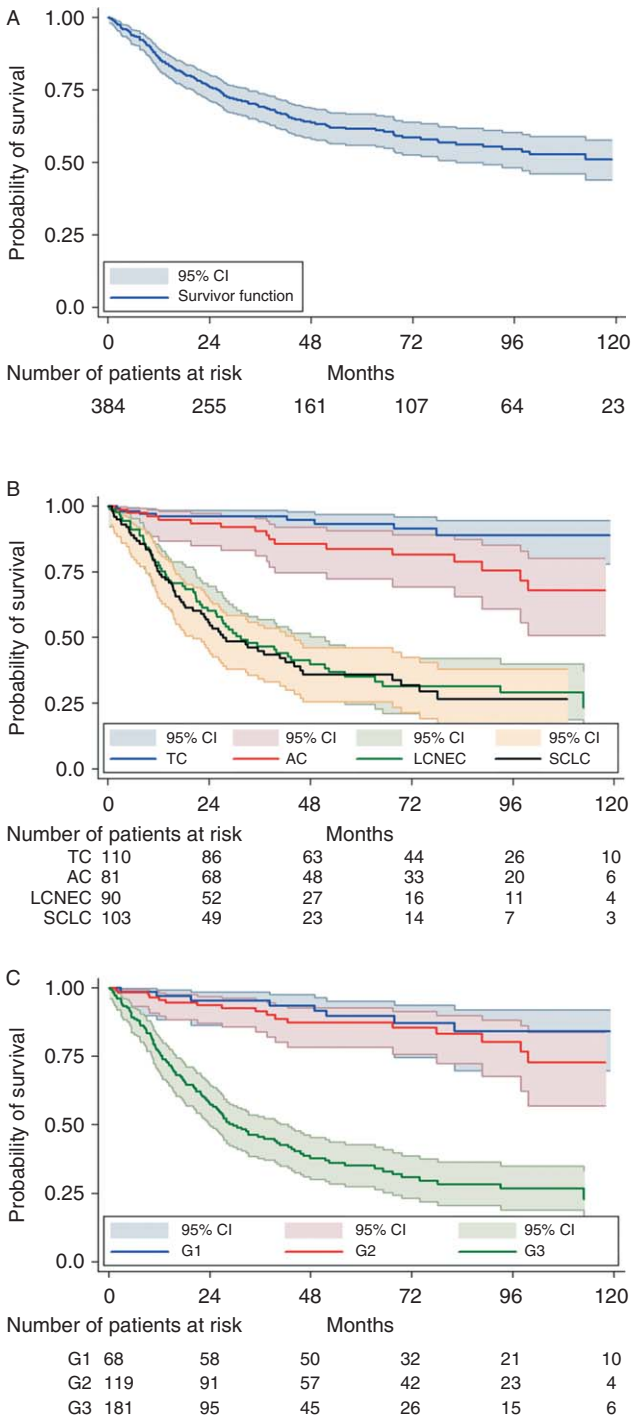

Figure 2
Kaplan–Meier survival curves for lung neuroendocrine tumors overall, by WHO 2004 classification and WHO 2010 grading. (A) The Kaplan–Meier survival curve is shown with the number of patients at risk given below the graph ($n=384$). (B) The neoplasms were grouped by WHO 2004 definition and Kaplan–Meier survival was calculated: 95% CIs are shown. The number of patients at risk is given below the graph ($n=384$). (C) The neoplasms were grouped by WHO 2010 grading for digestive tumors and Kaplan–Meier survival was calculated: 95% CIs are shown. The number of patients at risk is given below the graph ($n=368$).

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

Table 2 Predictors of overall survival at univariable analysis

Variable	Valid cases	All	Deaths, *n* (%)	Death rate per 100 person-year (95% CI)	HR (95% CI)	*P*[a]	*P*[b]	Royston explained variation (95% CI)	Harrell C (95% CI)[c,d]
WHO 2004 class	384					**<0.001**		0.498 (0.298–0.687)	0.75 (0.68–0.82)
TC		110	9 (8.18)	1.554 (0.808–2.987)	1.0 (referent)				
AC		81	18 (22.22)	4.125 (2.599–6.548)	2.68 (1.108–19.89)	0.021	0.021		
LCNEC		90	59 (65.356)	19.337 (14.982–24.958)	11.37 (4.197–101.73)	<0.001	<0.001		
SCLC		103	62 (60.19)	22.077 (17.212–28.317)	12.82 (7.489–72.274)	<0.001	0.580		
Mitosis (four quantiles)	355					**<0.001**		0.499 (0.307–0.738)	0.78 (0.71–0.85)
1		103	8 (7.77)	1.435 (0.717–2.870)	1.0 (referent)				
>1–<10		79	19 (24.05)	4.487 (2.862–7.035)	3.14 (1.14–8.64)	0.026	0.026		
10–55		86	51 (59.30)	16.849 (12.805–22.170)	10.71 (5.26–21.81)	<0.001	<0.001		
>55		87	51 (58.60)	22.617 (17.188–29.759)	13.75 (6.77–27.90)	<0.001	0.540		
Ki67 (four quantiles)	368					**<0.001**		0.443 (0.268–0.688)	0.72 (0.65–0.79)
<3		94	10 (10.64)	1.744 (0.938–3.241)	1.0 (referent)				
3–<25		93	18 (19.35)	4.443 (2.799–7.052)	3.39 (0.72–7.93)	0.155	0.155		
25–<80		95	69 (72.63)	23.379 (18.465–29.601)	11.77 (5.15–26.89)	<0.001	<0.000		
≥80		86	45 (52.33)	18.557 (13.855–24.854)	9.29 (3.80–22.72)	<0.001	0.003		
Ki67 (ROC values, 4)	368					**<0.001**		0.452 (0.280–0.668)	0.71 (0.63–0.78)
<4		113	15 (13.27)	2.267 (1.367–3.761)	1.0 (referent)				
4–<20		66	9 (13.64)	3.060 (1.592–5.881)	1.26 (0.84–1.87)	0.262	0.262		
20–<65		70	50 (71.43)	21.970 (16.652–28.988)	8.48 (4.89–14.72)	<0.001	<0.000		
≥65		119	68 (57.14)	20.428 (16.106–25.909)	7.96 (5.05–12.54)	<0.001	<0.348		
Necrosis[e]	382					**<0.001**		0.427 (0.250–0.651)	0.72 (0.65–0.79)
Absent		154	16 (10.739)	1.934 (1.184–3.157)	1.0 (referent)				
Focal		70	30 (42.86)	10.370 (7.251–14.832)	5.22 (4.09–6.67)	<0.001	<0.001		
Diffuse		158	101 (63.92)	20.873 (17.175–25.368)	9.74 (7.81–12.15)	<0.001	0.001		
ENETS/WHO 2010	368					**<0.001**		0.433 (0.010–0.296)	0.72 (0.65–0.78)
G1		68	9 (13.24)	2.090 (1.087–4.017)	1.0 (referent)				
G2		119	19 (15.97)	3.468 (2.212–5.437)	1.58 (0.49–5.15)	0.444	0.444		
G3		181	114 (62.98)	21.204 (17.648–25.477)	8.89 (3.82–20.73)	<0.001	<0.001		
Mitosis (three quantiles)	355					**<0.001**		0.458 (0.228–0.700)	0.77 (0.70–0.84)
2		139	13 (9.35)	1.70 (0.990–2.938)	1.0 (referent)				
>2–47		102	52 (50.98)	12.120 (9.235–15.905)	6.74 (5.01–9.06)	<0.001	<0.001		
>47		114	64 (56.14)	20.130 (15.756–25.719)	10.56 (7.41–15.04)	<0.001	<0.001		
Ki67 (three quantiles)	368					**<0.001**		0.313 (0.150–0.498)	0.71 (0.63–0.78)
<5		124	16 (12.90)	2.24 (1.375–3.664)	1.0 (referent)				
5–<65		125	58 (46.40)	12.328 (9.531–15.947)	4.97 (3.13–7.88)	<0.001	<0.001		
≥65		119	68 (57.14)	20.428 (16.106–25.909)	8.13 (4.63–14.25)	<0.001	<0.001		
Ki67 (ROC values, 3)	368					**<0.001**		0.447 (0.197–0.627)	0.73 (0.67–0.79)
<4		113	15 (13.27)	2.267 (0.367–3.761)	1.0 (referent)				
4–<25		74	13 (17.57)	4.102 (2.381–7.064)	1.63 (0.76–3.49)	0.021	0.021		
≥25		181	114 (62.98)	21.204 (17.648–25.477)	8.21 (4.62–14.62)	<0.001	<0.001		
Stage (AJCC 2010)	352					**<0.001**		0.002 (0.003–0.220)	0.63 (0.55–0.71)
IA		110	23 (20.91)	4.480 (2.977–6.742)	1.0 (referent)				
IB		65	19 (29.23)	5.990 (3.821–9.391)	1.34 (0.88–2.03)	0.170	0.170		

Table 2 *Continued*

Variable	Valid cases	All	Deaths, *n* (%)	Death rate per 100 person-year (95% CI)	HR (95% CI)	*P*[a]	*P*[b]	Royston explained variation (95% CI)	Harrell C (95% CI)[c,d]
IIA		69	29 (42.03)	10.404 (7.230–14.971)	2.18 (1.22–3.90)	0.009	0.086		
IIB		17	9 (52.94)	10.109 (5.260–19.429)	2.41 (1.75–3.31)	<0.001	0.739		
IIIA		68	44 (64.71)	21.640 (5.260–19.429)	4.42 (2.75–7.11)	<0.001	0.088		
IIIB		7	4 (57.14)	20.074 (7.534–53.486)	4.16 (1.03–16.78)	0.045	0.907		
IV		16	7 (43.75)	13.687 (6.525–28.710)	2.71 (2.04–3.59)	<0.001	0.467		
Four stage (AJCC 2010)	352					**<0.001**[f]		0.081 (0.010–0.296)	0.65 (0.57–0.73)
I		175	42 (24.00)	5.057 (3.737–6.843)	1.0 (referent)				
II		86	38 (44.19)	10.332 (7.518–14.200)	1.97 (1.23–3.16)	0.005	0.005		
III		75	48 (64.00)	21.500 (16.202–28.530)	3.90 (2.47–6.17)	<0.001	0.001		
IV		16	7 (43.75)	13.687 (6.525–28.710)	2.40 (1.97–2.93)	<0.001	<0.001		

WHO, World Health Organization; TC, typical carcinoid; AC, atypical carcinoid; LCNEC, large-cell neuroendocrine carcinoma; SCLC, small-cell lung carcinoma; ENETS, European Neuroendocrine Tumor Society; Mitosis, number of mitoses per 2 mm^2 according to WHO 2004 (Travis *et al.* 2004); Ki67, Ki67 index defined according to ENETS/WHO 2010 (Bosman *et al.* 2010).
[a]Cox models were used to calculate two-sided *P* values vs referent.
[b]Cox models were used to calculate two-sided *P* values with the previous class as the reference.
[c]Comparison of model performance was done informally.
[d]All not significant.
[e]Necrosis: focal, presence of <10% of sample, diffuse, more than 10% of sample; AJCC 2010, American Joint Cancer Committee (Edge *et al.* 2010).
[f]Log-rank test.

Survival and grading systems: multivariable analysis

Given that both the WHO 2004 and ENETS/WHO 2010 grading and the grading variables such as necrosis, mitosis, and Ki67 displayed high collinearity, five non-nested multivariable models were fitted including the WHO 2004 and the ENETS/WHO 2010 grading classifications, the grading variables in three tiers, and noncollinear predictors (Models A, B, C and D, Table 3). All models performed remarkably well, with well-explained variations, high shrinkage coefficient, and optimal discrimination at Harrell C statistic. In all models, the grading variables resulted as independent predictors of survival; however, only mitosis in three quantiles (Model B), necrosis (Model C), and Ki67 in three cutoffs identified by the ROC curve (Model E) retained the statistically significant separation capacity observed at univariable analysis (Table 2). Of the other variables tested, older age and Stage in four groups (Stages III and IV only) consistently resulted as independent predictors of survival (Table 3). Three further models with four tiers distribution (mitosis in four quantiles, Ki67 in four quantiles, and Ki67 in four cutoffs identified by the ROC curve) were also fitted and showed comparable performances, but confirmed to be ineffective for class separation (Models F–H, Supplementary Table 2, see section on supplementary data given at the end of this article).

Generating and testing a new grading proposal

The three variables mentioned earlier (mitotic count, necrosis, and Ki67) with cutoff based on tertile distribution for mitosis and ROC curve values for Ki67 were used to generate a grading system (Table 4). To stratify patients accordingly, 348 cases with information of all variables and follow-up were stratified in tertiles for progressive death rate based on the association of the three parameters (Table 5). The grade definition was as follows: G1 if two of three markers were at level 1; G2 if two of three markers were at level 2; and G3 if two of three markers were at level 3. Cases were reclassified resulting in 134 G1, 59 G2, and 155 G3 and tested for survival analysis. All TC ($n=105$) were reclassified as G1; of AC ($n=75$), 29 were G1, 45 were G2, and one were G3; of LCNEC ($n=86$), eight were G2 and 78 G3; of SCLC ($n=82$), six were G2 and 76 were G3. At univariable analysis, the new grading performed remarkably well showing high HR (G1 vs G2 4.42, $P<0.001$, 95% CI 2.59–7.56; G1 vs G3 11.37, $P<0.001$, 95% CI 8.80–14.69; G2 vs G3 2.57, $P<0.001$, 95% CI 1.66–3.97), high Harrell's C statistics (0.76,

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

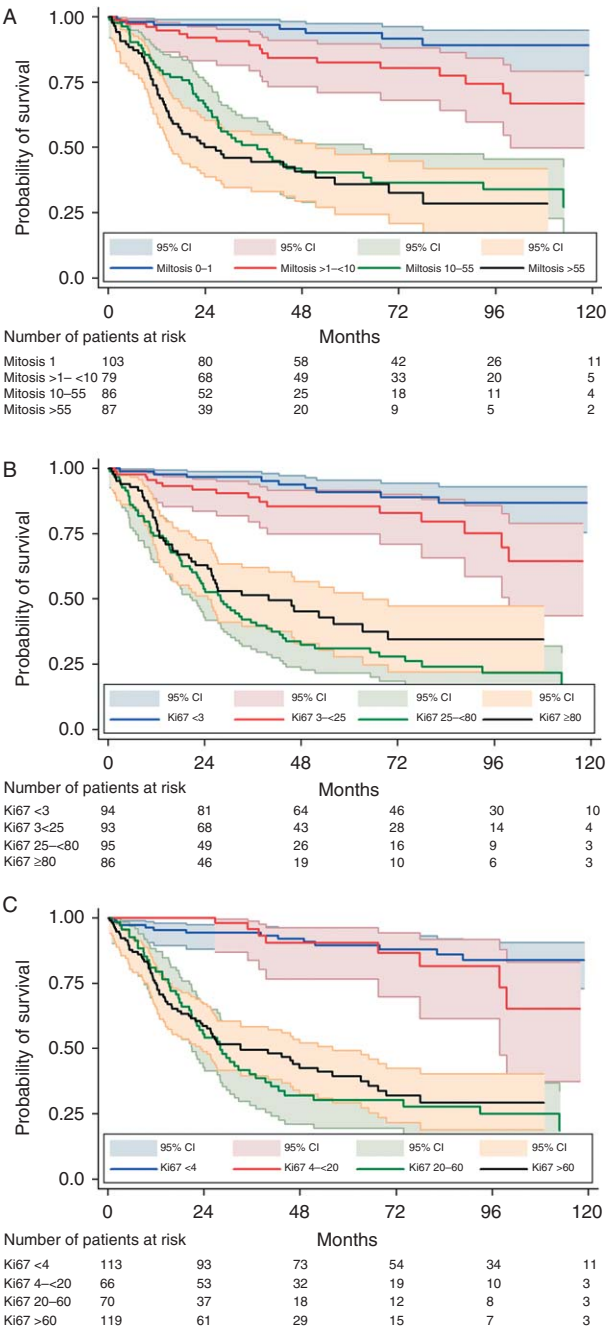

Figure 3
Kaplan–Meier survival curves for lung neuroendocrine tumors overall, by grading variables in four tiers. (A) The neoplasms were grouped by mitotic count in four quantiles distribution (see Table 2; $n=355$), (B) by Ki67% in four quantiles distribution (see Table 2; $n=368$), and (C) by Ki67% in four cutoffs identified by the ROC curve distribution (see Table 2; $n=368$): 95% CIs are shown. The number of patients at risk is given below the graphs.

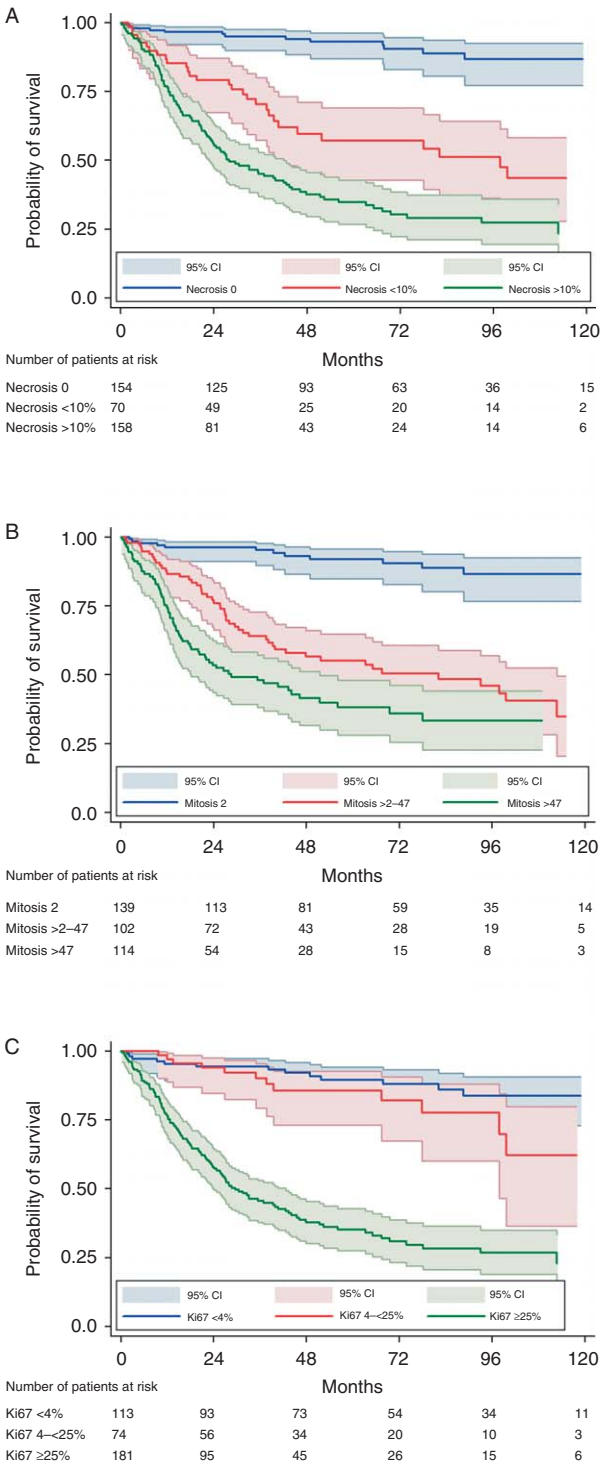

Figure 4
Kaplan–Meier survival curves for lung neuroendocrine tumors overall, by grading variables in three tiers. (A) The neoplasms were grouped by the presence of necrosis (see Table 2; $n=382$), (B) by mitotic count in three quantiles distribution (see Table 2; $n=355$), and (C) by Ki67% in three cutoffs identified by the ROC curve distribution (see Table 2; $n=368$): 95% CIs are shown. The number of patients at risk is given below the graphs.

$P<0.001$, 95% CI 0.69–0.82), and high Royston explained variation (0.49, 95% CI 0.30–0.70) (Fig. 5). Similar efficacy was demonstrated at multivariable modeling as performed earlier, with high HR (G1 vs G2 3.72, $P<0.001$, 95% CI

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

Table 3 Multivariable modeling for grading variables for overall survival of patients with lung neuroendocrine tumors

Model	Variable	HR (95% CI)	*P*[a]	*P*[b]	Model *P*	Shrinkage coefficient	Royston explained variation (95% CI)	Harrell's C (95% CI)[c]
A	WHO 2004 class		**<0.001**		**<0.001**	0.914	0.447 (0.152–0.635)	0.76 (0.68–0.84)
	TC	1.0 (referent)	1.0 (referent)					
	AC	2.65 (1.13–6.18)	0.024	0.024				
	LCNEC	7.46 (3.79–14.67)	<0.001	<0.001				
	SCLC	8.99 (3.24–24.99)	<0.001	0.505				
	Age	1.04 (1.03–1.06)	<0.001					
	Sex	0.98 (0.74–0.1.29)	0.870					
	Site	0.061 (0.56–1.01)	0.061					
			<0.001					
	Stage (AJCC 2010)							
	I	1.0 (referent)	1.0 (referent)					
	II	1.87 (0.89–3.94)	0.098					
	III	3.80 (2.46–5.85)	<0.001					
	IV	3.27 (1.31–8.20)	<0.011					
B	Mitosis (three quantiles)		**<0.001**		**<0.001**	0.986	0.362 (0.152–0.635)	0.75 (0.66–0.83)
	2	1.0 (referent)	1.0 (referent)					
	>2–47	4.08 (2.44–6.80)	<0.001	<0.001				
	>47	7.00 (3.32–14.75)	<0.001	<0.001				
	Age	1.04 (1.03–1.06)	<0.001					
	Sex	1.06 (0.72–1.55)	0.763					
	Site	0.74 (0.45–1.22)	0.244					
			<0.001					
	Stage (AJCC 2010)							
	I	1.0 (referent)	1.0 (referent)					
	II	1.92 (0.75–4.94)	0.173					
	III	4.01 (2.34–6.87)	<0.001					
	IV	2.88 (1.33–6.26)	0.008					
C	Necrosis[d]		**<0.001**		**<0.001**	0.942	0.435 (0.260–0.643)	0.76 (0.68–0.84)
	Absent	1.0 (referent)	1.0 (referent)					
	Focal	4.82 (3.12–7.44)	<0.001	<0.001				
	Diffuse	6.49 (3.84–10.98)	<0.001	0.009				
	Age	1.04 (1.03–1.06)	<0.001					
	Sex	1.13 (0.82–1.57)	0.449					
	Site	0.80 (0.59–1.08)	0.142					
			<0.001					
	Stage (AJCC 2010)							
	I	1.0 (referent)	1.0 (referent)					
	II	1.89 (0.83–4.27)	0.129					
	III	3.49 (2.33–5.24)	<0.001					
	IV	2.58 (1.16–5.74)	0.020					

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

Table 3 *Continued*

Model	Variable	HR (95% CI)	P^a	P^b	Model *P*	Shrinkage coefficient	Royston explained variation (95% CI)	Harrell's C (95% CI)[c]
D	ENETS/WHO 2010		**<0.001**		**<0.001**	0.948	0.331 (0.168–0.618)	0.74 (0.66–0.83)
	G1	1.0 (referent)	1.0 (referent)					
	G2	1.10 (0.55–2.19)	0.783	0.783				
	G3	5.98 (3.41–10.47)	<0.001	<0.001				
	Age	1.04 (1.02–1.05)	<0.001					
	Sex	1.00 (0.80–1.27)	0.968					
	Site	0.70 (0.49–1.00)	0.052					
			<0.001					
	Stage (AJCC 2010)							
	I	1.0 (referent)	1.0 (referent)					
	II	1.80 (0.90–3.60)	0.098					
	III	4.03 (2.89–5.62)	<0.001					
	IV	3.44 (1.79–6.63)	<0.001					
E	Ki67 (ROC values, 3)		**<0.001**		**<0.001**	0.909	0.388 (0.168–0.569)	0.75 (0.67–0.84)
	<4	1.0 (referent)	1.0 (referent)					
	4–<25	1.17 (0.91–1.50)	0.223	0.023				
	≥25	5.99 (4.28–8.38)	<0.001	<0.001				
	Age	1.04 (1.02–1.05)	<0.001					
	Sex	1.01 (0.78–1.30)	0.968					
	Site	0.70	0.049					
			<0.001					
	Stage (AJCC 2010)							
	I	1.0 (referent)	1.0 (referent)					
	II	1.80 (0.90–3.62)	0.098					
	III	4.02 (2.87–5.64)	<0.001					
	IV	3.40 (1.82–6.36)	<0.001					

WHO, World Health Organization; TC, typical carcinoid; AC, atypical carcinoid; LCNEC, large-cell neuroendocrine carcinoma; SCLC, small-cell lung carcinoma; ENETS, European Neuroendocrine Tumor Society; Mitosis, number of mitoses per 2 mm^2 according to WHO 2004 (Travis *et al*. 2004); Ki67, Ki67 index defined according to ENETS/WHO 2010 (Bosman *et al*. 2010).
[a]Cox models were used to calculate two-sided *P* values vs referent.
[b]Cox models were used to calculate two-sided *P* values with the previous class as the reference.
[c]Comparison of model performance was done informally.
[d]Necrosis: focal, presence of <10% of sample, diffuse, more than 10% of sample; AJCC 2010, American Joint Cancer Committee (Edge *et al*. 2010).

Table 4 Grading parameters with cutoff definitions based on present cohort findings (see note for application details)

	Variable		
Grade	Mitotic count (10HPF)[a]	Ki67 (%)[b]	Necrosis (%)[c]
G1	2	<4	No
G2	>2–47	4–<25	<10
G3	>47	≥25	>10

[a]10HPF, ten high-power field = 2 mm^2, to be assessed in at least 50 fields at 40× in areas of highest mitotic density.
[b]Ki67%: MIB antibody, as percentage of 500–2000 cells counted in areas of highest nuclear labeling.
[c]Necrosis as % of sample: focal, presence of <10% of sample, diffuse, more than 10% of sample. For grade definition by parameters association see Table 5 and text; for fractional values, approximate to the lower for ≤0.5 and to the higher for >0.5.

1.93–7.16; G2 vs G3 2.33, $P<0.001$, 95% CI 1.81–3.00), high Harrell's C statistics (0.75, $P<0.001$, 95% CI 0.66–0.83), and high Royston explained variation (0.37, 95% CI 0.15–0.64).

Discussion

Grading is an essential instrument to predict cancer behavior. Its strength directly stems from morphological parameters, cheap and prognostically effective tools so far unsurpassed in the real-life management of patients. Grading efficacy depends on the degree of its reproducibility, its limit residing in its exquisitely qualitative nature. Our aim was to provide a clinically efficient grading system that may improve the current WHO 2004 classification (Travis *et al.* 2004).

Here, we demonstrated that a three-tiers proliferation-based system is effective in predicting patient survival when tested in a large cohort of early-stage lung neuroendocrine cancers. Three evidences support this statement. First, WHO 2004 efficiently separated cancer patients as stratified in carcinoids (TC vs AC vs LCNEC/SCLC), but was ineffective in separating high-grade LCNEC vs SCLC (Fig. 1B). This observation largely confirmed previously reported data (Marchevsky *et al.* 2001, Asamura *et al.* 2006, den Bakker *et al.* 2010, Righi *et al.* 2010, Ha *et al.* 2012). Second, when using prognostically effective proliferation-associated tools like mitotic count and Ki67, a three-tiers system emerged as statistically efficient in predicting survival. These variables identified an intermediate-grade cancer-class which fares better than high-grade cancer patients and worse than low-grade (Fig. 3B and C). Third, the indirect morphological sign of malignancy, the presence of necrosis, similarly identified a three-tiers cancer patient stratification (Fig. 3A). Both mitotic count and presence of necrosis are well-known malignancy-associated parameters and were used to build the WHO 2004 classification (Travis *et al.* 2004). Interestingly, necrosis is recommended for reporting also in digestive neuroendocrine cancer implying its potential prognostic role (Klimstra 2013).

The grading proposal for lung neuroendocrine tumors stemming from the present data and including such three parameters proved simple in application and extremely effective in patient stratification (Tables 4 and 5 and Fig. 5). Since generated within the present cohort, this grading system necessarily requires confirmation of its

Table 5 The new grading with parameters association and death rates on present cohort (see text for application details)

	Ki67%			**Mitotic count** (10HPF)			**Necrosis**						
	0–<4	4–25	>25	2	>2–47	>47	No	<10%	>10%	**Death rate per 100 person-year**	**Variable combination**	**Deaths** (*n*)	**Death rate per 100 person-year** (95% CI)
Level	1	2	3	1	2	3	1	2	3				
G1	x			x			x			0.82–4.75	111	96	1.55 (0.77–3.99)
	x			x				x			112	9	1.53 (0.22–10.88)
	x				x		x				121	7	4.75 (1.19–18.99)
		x		x			x				211	29	0.82 (0.12–5.83)
		x			x		x				221	19	3.38 (1.09–10.49)
G2	x				x			x		6.08–15.52	122	7	11.22 (4.21–29.90)
		x		x			x				212	8	6.08 (1.52–24.31)
		x			x			x			222	11	8.15 (3.06–21.72)
		x				x			x		233	4	14.52 (3.63–58.04)
			x		x			x			322	13	11.47 (4.78–27.57)
G3		x			x				x	20.50–32.80	223	2	32.84 (4.63–233.10)
			x		x				x		323	45	20.50 (14.50–28.99)
			x			x		x			332	14	23.41 (11.7–46.81)
			x			x			x		333	99	20.96 (16.05–27.37)

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

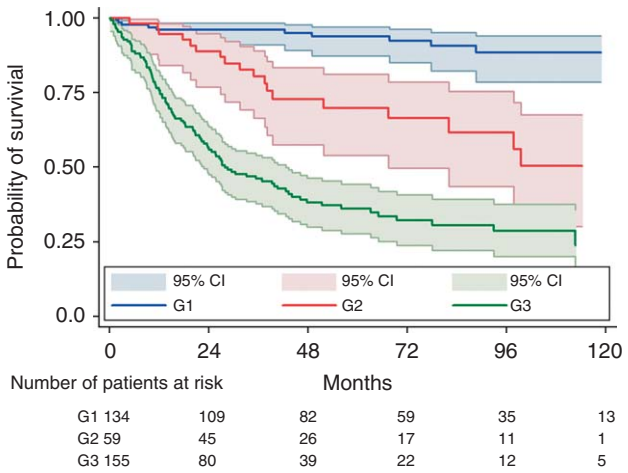

Figure 5
Kaplan–Meier survival curves for lung neuroendocrine tumors overall, by the new grading system inclusive of the three variables Ki67, mitosis, and necrosis as from Table 5 ($n=348$): 95% CIs are shown. The number of patients at risk is given below the graphs.

efficacy by studies on large, independent series. In specific, a larger sample will provide a more refined class separation. The current literature points to difficulties in accurately categorizing lung neuroendocrine cancers and supports the simplification of the WHO 2004 classification (Huang *et al*. 2002, Asamura *et al*. 2006, den Bakker *et al*. 2010, Ha *et al*. 2012). In addition, separating high-grade lung neuroendocrine cancers in two categories seems so far void of prognostic significance, given that LCNEC and SCLC patients experience similar poor outcome. This new grading, if confirmed in its efficacy, may well overcome such pitfalls.

So far Ki67 index is used for lung neuroendocrine cancer only for distinguishing TC/AC vs SCLC/LCNEC in small biopsy/cytology samples (Pelosi *et al*. 2005). Several previous studies applied Ki67 index to lung neuroendocrine tumors, some of them to test its efficacy vs the WHO 2004 classification (see Walts *et al*. (2012) and references therein Walts *et al*. (2012)). Contradictory results were obtained, likely reflecting the poor statistical power of investigated series and/or different Ki67 methods (Grimaldi *et al*. 2011, Walts *et al*. 2012, Zahel *et al*. 2012). Our data indicate that, similar to what experienced for digestive neuroendocrine neoplasms, the introduction of a reproducible and objective parameter (the Ki67 proliferation index) may well provide an effective grading tool. In addition, here we provided evidence that at least two methods commonly adopted in pathology practice can be effectively used for Ki67 determination in neuroendocrine cancer, as demonstrated by others (Yang *et al*. 2011, Tang *et al*. 2012, Walts *et al*. 2012). In particular, computer-assisted methods like the Aperio, now available in major centers, can be safely adopted thus reducing the burden of the quantitative assessment, the major criticism so far moved to Ki67-based grading systems. Of interest, similar efficacy was demonstrated for Ki67 eyeballing assessment (Tang *et al*. 2012). Finally, in light of previous evidence (Pelosi *et al*. 2005), the use of Ki67 as further grading variable will hopefully be of help in establishing the diagnosis of intermediate malignant neuroendocrine cancer (TC vs AC) in small biopsies too.

The ENETS/WHO 2010 grading for digestive neuroendocrine neoplasms (Bosman *et al*. 2010, Rindi *et al*. 2010) was ineffective in this large lung cohort (Fig. 2C). This suggests that tissue-specific unknown features are relevant and require site-specific cutoffs for Ki67 and mitosis indexes, as recently observed for pancreas (Scarpa *et al*. 2010, Rindi *et al*. 2012). Nonetheless, our data are proof of concept that a three-tiers proliferation-based grading system may work for prognostic stratification of neuroendocrine cancer patients also outside the digestive system.

This cohort was composed by early-stage, surgical-only neuroendocrine tumors and not designed to test the efficacy of the staging system (Edge *et al*. 2010). Nonetheless, the current staging accurately predicted patient survival for grouped Stages 1–3 (Supplementary Figure 1). The fact that no discrimination capacity was observed between sub-stage groups (Table 2 and Supplementary Table 1) suggests the need for a simpler system for lung neuroendocrine tumors, as observed in the digestive tract (Scarpa *et al*. 2010, Rindi *et al*. 2012).

This is the largest series ever used to test the efficacy of a Ki67-based grading system in lung neuroendocrine cancer. Though imperfect, this cohort is intended to be an acceptable compromise between patient selection bias and diagnostic accuracy. Indeed, our series was balanced in four equally populated lung neuroendocrine cancer classes according to WHO 2004, and provided effectively comparable surgical cases. The present series, however, suffers from the important exclusion of non-surgical, advanced, high-grade neuroendocrine cancer cases. Incidentally, our data indicate that 'early', high-grade neuroendocrine cancer patients may experience relatively long survival after curative oncological surgery.

The major limit of this study is its retrospective nature and consequent potentially non-homogenous data collection. Some variation in surgical approach at different Institution may have resulted in different efficacy, or

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-13-0246

Published by Bioscientifica Ltd.

different therapy policy in different centers may also have impacted the observed survival. A second important limit is the case selection based on curative surgery, making this cohort unsuitable for aims different from those investigated here. Finally, the present database did not contain information for recurrence – free survival, a relevant measure of low-grade neuroendocrine tumors. This aspect should be investigated in specifically designed future studies.

In conclusion, the present data indicate that Ki67 is an effective grading tool for lung neuroendocrine cancer. A three-tiers grading system based on Ki67 index, mitotic count, and presence of necrosis with cutoffs specifically generated for lung neuroendocrine tumors is prognostically effective and accurate, though in need of validation, supporting its introduction in the clinical practice.

Supplementary data
This is linked to the online version of the paper at http://dx.doi.org/10.1530/ERC-13-0246.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This work was in part supported by internal university grants to G Rindi (University of Parma and line D1/2010-2011, Università Cattolica).

Author contribution statement
G Rindi, C Klersy and F Inzani contributed equally to this work.

Acknowledgements
We are thankful to Prof. J Rosai (CDI, Milano, I) for the use of Aperio system by G Fellegara and to Mr L Tinelli (Nuova Impronta, Parma, I) for the professional art work.

References

Asamura H, Kameya T, Matsuno Y, Noguchi M, Tada H, Ishikawa Y, Yokose T, Jiang SX, Inoue T, Nakagawa K *et al.* 2006 Neuroendocrine neoplasms of the lung: a prognostic spectrum. *Journal of Clinical Oncology* **24** 70–76. (doi:10.1200/JCO.2005.04.1202)

den Bakker MA, Willemsen S, Grunberg K, Noorduijn LA, van Oosterhout MF, van Suylen RJ, Timens W, Vrugt B, Wiersma-van Tilburg A & Thunnissen FB 2010 Small cell carcinoma of the lung and large cell neuroendocrine carcinoma interobserver variability. *Histopathology* **56** 356–363. (doi:10.1111/j.1365-2559.2010.03486.x)

Bosman F, Carneiro F, Hruban RH & Theise ND 2010 In *WHO Classification of Tumours of the Digestive System*. Lyon, France: IARC Press.

Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL & Trotti A 2010 In *AJCC Cancer Staging Manual*. New York, NY, USA: Springer.

Grimaldi F, Muser D, Beltrami CA, Machin P, Morelli A, Pizzolitto S, Talmassons G, Marciello F, Colao AA, Monaco R *et al.* 2011 Partitioning of bronchopulmonary carcinoids in two different prognostic categories by ki-67 score. *Frontiers in Endocrinology* **2** 20. (doi:10.3389/fendo.2011.00020)

Ha SY, Han J, Kim WS, Suh BS & Roh MS 2012 Interobserver variability in diagnosing high-grade neuroendocrine carcinoma of the lung and comparing it with the morphometric analysis. *Korean Journal of Pathology* **46** 42–47. (doi:10.4132/KoreanJPathol.2012.46.1.42)

Harrell FE Jr, Lee KL & Mark DB 1996 Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* **15** 361–387. (doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4)

Huang Q, Muzitansky A & Mark EJ 2002 Pulmonary neuroendocrine carcinomas. A review of 234 cases and a statistical analysis of 50 cases treated at one institution using a simple clinicopathologic classification. *Archives of Pathology & Laboratory Medicine* **126** 545–553.

Iyoda A, Hiroshima K, Nakatani Y & Fujisawa T 2007 Pulmonary large cell neuroendocrine carcinoma: its place in the spectrum of pulmonary carcinoma. *Annals of Thoracic Surgery* **84** 702–707. (doi:10.1016/j.athoracsur.2007.03.093)

Klimstra DS 2013 Pathology reporting of neuroendocrine tumors: essential elements for accurate diagnosis, classification, and staging. *Seminars in Oncology* **40** 23–36. (doi:10.1053/j.seminoncol.2012.11.001)

Marchevsky AM, Gal AA, Shah S & Koss MN 2001 Morphometry confirms the presence of considerable nuclear size overlap between "small cells" and "large cells" in high-grade pulmonary neuroendocrine neoplasms. *American Journal of Clinical Pathology* **116** 466–472. (doi:10.1309/H40B-8W14-4Q47-03EP)

Newson RB 2010 Comparing the predictive powers of survival models using Harrell's C or Somers' D. *Stata Journal* **10** 339–358.

Pelosi G, Rodriguez J, Viale G & Rosai J 2005 Typical and atypical pulmonary carcinoid tumor overdiagnosed as small-cell carcinoma on biopsy specimens: a major pitfall in the management of lung cancer patients. *American Journal of Surgical Pathology* **29** 179–187. (doi:10.1097/01.pas.0000149690.75462.29)

Righi L, Volante M, Tavaglione V, Bille A, Daniele L, Angusti T, Inzani F, Pelosi G, Rindi G & Papotti M 2010 Somatostatin receptor tissue distribution in lung neuroendocrine tumours: a clinicopathologic and immunohistochemical study of 218 'clinically aggressive' cases. *Annals of Oncology* **21** 548–555. (doi:10.1093/annonc/mdp334)

Rindi G, Kloppel G, Couvelard A, Komminoth P, Korner M, Lopes JM, McNicol AM, Nilsson O, Perren A, Scarpa A *et al.* 2007 TNM staging of midgut and hindgut (neuro) endocrine tumors: a consensus proposal including a grading system. *Virchows Archiv* **451** 757–762. (doi:10.1007/s00428-007-0452-1)

Rindi G, Arnold R, Capella C, Klimstra DS, Klöppel G, Komminoth P & Solcia E 2010 Nomenclature and classification of digestive neuroendocrine tumours. In *World Health Organization Classification of Tumours, Pathology and Genetics of Tumours of the Digestive System*, pp 10–12. Eds F Bosman & F Carneiro. Lyon: IARC Press.

Rindi G, Falconi M, Klersy C, Albarello L, Boninsegna L, Buchler MW, Capella C, Caplin M, Couvelard A, Doglioni C *et al.* 2012 TNM staging of neoplasms of the endocrine pancreas: results from a large international cohort study. *Journal of the National Cancer Institute* **104** 764–777. (doi:10.1093/jnci/djs208)

Scarpa A, Mantovani W, Capelli P, Beghelli S, Boninsegna L, Bettini R, Panzuto F, Pederzoli P, delle Fave G & Falconi M 2010 Pancreatic endocrine tumors: improved TNM staging and histopathological grading permit a clinically efficient prognostic stratification of patients. *Modern Pathology* **23** 824–833. (doi:10.1038/modpathol.2010.58)

Published by Bioscientifica Ltd.

Schneider CA, Rasband WS & Eliceiri KW 2012 NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9** 671–675. (doi:10.1038/nmeth.2089)

Tang LH, Gonen M, Hedvat C, Modlin IM & Klimstra DS 2012 Objective quantification of the Ki67 proliferative index in neuroendocrine tumors of the gastroenteropancreatic system: a comparison of digital image analysis with manual methods. *American Journal of Surgical Pathology* **36** 1761–1770. (doi:10.1097/PAS.0b013e318263207c)

Travis WD, Gal AA, Colby TV, Klimstra DS, Falk R & Koss MN 1998 Reproducibility of neuroendocrine lung tumor classification. *Human Pathology* **29** 272–279. (doi:10.1016/S0046-8177(98)90047-8)

Travis WD, Brambilla E, Müller-Hermelink HK & Harris CC Eds 2004 In *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart*. Lyon, France: IARC Press.

Travis WD, Giroux DJ, Chansky K, Crowley J, Asamura H, Brambilla E, Jett J, Kennedy C, Rami-Porta R, Rusch VW *et al*. 2008 The IASLC Lung Cancer Staging Project: proposals for the inclusion of broncho-pulmonary carcinoid tumors in the forthcoming (seventh) edition of the TNM Classification for Lung Cancer. *Journal of Thoracic Oncology* **3** 1213–1223. (doi:10.1097/JTO.0b013e31818b06e3)

Walts AE, Ines D & Marchevsky AM 2012 Limited role of Ki-67 proliferative index in predicting overall short-term survival in patients with typical and atypical pulmonary carcinoid tumors. *Modern Pathology* **25** 1258–1264. (doi:10.1038/modpathol.2012.81)

Yang Z, Tang LH & Klimstra DS 2011 Effect of tumor heterogeneity on the assessment of Ki67 labeling index in well-differentiated neuroendocrine tumors metastatic to the liver: implications for prognostic stratification. *American Journal of Surgical Pathology* **35** 853–860. (doi:10.1097/PAS.0b013e31821a0696)

Zahel T, Krysa S, Herpel E, Stenzinger A, Goeppert B, Schirmacher P, Hoffmann H, Schnabel PA & Warth A 2012 Phenotyping of pulmonary carcinoids and a Ki-67-based grading approach. *Virchows Archiv* **460** 299–308. (doi:10.1007/s00428-012-1194-2)

Received in final form 2 October 2013
Accepted 25 October 2013

Published by Bioscientifica Ltd.