JNCI 16-0952R2

Article

Design and Validation of the GI-NEC Score to Prognosticate Overall Survival in Patients with High-Grade Gastrointestinal Neuroendocrine Carcinomas

Authors

Angela Lamarca[1], Thomas Walter[2], Marianne Pavel[3], Ivan Borbath[4], Patricia Freis[2], Barbara Nuñez[5], Alexa Childs[6], Mairéad G McNamara[1,7], Richard A Hubner[1], Rocio Garcia-Carbonero[5], Tim Meyer[6], Juan W. Valle[1,7,*], Jorge Barriuso[1,8,*]

[1] Department of Medical Oncology, The Christie NHS Foundation Trust (ENETS Centre of Excellence), Manchester, United Kingdom

[2] Department of Medical Oncology, Hospices Civils de Lyon Edouard Herriot Hospital, University of Lyon, Lyon, France

[3] Department of Hepatology and Gastroenterology, Charite´ University Medicine Berlin, Berlin, Germany

[4] Department of Gastroenterology, Cliniques Universitaires Saint-Luc, Bruxelles, Belgium

[5] Department of Medical Oncology, Doce de Octubre University Hospital, Madrid, Spain

[6] Department of Medical Oncology, Royal Free London NHS Foundation Trust (ENETS Centre of Excellence), London, United Kingdom

[7] Institute of Cancer Sciences, University of Manchester, Manchester, United Kingdom

[8] Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom

*These authors contributed equally to this work.

Corresponding author: Juan W Valle & Jorge Barriuso

- Professor Juan W Valle

  Institute of Cancer Sciences, University of Manchester, Manchester Academic Health Sciences Centre (MAHSC), Department of Medical Oncology, The Christie NHS Foundation Trust, Wilmslow Road, Manchester M20 4BX; Tel: +44 (0)161 446 8106; Fax: +44 (0)161 446 3468; e-mail: juan.valle@manchester.ac.uk

- Dr. Jorge Barriuso

  Faculty of Life Sciences, University of Manchester, Manchester, UK; Dover Street, M13 9PL, Manchester, UK; Phone number: 0044 (0)1612751586; Fax number: 0044 (0)1612751574; Jorge.Barriuso@manchester.ac.uk

Running head: GI-NEC Prognostic Score: Design and Validation

Abstract

Background

Prognostic markers for risk-stratification of patients with gastrointestinal high-grade neuroendocrine carcinomas (GI-NECs) are lacking; we designed and validated a prognostic score for overall survival (OS).

Methods

Consecutive patients, diagnosed in five neuroendocrine specialist European Centres were included. Patients were divided into three cohorts: a training cohort (TC), an external validation cohort (EVC) and a prospective validation cohort (PVC). Prognostic factors were identified by using Log-rank test, Cox-regression and logistic regression analyses. The derived score was internally and externally validated. All statistical tests were two-sided.

Results

Of 395 patients screened, 313 were eligible (TC: 109 patients, EVC: 184 patients and PVC: 20 patients). The derived prognostic score included five variables (presence of liver metastases, alkaline phosphatase (ALK), lactate dehydrogenase (LDH), ECOG performance status (PS) and Ki67). On multivariable analysis, the score was prognostic for OS (HR 1.86, 95%CI 1.47-2.35; p<0.001) and had good discrimination (C-index, 0.76) and calibration (mean error, 0.021; percentile 90, 0.037) in the TC. These results were validated in the EVC and PVC; in which it was able to prognosticate for OS when adjusted for other prognostic variables in the multivariable analysis (HR 1.85 (95%CI 1.27-2.71), p-value 0.001 and HR 4.51 (95%CI 1.87-10.87), p-value 0.001, respectively). The score classified patients into two groups with incremental risk of death: group A (0-2 points; 181 patients (63.9%); median OS 19.4 months [95%CI 16.1-25.1]) and B (3-6 points; 102 patients (36.1%); median OS 5.2 months [95%CI 3.6-6.9]).

Conclusion

The GI-NEC score identifies two distinct patient cohorts; it provides a tool for clinicians when making treatment decisions and may be used as a stratification factor in future clinical trials.

Introduction

Neuroendocrine malignancies from the gastrointestinal (GI) tract are relatively rare, although the incidence has been rising during recent years (1). Due to the impact on treatment strategy and survival, patients with neuroendocrine malignancies are classified according to both tumor morphology and assessment of proliferation according to WHO/ENETS guidelines. Morphology is classed as well- or poorly-differentiated; proliferation is assessed objectively by Ki-67 or mitotic count. Malignancies with a Ki-67 index >20% are considered high-grade neuroendocrine carcinomas (NEC-G3) (2-4). Well-differentiated tumours with a proliferation index of >20% have recently been described, so called NET-G3, as a discrete entity with clinical relevance (5, 6).

Gastrointestinal NECs (GI-NECs) represent only 5-10% of all digestive neuroendocrine malignancies (7, 8) and arise mainly from the stomach, pancreas, or colon (9-11). They are usually diagnosed in advanced stages, when only palliative treatment is available. In contrast to well-differentiated GI-NETs (1), the median survival of patients with NECs (all stages) is clinically significantly shorter (estimated to be around 12-17 months) due to their aggressive natural history (6, 12). Distant disease is present at initial diagnosis in 57% of patients and impacts on survival: the median survival is 38 months (95% confidence interval [95% CI] 31-45 months), 16 months (95% CI 15-17 months) and 5 months (95% CI 4.7-5.4 months) for patients with localised, locally advanced and metastatic disease, respectively (1, 13).

Management of GI-NECs with advanced disease, is based on systemic cytotoxic chemotherapy due to high mitotic activity and rapid rate of disease progression. Current treatments are based on data from small cell lung cancer, such as cisplatin-etoposide or carboplatin-etoposide (14). First-line treatment for GI-NECs has remained unchanged since the early 1990s, when high tumour response rates were reported with etoposide-platinum combination (41-67%) (15). In addition, a number of small retrospective series have published results of other chemotherapy regimens (temozolomide-based (12, 16), taxane-based (12), 5-

FU-based (17, 18) or topotecan (19)) after failure of platinum-etoposide therapy in patients with NECs (12, 16-20). Results from ongoing studies such as the ECOG-ACRIN 2142 trial (NCT02595424), which plans to randomise 126 patients diagnosed with NEC-G3 to first-line cisplatin/etoposide or capecitabine/temozolomide may clarify the most suitable treatment for this population of patients.

Prognostic factors have been reported in small series, but have not been externally validated. These include Ki67 (12), Eastern Cooperative Oncology Group performance status (ECOG-PS) (12, 21), elevated lactate dehydrogenase (LDH) (12, 21), primary site (12, 21, 22), thrombocytosis (12) and tumour morphology (6). It is worth highlighting that survival reported between different series vary from 20 months (16, 18, 20) to 3 months (19) suggesting marked heterogeneity of the patient populations. In addition, clinicians are lacking tools to identify patients who may have longer survival and therefore may benefit from active treatment or inclusion in clinical trials.

In this study, the aim was to design and validate a score prognostic for overall survival (OS) in patients with GI-NECs.

Methods

Study Design

Three cohorts of patients were analysed: a Training Cohort (score was designed and internally validated), an External Validation Cohort (score was externally validated) and a Prospective Validation Cohort (score was prospectively validated). Approval for data collection was obtained independently by each one of the institutions involved as per local practice.

Patients in cohort 1: Training cohort

All consecutive patients diagnosed with GI-NEC between January 1997 and June 2014 at The Christie NHS Foundation Trust (Manchester, UK) were included in this retrospective cohort. Eligible patients where those with a diagnosis of GI-NEC (including patients with an unknown

primary in whom the primary tumor was suspected to be of GI origin) with a Ki67 >20%; and available survival data. Site of primary tumor was classified as foregut (oesophagus, stomach and proximal duodenum; excluding pancreas), midgut (distal duodenum, appendix and proximal colon), pancreas, hindgut (colon and rectum) or unknown primary according to clinical information available. Patients with ECOG-PS 4 and patients with mixed adeno-neuroendocrine carcinoma were excluded. Demographic characteristics together with ECOG-PS, stage (23, 24), primary GI tumour site, sodium, alkaline phosphatase (ALP), LDH and Ki67 at time of first diagnosis were collected for identification of prognostic factors and design of the prognostic score (25, 26). The primary end-point was OS, defined as the time between first diagnosis of NEC and death (or last follow-up with no death).

Patients in cohort 2: External Validation Cohort

Patients diagnosed with GI-NECs who were seen between April 2000 and December 2015 were identified retrospectively in five different European Countries (France, Belgium, Germany, Spain and UK) from centres with expertise in neuroendocrine malignancies. The same inclusion criteria as the Training Cohort were used and the same baseline and demographic characteristics were collected. Patients with missing data in any one of the items included in the score were excluded and considered ineligible. The primary end-point was OS. A sample size for the external validation was estimated to replicate the hazard ratio achieved by the score (as a discrete variable) in the multivariable analysis from the Training Cohort (HR 1.9); a minimum of 82 patients was required to externally validate the score (power: 80% and two-tailed α-error: 0.05).

Patients in cohort 3: Prospective Validation Cohort

The Prospective Validation Cohort included all consecutive patients diagnosed with GI-NECs seen at The Christie NHS Foundation Trust between July 2014 and November 2015. Patients were identified and data was collected prospectively. The same inclusion criteria and data

collection items as in the External Validation Cohort were applied; patients with missing data for any of the score items were excluded. The primary end-point was OS.

Description of baseline characteristics and comparison between cohorts

The median, with range and/or 95% CI, was calculated for continuous/discrete variables. Percentages were employed for distribution of categorical variables. Chi-square and T-Tests were used for comparison of baseline characteristics, as appropriate. Comparisons with p-value ≤0.05 were considered statistically significant.

Statistical analysis

Statistical analysis was performed with Stata v.12 and RStudio packages. All statistical tests were two-sided.

Design of the prognostic score (Training Cohort)

Cox-regression, Kaplan-Meier, log-rank test and logistic regression were employed for identification of relevant factors impacting on OS. Proportionality of hazards assumption was assessed by graphic methods such as Log-log plot of survival and Kaplan–Meierobserved/ predicted survival plot. A maximum regression model including all the statistically significant variables in the univariate Cox-regression (defined as p-value ≤0.05) and previously-defined "variables of interest" (ECOG-PS, stage and Ki67) was designed. Treatment-related variables (such as administration of chemotherapy or radiological response) were excluded from the maximum regression model. The prognostic score included selected items from the maximum regression model, which were chosen by the *allsets* Stata command (selecting the model with lower Akaike Information Criterion (AIC), lower number of variables and without duplicated clinical information).

Once the variables to be included in the prognostic score were identified, a score nomogram was built. First, continuous/discrete variables were categorised taking into account the most suitable cutoff according to Receiver Operating Characteristic (ROC) analysis for prediction of death. Second, Kaplan-Meier curve for prediction of OS, to confirm the tested

cutoff as the most informative one in terms of our primary end-point (OS). Finally, the punctuation for each item's category was selected to be proportional to the hazard ratio achieved in the multivariable Cox-regression analysis.

Accuracy of prediction of risk of survival: prognostic score compared to the maximum regression model (Training Cohort)

The impact of the score on survival was confirmed by multivariable Cox-regression adjusted for other variables with known prognostic impact that were not included in the score. ROC curve comparison test (by comparison of Area Under the Curve (AUC)) was employed to compare the accuracy of prediction of survival at 3, 6, 9, 12, 18 and 24 months of the maximum model (considered the gold standard) and the prognostic score. Comparisons with p-values ≤0.05 were considered statistically significant.

Internal validation of the prognostic score (Training Cohort)

Internal validation of the score was performed by Bootstrap-corrected Harrell Concordance Index (C-Index) calculation and Resampling Model Calibration. For interpretation of results of the C-Index, an index of 1 was considered to be perfectly discriminating, while a C-Index of 0.5 was as good as a random estimation.

External validation of the prognostic score (External and Prospective Validation Cohorts)

The previously-designed score was applied to all patients in the External and Prospective Validation Cohorts. Cox-regression, Kaplan-Meier and log-rank tests were employed to identify factors impacting on OS. Multivariable Cox-regression analyses performed in the Training Cohort were reproduced to validate the impact of the score on OS.

Results

A total of 395 patients were identified and considered for eligibility. Of these, 313 patients with GI-NECs were eligible and included in the final analysis: Training Cohort (109 patients),

External Validation Cohort (184 patients) and Prospective Validation Cohort (20 patients) (See CONSORT diagram, **Figure 1**).

Training Cohort: Patient's baseline characteristics, prognostic factors and design of prognostic score

The Training Cohort included 109 patients; baseline characteristics are summarised in **Table1**. By the end of follow-up, 83.5% of patients had died. The median estimated OS was 11.3 months (95% CI 8.3-14.03).

In addition to Ki67 (p-value 0.15) and ECOG-PS (p-value 0.046 for ECOG-PS 2 compared to ECOG-PS 0), which were previously defined as "variables of interest"; the following seven variables were found to be prognostic for OS by univariate Cox-regression analysis (**Supplementary Table 1**): stage (p-value 0.008 for metastatic stage compared to localized stage), LDH (p-value <0.001), sodium (p-value 0.003), ALP (p-value <0.001), number of metastatic sites (p-value 0.021), presence of liver metastases (p-value <0.001) and presence of lung metastases (p-value 0.031). All these variables were included in the multivariable maximum regression model (**Supplementary Table 1**).

Following analysis of 511 potential combinations, the most informative reduced model was selected to design the prognostic score. The selected reduced model included five variables (liver metastases, ECOG-PS, Ki67, LDH and ALP (**Supplementary Table1**) and had the following characteristics: AIC of 71.6, AUC of 0.839 and Hosmer-Lemeshow goodness-of-fit of 0.977. Punctuation for each item was selected taking into account the HR from the multivariable analysis of the reduced model (**Supplementary Table 1**) and **Figure 2**). According to the sum of the five items, patients could be assigned a minimum of 0 to a maximum of 6 points. Risk of death, increased with each extra point added into the score (**Supplementary Figure 1**). Distribution of the prognostic score within TC population is summarized in **Supplementary Table 2**.

Internal Validation of the Score and Accuracy of prediction of risk of survival: prognostic score compared to the maximum regression model (Training Cohort)

The score was internally validated showing good discrimination (C-Index of 0.76) and calibration (mean error 0.021; percentile 90 0.037; **Figure 3**).

There was good accuracy for prediction of risk of survival when compared to the maximum regression model. There were no differences between the prognostic score and the maximum regression model at 3, 6, 9, 12, 18 and 24 months (**Supplementary Figure 2**).

Impact of the Score on OS

Multivariable Cox-regression analysis, including the score as one of the variables (**Table 2**), was performed to assess its impact on survival when adjusted for other variables not included in the score such as sodium (independent prognostic factor identified during the score building process (**Supplementary Table1**) and stage (variable of interest). However lung metastases were another of the independent prognostic factor identified during the score building process (**Supplementary Table1**), this variable was not included in this multivariable analysis (**Table 2**), because this clinical information was already provided by other variables such as presence of liver metastases (included in the score) and stage (defined as variable of interest and therefore included in multivariable analysis). The score was an independent prognostic factor for OS (HR 1.86, 95% CI 1.47-2.35; p-value <0.001) when the analysis was adjusted for other variables, such as stage and sodium (**Table 2**).

External and Prospective Validation of the Score

One hundred and seventy-eight and 20 patients were eligible for inclusion in the External Validation and in the Prospective Validation Cohorts, respectively. Baseline characteristics of these validation cohorts are shown in **Table 1**. Distribution of the prognostic score within these two populations of patients is summarized in **Supplementary Table 2**.

The multivariable analysis (including the score as a prognostic variable) performed on the Training Cohort was replicated with each of the validation cohorts (**Table 2**). The score

remained an independent prognostic factor in both validation cohorts when adjusted for stage and sodium (previously showed to be prognostic variables in the univariate analysis): HR 1.85 (95%CI 1.27-2.71; p-value 0.001) and HR 4.51 (95%CI 1.87-10.87; p-value 0.001), for EVC and PVC, respectively.

Deriving the Score in clinical practice

The prognostic score has been defined as a discrete variable, with a range from 0 to 6 points. For applicability in clinical practice, the score classified patients into two groups, with statistically significant differences in OS: group A (0-2 points; good prognosis patients; 181 patients (63.9%); median OS 19.4 months [95%CI 16.1-25.1]) and group B (3-6 points; poor prognosis patients; 102 patients (36.1%); median OS 5.2 months [95%CI 3.6-6.9]) (**Figure 4**). When the multivariable analysis was repeated using the score as a dichotomised variable, the score (group A vs. B) was an independent prognostic factor in all patient cohorts (**Table 2**). Distribution of the prognostic score as a dichotomised variable within all populations of patients included in this is summarized in **Supplementary Table 2**.

Receipt of chemotherapy had an impact on OS in both prognostic groups and was an independent positive prognostic factor (HR for patients receiving chemotherapy was 0.31 (95% CI 0.20-0.48); p-value <0.001) when adjusted for the prognostic score, stage and sodium in the multivariable analysis. This resulted in the following median OS: Group A treated with chemotherapy (129 patients, median OS 20.7 months (95% CI 16.7-26.4), Group A not treated with chemotherapy (24 patients, median OS 11.9 months (95% CI 4.6-81.1), Group B treated with chemotherapy (55 patients, median OS 8.1 months (95% CI 5.4-10.8), Group B not treated with chemotherapy (26 patients, median OS 2.1 months (95% CI 1.4-2.4).

Discussion

This study represents one of the largest series of patients diagnosed with high grade NECs to date (6, 12). High-grade NECs are known to consist of a heterogeneous population (27), in

whom prognosis depends on multiple factors (6, 12, 21), rather than unique variables; thus our interest in developing a prognostic score which would combine the impact of multiple variables (all objective, with the exception of performance status, which is subjective) in one unique measurement.

The final GI-NEC Score showed prognostic impact, and its ability to identify a patient group with an inferior outcome was validated in an external cohort. This study constitutes a collaborative effort between five European Centres with recognised expertise in neuroendocrine malignancies giving credence to the results presented.

The direct application of the GI-NEC Score identified two subgroups, a good prognosis (A) and poor prognosis (B) group with marked differences in OS. The magnitude of difference indicates that this score may not only have a role in determining prognosis and aiding treatment discussion between clinicians and patients, but may also influence clinical trial design with respect to patient stratification.

Both prognostic groups benefitted from treatment with systemic chemotherapy. This benefit was more pronounced in group B patients (4-fold OS increase) than in group A patients (2-fold OS increase); however, differences due to patient selection bias cannot be excluded. Group B identified cases with behaviour similar to Small Cell lung Cancer (SCLC) in terms of prognosis, who may derive the greatest benefit from classical platinum plus etoposide chemotherapy combinations.

The identification of two patient cohorts with apparently-different clinical courses highlights the need to consider different management strategies between them. For instance, patients classified in group A, may be suitable for treatment options beyond platinum-etoposide chemotherapy including alternative chemotherapy combinations for patients with a lower Ki67 or morphologically well-differentiated tumours (6, 12); or locoregional-directed therapy in cases with liver-predominant disease. In essence, although there is often urgency in starting platinum-etoposide chemotherapy in patients with GI-NEC, patients in group A may

benefit from case-by-case discussion within a NET multidisciplinary meeting regarding other treatment options. Furthermore, due to the absence of available clinical trials in this setting and the more-favourable survival of this particular group, these patients would be well-suited for early-phase trials.

It is worth mentioning that the most suitable Ki67 cut-off for making clinical management decisions remains unclear. In the NORDIC retrospective study, it was suggested that a Ki67 cut-off of 55% was informative for choosing between platinum-based or temozolomide-based treatment (12). In contrast, the current series identified a different cut-off (80%) as the most informative one with respect to OS (rather than treatment response). These results highlight that it is not advisable to make informed treatment decisions reliant on Ki67 alone. Morphology should also be considered, as suggested by Heetfeld and colleagues (6).

The Group A population still likely constitutes several different subgroups with differing molecular biology. These patients may derive benefit from further characterisation (28). A worldwide collaborative effort to better understand the nature of this pathology should be prioritised by the NET research community.

On the other hand, patients classified in the poor prognosis group, in the absence of other alternatives, may derive the most benefit from therapy traditionally given to patients with SCLC. Our results suggest that in this poor-prognostic subgroup, there is an urgent priority to initiate chemotherapy treatment; which may be less pressing in Group A patients.

This new prognostic tool may become a cornerstone in the stratification of patients for inclusion into clinical trials as a randomisation factor, allowing adjustment for multiple prognostic factors at once. This will allow for efficient reduction of stratification variables and may potentially reduce trial sample size and facilitate subgroup analysis. Trial design may also be influenced to develop distinct therapeutic approaches based on the clinical characteristics of both subgroups when evaluating and introducing new therapies. For instance, if we were to

design a trial with new immunotherapies (such us anti-PD1 targeted therapy, which has recently shown benefit in Merkel cell carcinoma (29)) a different approach may be considered in group B patients with induction chemotherapy before introducing the immunotherapy to achieve clinical control of the disease, thus optimising the use of chemotherapy. However, most patients in group A would not need this sequential approach.

The limitations of this study are the ones which inherently apply to population data which are collected retrospectively (Training Cohort and External Validation Cohort), as shown by imbalances identified between patient cohorts (such as primary tumour site, Ki67 and median OS). Despite of the small sample size of the Prospective Validation cohort, the score retains its prognostic value; however, it would still require confirmation in a longer prospective series, ideally within a clinical trial. In addition, the depth of this analysis is limited to the use of clinical data; i.e. chromogranin-A and neurone-specific enolase, were not included due to data unavailability. The absence of single central pathology review is attenuated by the documented expertise of these centres and their pathology departments. The fact that patients with no survival data or at least one of the score items missing were excluded from the EVC and PVC could introduce a selection bias. Finally, tumour morphology (well differentiated high grade vs. poorly differentiated high grade) was not incorporated into this score due to non-availability of this information for the majority of patients. Patients with the emerging entity of NET-G3 are most likely to be included in Group A; the high Ki67 cut-off employed in this study (80%) may negate this limitation due to the low rate of patients with well-differentiated morphology who would be expected to have a Ki67 of >80% (6).

In summary, in this study, the GI-NEC Score was designed and validated for patients with GI-NECs for prognostication of OS, as a tool to aid clinician selection of patients for treatment or clinical trial inclusion. This score could also be incorporated as a stratification factor in future comparative trials, to assist adjustment for prognostic factors and secure comparability between arms.

Funding: A Lamarca was part-funded by Pancreatic Cancer Research Fund and SEOM (Spanish Society of Medical Oncology); Jorge Barriuso was part-funded by SEOM (Spanish Society of Medical Oncology). TM is part funded by the NIHR UCLH Biomedical Research Centre. AC is funded by the UCL Comprehensive Cancer Imaging Centre.

Notes: The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

This study has been partially presented at ASCO (American Society of Clinical Oncology) Annual Meeting 2015, UK and Ireland Neuroendocrine Tumour Society (UKINETS) 2015, European Neuroendocrine Tumour Society (ENETS) 2016 and ASCO Annual Meeting 2016 (Merit Award Winner).

Authors declare no conflict of interest related to this study

REFERENCES

1. Yao JC, Hassan M, Phan A, et al.: One hundred years after "carcinoid": epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. J Clin Oncol 26:3063-3072, 2008
2. Brenner B, Tang LH, Shia J, et al.: Small cell carcinomas of the gastrointestinal tract: clinicopathological features and treatment approach. Semin Oncol 34:43-50, 2007
3. Scoazec JY, Couvelard A: [The new WHO classification of digestive neuroendocrine tumors]. Ann Pathol 31:88-92, 2011
4. Rindi G, Arnold R, Bosman FT, et al. Nomenclature and classification of neuroendocrine neoplasms of the digestive system. In: WHO Classification of Tumours of the

Digestive System, 4th ed, Bosman TF, Carneiro F, Hruban RH, Theise ND (Eds), International Agency for Research on cancer (IARC), Lyon 2010. p.13, 2013

5. Tang LH, Basturk O, Sue JJ, et al.: A Practical Approach to the Classification of WHO Grade 3 (G3) Well-differentiated Neuroendocrine Tumor (WD-NET) and Poorly Differentiated Neuroendocrine Carcinoma (PD-NEC) of the Pancreas. Am J Surg Pathol, 2016
6. Heetfeld M, Chougnet CN, Olsen IH, et al.: Characteristics and treatment of patients with G3 gastroenteropancreatic neuroendocrine neoplasms. Endocr Relat Cancer 22:657-664, 2015
7. Ahlman H, Nilsson O, McNicol AM, et al.: Poorly-differentiated endocrine carcinomas of midgut and hindgut origin. Neuroendocrinology 87:40-46, 2008
8. Bettini R, Boninsegna L, Mantovani W, et al.: Prognostic factors at diagnosis and value of WHO classification in a mono-institutional series of 180 non-functioning pancreatic endocrine tumours. Ann Oncol 19:903-908, 2008
9. Brenner B, Tang LH, Klimstra DS, et al.: Small-cell carcinomas of the gastrointestinal tract: a review. J Clin Oncol 22:2730-2739, 2004
10. Galanis E, Frytak S, Lloyd RV: Extrapulmonary small cell carcinoma. Cancer 79:1729-1736, 1997
11. Lee SS, Lee JL, Ryu MH, et al.: Extrapulmonary small cell carcinoma: single center experience with 61 patients. Acta Oncol 46:846-851, 2007
12. Sorbye H, Welin S, Langer SW, et al.: Predictive and prognostic factors for treatment and survival in 305 patients with advanced gastrointestinal neuroendocrine carcinoma (WHO G3): the NORDIC NEC study. Ann Oncol 24:152-160, 2013
13. Sorbye H, Strosberg J, Baudin E, et al.: Gastroenteropancreatic high-grade neuroendocrine carcinoma. Cancer 120:2814-2823, 2014

14. Mitry E, Baudin E, Ducreux M, et al.: Treatment of poorly differentiated neuroendocrine tumours with etoposide and cisplatin. Br J Cancer 81:1351-1355, 1999

15. Moertel CG, Kvols LK, O'Connell MJ, et al.: Treatment of neuroendocrine carcinomas with combined etoposide and cisplatin. Evidence of major therapeutic activity in the anaplastic variants of these neoplasms. Cancer 68:227-232, 1991

16. Hadoux J, Malka D, Planchard D, et al.: Post-first-line FOLFOX chemotherapy for grade 3 neuroendocrine carcinoma. Endocr Relat Cancer 22:289-298, 2015

17. Hentic O, Hammel P, Couvelard A, et al.: FOLFIRI regimen: an effective second-line chemotherapy after failure of etoposide-platinum combination in patients with neuroendocrine carcinomas grade 3. Endocr Relat Cancer 19:751-757, 2012

18. Hadoux J PDGJea: Oxaliplatin-based chemotherapy for grade 3 neuroendocrine carcinoma after failure of platinum-based chemotherapy. European Neuro Endocrine Tumour Society: Abstract J2, 2013., in , 2016

19. Olsen IH, Knigge U, Federspiel B, et al.: Topotecan monotherapy in heavily pretreated patients with progressive advanced stage neuroendocrine carcinomas. J Cancer 5:628-632, 2014

20. Welin S, Sorbye H, Sebjornsen S, et al.: Clinical effect of temozolomide-based chemotherapy in poorly differentiated endocrine carcinoma after progression on first-line chemotherapy. Cancer 117:4617-4622, 2011

21. Yamaguchi T, Machida N, Morizane C, et al.: Multicenter retrospective analysis of systemic chemotherapy for advanced neuroendocrine carcinoma of the digestive system. Cancer Sci 105:1176-1181, 2014

22. Strosberg JR, Cheema A, Weber J, et al.: Prognostic validity of a novel American Joint Committee on Cancer Staging Classification for pancreatic neuroendocrine tumors. J Clin Oncol 29:3044-3049, 2011

23. Edge SBDRCCCFAGGFLTA: AJCC Cancer Staging Manual, in , Springer-Verlag New York, 2010

24. Salazar R, Wiedenmann B, Rindi G, et al.: ENETS 2011 Consensus Guidelines for the Management of Patients with Digestive Neuroendocrine Tumors: an update. Neuroendocrinology 95:71-73, 2012

25. Maestu I, Pastor M, Gomez-Codina J, et al.: Pretreatment prognostic factors for survival in small-cell lung cancer: a new prognostic index and validation of three known prognostic indices on 341 patients. Ann Oncol 8:547-553, 1997

26. Cerny T, Blair V, Anderson H, et al.: Pretreatment prognostic factors and scoring system in 407 small-cell lung cancer patients. Int J Cancer 39:146-149, 1987

27. Velayoudom-Cephise FL, Duvillard P, Foucan L, et al.: Are G3 ENETS neuroendocrine neoplasms heterogeneous? Endocr Relat Cancer %19;20:649-657, 2013

28. Emily K.Bergsland RRPSJSRMBAO: Genomic profiling to distinguish poorly differentiated neuroendocrine carcinomas arising in different sites. ASCO Annual Meeting 2016 J Clin Oncol 34, 2016 (suppl; abstr 4020), 2016 (abstr)

29. Nghiem PT, Bhatia S, Lipson EJ, et al.: PD-1 Blockade with Pembrolizumab in Advanced Merkel-Cell Carcinoma. N Engl J Med %19., 2016

Table 1: Baseline characteristics of cohort of patients included in the study*

Variable	Training Cohort (n=109)	External Validation Cohort (n=184)	Prospective Validation Cohort (n=20)
	N (%)	**N** (%)	**N (%)**
Age, Median (range)	67.7 (16.3-84.1)	59.1 (22.2-85.5)	67.7 (22.9-82.7)
Gender			
Male	67 (61.5)	115 (62.5)	15 (75.0)
Female	42 (38.5)	69 (37.5)	5 (25.0)
Stage			
Localised	11 (10.1)	9 (4.9)	2 (10.0)
Locally advanced	7 (6.4)	26 (14.1)	3 (15.0)
Metastatic	91 (83.5)	149 (80.9)	15 (75.0)
Site of primary tumour			
Foregut	21 (19.3)	33 (17.9)	1 (5.0)
Midgut	5 (4.6)	8 (4.4)	4 (20.0)
Pancreas	20 (18.4)	75 (40.8)	4 (20.0)
Hindgut	31 (28.4)	31 (16.9)	5 (25.0)
Unknown primary	32 (29.4)	37 (20.1)	6 (30.0)
Site of metastases			
Median (range)	1 (0-4)	1 (0-6)	1 (0-4)
Liver (Yes)	70 (64.2)	123 (66.9)	12 (60.0)
Lungs (Yes)	10 (9.2)	12 (6.5)	3 (15.0)

Lungs (unknown)	0	79	0
Sodium, mmol/L			
Median (range)	139 (124-145)	138 (129-144)	139.5 (133-143)
Unknown	0	115	0
Alkaline Phosphatase, IU/l			
Median (range)	109 (45-2.03x10^3)	129.5 (30-2.3x10^3)	103 (64-2.04x10^3)
Unknown	14	0	0
Lactate dehydrogenase, IU/l			
Median (range)	463 (258-11.1x10^3)	347 (107-3.8x10^3)	450 (258-1.4x10^3)
Unknown	17	0	0
Ki67, %			
Median (range)	70 (60-70)	55 (50-62)	80 (41-80)
Unknown	15	0	0
ECOG Performance Status			
0	28 (25.9)	65 (35.3)	3 (15.0)
1	57 (52.8)	82 (44.6)	10 (50.0)
2	17 (15.7)	27 (14.7)	3 (15.0)
3	6 (5.6)	10 (5.4)	4 (20.0)
Unknown	1 (0.1)	0 (0.0)	0 (0.0)

Palliative Chemotherapy			
Yes	76 (69.7)	118 (64.1)	14 (70.0)
Unknown	0 (0.0)	49 (26.6)	0 (0.0)
Follow-up, median (95% CI)	9.7 (7.02-12.3)	11.4 (8.6-14.7)	5.01 (3.5-12.7)
Status, dead	91 (83.5)	127 (69.1)	11 (55.0)
Overall survival, median (95% CI)	11.3 (8.3-14.03)	16.04 (13.7-19.03)	10.1 (3.5-nr)

* Overall survival medians are estimated with Kaplan-Meier. Normal ranges for laboratory results were as follows sodium 135-145 mmol/L; alkaline phosphatase 25-110 IU/L; lactate dehydrogenase 200-550 IU/L. Nr=not reached, CI=confidence interval.

Table 2: Multivariable prognostic analysis for overall survival (Cox regression). *

	Training Cohort		External Validation Cohort		Prospective Validation Cohort		All Cohorts	
	HR (95% CI)	**P-value†**	**HR (95% CI)**	**P-value†**	**HR (95% CI)**	**P-value†**	**HR (95% CI)**	**P-value†**
Score as a discrete variable								
Stage								
Localised	1 (Ref)	-	1 (Ref)	-	1 (Ref)	-	1 (Ref)	-
Locally advanced	0.88 (0.16 - 4.91)	0.88	0.81 (0.19-3.36)	0.77	0.17 (0.01-5.63)	0.32	0.84 (0.32-2.19)	0.88
Metastatic	2.11 (0.73 – 6.12)	0.17	0.91 (0.24-3.45)	0.88	0.11 (0.01-1.73)	0.12	1.28 (0.59-2.73)	0.45
Sodium, continuous variable	1.02 (0.95 - 1.08)	0.64	0.97 (0.86-1.09)	0.61	0.77 (0.57-1.05)	0.11	1.01 (0.96-1.06)	0.72
Score, discrete variable	1.86 (1.47 - 2.35)	< 0.001	1.85 (1.27-2.71)	0.002	4.51 (1.87-10.87)	0.001	1.95 (1.62-2.36)	<0.001
Score as a dichotomised variable								
Stage								
Localised	1 (Ref)	-	1 (Ref)	-	1 (Ref)	-	1 (Ref)	-
Locally advanced	1.29 (0.23 - 7.24)	0.77	0.82 (0.21-3.38)	0.93	0.32 (0.02-7.22)	0.47	0.88 (0.34-2.29)	0.79
Metastatic	2.57 (0.89 – 7.46)	0.08	1.27 (0.35-4.61)	0.62	0.15 (0.01-2.16)	0.16	1.64 (0.77-3.51)	0.20

Sodium, continuous variable	0.99 (0.93-1.06)	0.79	0.97 (0.86-1.09)	0.77	0.87 (0.67-1.13)	0.29	0.99 (0.94-1.04)	0.74
Score								
Group A (0-2 points)	1 (Ref)	-	1 (Ref)	-	1 (Ref)	-	1 (Ref)	-
Group B (3-6 points)	3.61 (1.96-6.65)	<0.001	3.39 (1.57-7.31)	<0.001	14.33 (2.28 -90.13)	0.005	3.94 (2.53-6.13)	<0.001

* Only patients with all data available for all the variables included in the multivariable analysis were included: Training Cohort (79 patients), External validation Cohort (63 patients), Prospective Validation Cohort (20 patients) and all cohorts (162 patients). CI=confidence interval; HR=hazard ratio.

† [Please name the statistical test used to calculate the P values and say if that test was two-sided.]

FIGURES

Figure 1: CONSORT diagram: patient flow for each cohort of patients included in the design and validation of the prognostic score. N: number of patients; GI: gastrointestinal; NEC: neuroendocrine carcinoma (high grade); ECOG-PS: Eastern Cooperative Oncology Group performance status; OS: overall survival.

Figure 2: Nomogram for calculation of the prognostic score. Example: patient with ECOG-PS of 1 (0 points), liver metastases (1 point), ALP of 70 (0 points), LDH of 840 (1 point) and Ki67 of 75 (0 points) scores a total of 2 points. ALP: alkaline phosphatase; LDH: lactate dehydrogenase, ECOG-PS: Eastern Cooperative Oncology Group performance status.

Figure 3: The Prognostic Score showed good calibration (Mean (error) = 0.021; Percentile 90= 0.037).

Figure 4: Prognostic score defining two populations of patients with statistically significant differences in overall survival: this is shown when all cohorts are analysed together (**A**) and also in each cohort separately: Training Cohort (**B**), External Validation Cohort (**C**) and Prospective Validation Cohort (**D**). Log-rank test p-value is provided, for multivariable Cox Regression p-value please refer to **Table 2**. Only those patients with available data for calculation of the GI-NEC Score (all items were required) and survival data available are represented in these graphics, represented by 283 patients out of the total 313 patients included in All Cohorts, 79 out of 109 in the Training Cohort, 184 out of 184 in the External Validation Cohort and 20 out of 20 in the patients from the Prospective Validation Cohort), patients with at least one of the items or survival data missing (i.e. 30 patients from the Training Cohort (**B**) and therefore 30 patients from the All Cohorts (**A**)) are excluded. All statistical tests were two-sided.