



Published in final edited form as:

Ann Surg. 2014 February ; 259(2): 204–212. doi:10.1097/SLA.0b013e31828f3174.

A Single Institution's 26-Year Experience With Nonfunctional Pancreatic Neuroendocrine Tumors:

A Validation of Current Staging Systems and a New Prognostic Nomogram

Trevor A. Ellison, MD^{*}, Christopher L. Wolfgang, MD, PhD^{*}, Chanjuan Shi, MD, PhD[§], John L. Cameron, MD^{*}, Peter Murakami, ScM[¶], Liew Jun Mun, MBBS[†], Aatur D. Singhi, MD, PhD[†], Toby C. Cornish, MD, PhD[†], Kelly Olino, MD^{*}, Zina Meriden, MD[†], Michael Choti, MD, MBA^{*‡}, Luis A. Diaz, MD[‡], Timothy M. Pawlik, MD, MPH^{*}, Richard D. Schulick, MD^{||}, Ralph H. Hruban, MD^{†‡}, and Barish H. Edil, MD^{||}

^{*}Department of Surgery, The Sol Goldman Pancreatic Cancer Research Center, the Johns Hopkins University School of Medicine, Sidney Kimmel Comprehensive Cancer Center, Baltimore, MD

[†]Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, the Johns Hopkins University School of Medicine, Sidney Kimmel Comprehensive Cancer Center, Baltimore, MD

[‡]Department of Oncology, The Sol Goldman Pancreatic Cancer Research Center, the Johns Hopkins University School of Medicine, Sidney Kimmel Comprehensive Cancer Center, Baltimore, MD

[§]Vanderbilt University Medical Center, Department of Pathology, Microbiology and Immunology, Nashville, TN

[¶]The Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, Baltimore, MD

^{||}The University of Colorado, Department of Surgery, Aurora, CO

Abstract

Objective—To validate the 2010 American Joint Committee on Cancer (AJCC) and 2006 European Neuroendocrine Tumor Society (ENETS) tumor staging systems for pancreatic neuroendocrine tumors (PanNETs) using the largest, single-institution series of surgically resected patients in the literature.

Background—The natural history and prognosis of PanNETs have been poorly defined because of the rarity and heterogeneity of these neoplasms. Currently, there are 2 main staging systems for PanNETs, which can complicate comparisons of reports in the literature and thereby hinder progress against this disease.

Copyright © 2013 by Lippincott Williams & Wilkins

Reprints: Barish H. Edil, MD, Department of Surgery, University of Colorado, 1665 Aurora Court, Room 3337, Aurora, CO 80045. barish.edil@ucdenver.edu.

Disclosure: The authors declare no conflicts of interest.

Methods—Univariate and multivariate analyses were conducted on the prognostic factors of survival using 326 sporadic, nonfunctional, surgically resected PanNET patients who were cared for at our institution between 1984 and 2011. Current and proposed models were tested for survival prognostication validity as measured by discrimination (Harrel's c-index, HCI) and calibration.

Results—Five-year overall-survival rates for AJCC stages I, II, and IV are 93% (88%–99%), 74% (65%–83%), and 56% (42%–73%), respectively, whereas ENETS stages I, II, III, and IV are 97% (92%–100%), 87% (80%–95%), 73% (63%–84%), and 56% (42%–73%), respectively. Each model has an HCI of 0.68, and they are no different in their ability to predict survival. We developed a simple prognostic tool just using grade, as measured by continuous Ki-67 labeling, sex, and binary age that has an HCI of 0.74.

Conclusions—Both the AJCC and ENETS staging systems are valid and indistinguishable in their survival prognostication. A new, simpler prognostic tool can be used to predict survival and decrease interinstitutional mistakes and uncertainties regarding these neoplasms.

Keywords

AJCC; ENETS; grade; nomogram; nonfunctional; pancreatic neuroendocrine tumors; staging systems

Pancreatic cancer is the second most common gastrointestinal malignancy in the United States and the fourth leading cause of deaths due to cancer in adults. Although the vast majority of pancreatic cancer cases are pancreatic adenocarcinoma, pancreatic neuroendocrine tumors (PanNETs) make up 3% to 5% of all pancreas cancers.¹ The incidence of PanNETs is less than 0.25 to 1 in 100,000 people per year in the United States and the incidence has almost doubled over the past few decades partly because of the increased use of computed tomographic scans.^{2–8} The natural history and prognosis of these tumors remain ill-defined because of their rarity and heterogeneity. PanNETs are heterogeneous in both histology and clinical presentation. Histology ranges from poorly differentiated tumors, which portend very limited survival, to well-differentiated tumors (representing approximately 90% of cases), which have clinical behavior ranging anywhere from indolent to highly malignant.⁹ Clinical presentation ranges from genetic syndromes (eg, multiple endocrine neoplasm type I, von-Hippel Lindau) to functional tumors (eg, insulinomas, glucagonomas, somatostatinomas, VIPomas) and to non-syndromic/nonfunctional neoplasms. The heterogeneity of these neoplasms can further be appreciated in the reported 5-year survivals of 97% for insulinomas to 30% for nonfunctioning metastatic tumors.¹⁰ Of note, another hurdle to the understanding of the incidence, natural history, and prognosis of this rare entity has been that all PanNETs have not been captured in national cancer registries. Certain PanNETs that would be included today were excluded from the national cancer registries before 2000 as they were considered to have a “benign” or “uncertain” clinical course.³

The history of classifying, staging, and grading PanNETs has undergone a complex evolution. Williams and Sandler¹¹ classified carcinoid neoplasms in 1963 based on the perceived embryological origin of these neoplasms in the fore-, mid- and hind-gut,¹² but it

was not until 1995 that Capella et al¹³ specifically classified PanNETs by histologic differentiation. The World Health Organization (WHO) then reclassified these tumors in 2000, 2004, and 2010 resulting in a system with both prognostic accuracy and reproducibility in clinical practice.^{14–16} In 2006, The European Neuroendocrine Tumor Society (ENETS) established and validated another staging and grading system in an attempt to rectify perceived faults in the WHO classification.¹⁶ In 2010, the American Joint Committee on Cancer (AJCC) first staged PanNETs and employed the same staging system as they used for exocrine pancreas malignancies.¹ In the face of all these staging systems, potential confusion and communication difficulties exist.¹⁷ A recent study found that the difference in T classification between the AJCC and ENETS systems were relatively frequent and heterogeneous although the differences did not lead to inappropriate application of any diagnostic, therapeutic, or follow-up protocols.¹⁸

In summary, the understanding of PanNETs has been hindered by its rarity, heterogeneity, and differing staging systems. We set out to use the largest series of a single institution's surgically resected, nonsyndromic, and nonfunctional PanNETs to validate and compare the AJCC and ENETS staging systems. We also propose to evaluate the importance of certain prognostic factors in PanNET staging to the end of possibility developing an alternative system with more accuracy in predicting survival. We limited our study to the nonsyndromic and nonfunctional PanNETs as they account for the largest portion (77%) of our surgical series.

METHODS

Institutional review board approval was obtained for this study. Demographic, operative, and clinical data were compiled from our institution's electronic patient record and our surgical department's prospectively collected pancreas database of surgically resected patients from 1984 to 2011. All patients who were diagnosed with PanNET by histopathology were included, whereas all functional PanNET patients (68 patients), syndromic patients (27 patients), patients who had incidental PanNETs found in the context of pancreatic or ductal adenocarcinoma (3 patients) or patients whose PanNETs were incidental to the pathology that resulted in the operation (24 patients) were excluded (all incidental PanNET ranged in size from 1 to 8 mm). This resulted in 326 nonsyndromic, nonfunctional PanNET patients. Information from the Johns Hopkins Hospital Cancer Registry (for the available years of 1995–2010) was collected for 240 of the 326 patients (74%) on recurrence, reoperations, and death. Death information was also obtained from the Social Security Death Index. Phone interviews were conducted for all patients with a response rate of 45% (146 patients) and data on survival status, date of death, tumor recurrence, and reoperation for recurrence was collected.

Ninety-five percent of the histopathology slides (309) were read since 2009 with a single pathologist rereading 86% (281) of the historical slides. The results of immunolabeling for Ki-67 were available for 85% of the cases and when rereviewing histopathology slides, the Ki-67 index was calculated using custom software written in ImageJ (Wayne Rasband, NIH, Bethesda, MD) to assist in the nuclear counts. Five images were acquired per case at a magnification of 400× using a Q-Color3 digital camera (Olympus, Center Valley, PA) on an

Olympus B-50 microscope (Center Valley, PA). Fields were selected that represented the highest density of Ki-67 positive cells and a mean of 978 (standard error of the mean = 34) nuclei were counted per case. For this study, necrosis was recorded as either present or absent.

Validation and comparison in survival prognostication of the 2010 AJCC and 2006 ENETS staging systems were accomplished using our patient population. Kaplan-Meier curves depicting overall survival (OS) were computed with the log-rank test being used to verify significance in the differences in survival curves. Median survival times were calculated per stage using a Weibull proportional hazards model. A “Harrell’s c-index” (HCI) was calculated for each staging system as a measure of the predictive accuracy of the survival outcome. The reported HCIs are the bias-corrected values using bootstrap validation. HCI test statistics were compared by taking their difference, generating a bootstrapped 95% confidence interval of the difference and then specifying whether that confidence interval included zero. Model calibration was evaluated by calculating the mean absolute error between the actual and predicted outcome of survival. Calibration statistics were compared in the same fashion as were the HCI test statistics. Univariate and multivariate analyses were performed on potential determinants of survival using Cox-proportional hazards and then these determinants were used to build additional predictive survival models (all models were based on complete case analysis). Diagnostic checks of colinearity, proportional hazards assumption, and goodness of fit supported the use of all of our survival models. Tumor size and grade (as measured by Ki-67 immunolabeling throughout this paper) were used as continuous variables to evaluate for any natural cutoffs relating to survival. Survival based on tumor size and grade was assessed by evaluating the association between the log-hazard of death to tumor size and grade. The association between grade and other pathologic features of the tumors was evaluated via *t* tests. Statistical analysis was performed using R version 2.12.2.^{19–21} A *P* value of less than 0.05 was considered statistically significant and all tests were 2-sided.

RESULTS

Patient Characteristics

A total of 326 patients were surgically resected at our institution for nonsyndromic and nonfunctional PanNETs between the dates of October 29, 1984, and May 18, 2011. Descriptions of this population regarding demographics, tumor characteristics, stages, and clinical data can be found in Table 1. Median and mean follow-up were 43 months and 56 months, respectively (range from 0 to 272 months) with 91 (28%) being followed to death.

Validation of AJCC and ENETS Staging Systems

Patients were assigned both an AJCC and ENETS stage (Tables 1 and 2) and Kaplan-Meier curves were generated on the basis of both staging systems (Fig. 1). The 5-year OS rates for the AJCC stages Ia, Ib, IIa, IIb, and IV (our population did not have any stage III patients, meaning those with tumors that were unresectable because of invasion of the celiac axis or superior mesenteric artery) were 96% (91%–100%), 92% (84%–100%), 76% (62%–93%), 73% (63%–84%) and 56% (42%–73%), respectively. Median survival times were 363 (172–

764), 314 (166–592), 98 (65–149), 114 (83–157), and 62 (45–85) months, respectively. When using only the 4-stage AJCC model with stages I, II, and IV, the 5-year OS rates were 93% (88%–99%), 74% (65%–83%), and 56% (42%–73%) with median survival times of 333 (201–553), 108 (84–139), and 62 (45–85) months, respectively. Statistically significant differences in survival were not present between the lower substages but were present between the rest of the stages (see “AJCC Staging System, 2010” in Table 3). The HCI was 0.68 for the full staging system and 0.69 for the 4-stage system.

There were not enough events in the substages of the ENETS system to power an evaluation of the full staging system, so only the 4-stage ENETS model was evaluated. The 5-year OS rates for ENETS stages I, II, III, and IV were 97% (92%–100%), 87% (80%–95%), 73% (63%–84%), and 56% (42%–73%), respectively. Median survival times were 402 (162–995), 195 (135–281), 108 (80–147), and 62 (45–85) months, respectively. When comparing all 4 ENETS stages together, statistically significant differences in survival were not present between the lower stages but were present between the higher stages (see “ENETS Staging System, 2006” in Table 3). The HCI for the 4-stage system was 0.68.

A 95% confidence interval for the estimated difference between the 2 HCI statistics from the 4-stage AJCC model and the 4-stage ENETS model included zero, so one cannot say that there is a significant difference in their ability to predict survival.

Evaluation of Current Size and Grade Cutoffs

Meaningful cutoffs in tumor size and grade (as measured by Ki-67 index) relating to survival were sought to compare our potential cutoffs with the current cutoffs used in the AJCC and ENETS staging systems. Associations between tumor size and grade (both measured as continuous variables) to the log-hazard of death were graphed (figures not shown). For tumor size, there appeared to be a natural size cutoff at 3 cm above which there was a relatively constant hazard of death. On the basis of this finding, we hypothesize that the size cutoff of 2 cm for AJCC and ENETS is not necessarily meaningful (potentially contributing to the insignificant survival difference between the lower stages in both the AJCC and ENETS models) whereas the 4-cm cutoff for ENETS matches relatively well with our 3-cm cutoff.

For grade, there was a relatively steep and constant positive association with the log-hazard of death. We concluded that grade carries a large measure of survival prognostication and that it loses some of its predictive ability on survival when categorized, so we treated grade as a continuous variable rather than a categorical variable for the rest of our analyses.

Analysis of Prognostic Factors on Survival

Univariate analysis was performed on demographic and pathologic variables that a pathologist would have available during the examination of a surgical specimen (Table 3). On the basis of the univariate analysis findings, multivariate models were then constructed. Because of the high likelihood that including all significant covariates from the univariate analysis in the multivariate models would result in overfit models (ie, too many covariates in relation to the primary outcome of death leading to a modeling of “random noise” rather than the relationships of interest), reduced multivariate models were fit using just a few

covariates. While having a number of significant univariate analysis variables to choose from, it was found that grade could be a surrogate for many of them (Table 4) and so grade was included in the models along with the other significant variables of age, sex, and positive margins. These reduced models included the multivariate AJCC reduced model, the multivariate ENETS reduced model, and the multivariate nonstaged reduced model. The nonstaged model was created to explore whether “un-grouping” tumor size, lymph node status, and distant metastasis status (which are “grouped” in defining stages in the AJCC and ENETS models) lead to an increase or decrease in the model’s predictive ability on survival. The HCIs for these models are predictably larger than those for the isolated AJCC and ENETS models (0.68), because these 3 models contain more information that predicts survival. The reduced models’ HCIs of 0.74 to 0.76 are within the range of novel prognostic systems for pancreatic cancer (HCI = 0.61)²² and breast cancer (HCI = 0.80).²³

The 95% confidence interval for the estimated difference between the HCI statistics for the reduced AJCC and ENETS models included zero, so we cannot say that there is a significant difference in the AJCC and ENETS ability to predict survival in these multivariate models. However, the 95% confidence interval for the estimated differences between the HCI statistics for the reduced AJCC and ENETS models as compared with the nonstaged model did not include zero, so we can say that both the reduced AJCC and ENETS models have a slightly better predictive ability on survival as compared with the nonstaged model.

Grade as a Prognostic Indicator

When categorizing grade according to the WHO cutoffs of 0% to 3% (grade I), 3% to 20% (grade II), and more than 20% (grade III),¹⁴ it was seen that grade was predictive of survival (Fig. 1) with 5-year OS for grades 1 to 3 being 85% (78%–92%), 78% (70%–88%), and 9% (2%–52%), respectively. Median survival times were 182 (133–249), 108 (83–140), and 20 (13–31) months, respectively.

To test the hypothesis that grade carried a lot more prognostic power when used as a continuous variable and that it might even supplement or supplant the AJCC and ENETS staging systems as a prognostic indicator, associations were established between Ki-67 index and all of the other pathologic variables (Table 4). Of note, it can be seen that below a Ki-67 of 2%, it is unlikely that a patient will be positive for any of the items listed in Table 4, whereas the reverse is true for a Ki-67 over 3%. It was also found that the Ki-67 index is significantly associated with recurrence (and recurrence is related to death with an HR = 2.07, 95% confidence interval = 1.19–3.61, $P = 0.01$). To evaluate the association between tumor size and Ki-67 index, the natural log of the Ki-67 index was compared with the natural log of tumor size (percentage association = 0.157, 95% confidence interval = 0.10–0.22, $P = 0.001$; the percentage association is interpreted as a change in tumor size by a factor of X is associated with a change in Ki-67 by a factor of $X^{0.157}$).

To further evaluate the hypothesis that grade might be the only pathologic variable that is relevant in predicting survival, a survival model was created on the basis of continuous grade as the only pathologic variable while controlling for the significant demographic variables of age and sex (the “Grade Only” model) (Table 3). To explore the possibility that just controlling the AJCC and ENETS models for age and sex would make them better

prognostic models than our “Grade Only” model, 2 more models were created that were the “AJCC Staging System Corrected for Age and Sex” and the “ENETS Staging System Corrected for Age and Sex.” Of note, the HCI for the “Grade Only” model (0.74) trended higher than its age- and sex-controlled AJCC and ENETS comparators (0.72 and 0.70), although there was no statistical significance between the predictive ability of the 3 models. One final model was fit, which was a slight modification of the “Grade Only” model. Distant metastases were added to the “Grade Only” model because metastases are a relatively easy factor to assess in patients as well as a factor likely to confer survival prognosis according to the multivariate nonstaged reduced model (Table 3). However, the “Grade Only Plus Distant Metastases Model’s” HCI only increased to 0.75 (Table 3).

Model Choice, Recalibration, and Nomogram

To select one of the new models to compare with the 2010 AJCC and 2006 ENETS models, we looked for a model with a higher HCI and that did not include the AJCC and ENETS staging. This narrowed the choices down to the “Grade Only” and the “Grade Only Plus Distant Metastases” models. The 95% bootstrapped confidence intervals for the estimated differences between the HCI statistics for the 2010 AJCC and 2006 ENETS models as compared with the other 2 models all included zero, so we cannot say that there is a significant difference in all 4 models’ ability to predict survival.

Having explored the discrimination of these 4 models, we next turned to the second half of validation, namely calibration. The mean absolute errors (calibration measurements) for the 2010 AJCC, 2006 ENETS, “Grade Only,” and “Grade Only Plus Distant Metastases” models were 0.0142, 0.0172, 0.0496, and 0.0607, respectively. The 95% bootstrapped confidence intervals for the estimated differences between the 2010 AJCC and 2006 ENETS models as compared with the other 2 models showed that calibration was statistically similar for the “Grade Only” model but statistically worse for the “Grade Only Plus Distant Metastases” model. For this reason, we chose the “Grade Only” model as our final comparison model. However, even with a statistically insignificant difference in the calibration between the 2010 AJCC, 2006 ENETS, and “Grade Only” models, we recalibrated our “Grade Only” model by dichotomizing age at a number of different cutoff points. The best discrimination (as measured by HCI) and calibration were seen at the cutoff age of 63 (≤ 63 or >63), where the HCI was 0.74 and the mean absolute error was 0.0277. This HCI and calibration statistic were, again, not statistically different from the HCI and calibration statistics of the 2010 AJCC and 2006 ENETS models when looking at the 95% bootstrapped confidence intervals for their estimated differences so we used this binary age model as our final “Proposed” model. We did not explore any further divisions in age, beyond dichotomization, as most of our patient population was between the ages of 40 and 74 with only 10 patients older than 80 years and 4 patients older than 85 years. Although this age cutoff may have tended toward improved calibration, we note that a graph of the log-hazard of death versus continuous age does not show any natural break at the age of 63. Of note, when we recalibrated our “Proposed” model, “sex” became insignificant at the $\alpha = 0.05$ level as a predictor of survival, but we retained “sex” in our final “Proposed” model for its statistically insignificant enhancement of survival prognostication in both discrimination and calibration (a model that only includes age as ≤ 63 years or >63 years and Ki-67 has an

HCI of 0.72 and mean absolute error of 0.0396). Finally, recalibration efforts also included categorizing grade by our Ki-67 cutoffs of 0% to 2%, 2% to 3%, and greater than 3% on top of the binary age model, but this resulted in worse discrimination (HCI = 0.63) with only a somewhat improved mean absolute error of 0.020, so this model was not pursued.

At last, a nomogram was produced that provides 5-year OS and median survival times according to our “Proposed” model (Fig. 2).

DISCUSSION

With the largest surgical series of nonsyndromic, nonfunctional PanNETS from a single institution, we assessed the validity (discrimination and calibration) of survival prognostication of the 2 most common PanNET staging systems. We found that both the 4-stage AJCC and ENETS models were equally valid (predictive ability of 68% for both and mean absolute errors in calibration of 0.0142 and 0.0172, respectively). We also built and then explored other predictive survival models and found one that has statistically similar discrimination and calibration (although the predictive ability clearly trends higher at 74% with calibration that trends slightly lower at a mean absolute error of 0.0277) but has the advantage of being much more parsimonious than either the AJCC or ENETS models. We believe that it would be reasonable to externally validate our new predictive survival model based on grade, age, and sex to assess its clinical utility in becoming a standard for survival prognosis.

The use of Ki-67 immunolabeling as a marker for grade has been widely used as one of several predictive factors of survival although there are no consistent Ki-67 index category cutoffs across the literature.²⁴ We discovered in comparing Ki-67 index to survival that there was a linear relationship and thus no meaningful Ki-67 index cutoffs in making Ki-67 index categories. For this reason, our model is different from other models, which include Ki-67, in that we treat the Ki-67 index as a continuous variable rather than a categorical variable and thereby retain all of the predictive potential of the Ki-67 index. The fact that our model is based almost exclusively on grade makes it the most parsimonious (having just one predictive factor that has any appreciable measurement error) predictive survival model when compared with the AJCC and ENETS models. Parsimony and validity both are important characteristics of useful predictive models.²⁵ Parsimony is important not just because it has been found that more complex models can be overly optimistic in their predictions²⁵ but also because it can lead to more accurate and reliable communication within and between institutions that may, in turn, help standardize and streamline treatment. Specifically, the only measured variable in the “Proposed” model is grade, whereas the measured variables for the AJCC and ENETS models include tumor size, lymph node status (further complicated by potential questions regarding what is an adequate lymph node harvest), and distant metastases (potentially complicated by questions regarding the thoroughness of the metastatic workup and interpretation of diagnostic images or biopsies). The WHO already suggests recording grade as mitoses per 10 high-power fields or as percentage of Ki-67 staining,¹⁴ so adopting this “Proposed” model should not change current pathology practices drastically.

Although our “Proposed” model is a predictive model for survival, we also believe that the Ki-67 index serves as a basis for a simplistic classification system describing the tumors’ propensity to be associated with the characteristics listed in Table 4. We suggest that a Ki-67 index of 0% to 2% be labeled as “low-grade,” 2% to 3% as “intermediate-grade,” and greater than 3% as “high-grade.” The classification system we propose is related to survival in that low-grade, intermediate-grade, and high-grade tumors have 5-year OSs of 85% (77%–93%), 84% (70%–100%), and 69% (59%–79%) with median survival times of 197 (134–288), 176 (81–382), and 86 (68–109) months, respectively. There are 24 patients in our population classified as intermediate-grade, and the intermediate-grade survival curve is not statistically different from the low-grade or high-grade survival curve (log rank test, $P = 0.804$ and 0.081 , respectively), whereas the low-grade and high-grade survival curves are statistically different (log rank test, $P = 0.001$). Although we use these grade categories for a general classification, we do not use grade categories for survival prediction in our “Proposed” model, as it tends toward poorer discrimination.

Besides our low-, intermediate-, and high-grade classification system potentially leading to a more simple, uniform, and prognostically accurate communication about these tumors, using our “Proposed” model for survival prognostication may lead to a change in the management of PanNETs in a specific sense if used in a preoperative setting. In our population, there were 12 patients who were staged as AJCC stage Ia (and 27 staged as AJCC stage I) or ENETS stage I while having a “high-grade” Ki-67 index more than 3%. Those who were AJCC or ENETS stage I and were “high-grade” had 5-year OSs of 69% (75%–82%) and 77% (71%–83%), respectively, which are much closer to “high-grade” tumor survival rates (69%, 59%–79%) than AJCC (93%, 88%–99%) or ENETS (97%, 92%–100%) stage I survival rates. In a situation where tissue biopsy reveals a “high-grade tumor” but the patient is stage I according to the AJCC or ENETS staging systems, a practitioner might be more aggressive in recommending surgery for appropriate surgical candidates or adjuvant therapy than what might be offered to AJCC or ENETS stage I patients.

We recognize several limitations in our study. First, our “Proposed” model will need to be validated on other postoperative patient populations before it can be considered for standard use. We concluded apparent and internal (via bootstrapping) validation of our model but external validation with other populations from other institutions remains to be assessed as patient selection, diagnostic, and therapeutic regimens and Ki-67 index calculations can vary by institution. Validation of both our predictive survival model and classification system can be accomplished through discrimination and calibration tests using retrospective cohorts of surgical patients at other institutions.^{4,5,26}

Second, we acknowledge that our “Proposed” model has an HCI that trends higher and a calibration that trends lower than the AJCC and ENETS models (although there is no statistical difference between any of these statistics in the 3 models). Although validation includes assessment of both discrimination and calibration, there is no direct comparison to say one is more important than another. In fact, the clinical utility of a model may stand even in the face of “less than good” model performance in these areas depending on the context and clinical judgment of its use.²⁵ For this reason, we believe that it is still important

to externally validate our “Proposed” model to see if the trending higher discriminative ability improves prognostication even in the face of the calibration that trends lower.

A third limitation of our “Proposed” model is that it requires a diagnostic tissue sample that on some biopsies may not be large enough to do the correct staining and counting for the Ki-67 index. This prognostic model was based on postsurgical candidates; so the model would need to be validated on preoperative patients as well before using the model to determine surgical treatment based on tissue biopsy only. Also, staining and counting must be uniform between institutions (we detailed our process in the Methods) and concordance tests will likely need to be performed.

Finally, a perceived limitation might be that we did not divide our cohort according to some time periods; however, we believe that this is not absolutely necessary. This is because our large cohort was followed over 26 years and received a standard surgical approach, whereas there has been little change in neoadjuvant or adjuvant therapies over the past 2 decades that have had a significant impact on improving the outcomes of those with metastatic disease.²⁷

While we have found a manner to improve survival prognostication in our postsurgical patients, we recognize that the great interest and research in PanNETS will continue to provide further elucidation into prognostication and treatment. We also recognize that there are other exciting avenues for PanNET prognostication including genetic alteration evaluation, which may be developed in future staging systems.²⁸

Acknowledgments

No support has been taken from any source.

REFERENCES

1. Edge, SB.; Byrd, DR. *AJCC Cancer Staging Manual*. 7th ed.. Philadelphia, PA: Springer; 2010. Exocrine and endocrine pancreas; p. 241-249.
2. Martin RCG, Kooby DA, Weber SM, et al. Analysis of 6,747 pancreatic neuroendocrine tumors for a proposed staging system. *J Gastrointest Surg*. 2011; 15:175–183. [PubMed: 21103949]
3. Niederle MB, Hackl M, Daserer K, et al. Gastroenteropancreatic neuroendocrine tumours: the current incidence and staging based on the WHO and European Neuroendocrine Tumour Society classification: an analysis based on prospectively collected parameters. *Endocr Relat Cancer*. 2010; 17:909–918. [PubMed: 20702725]
4. 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*. 2011; 29:3044–3049. [PubMed: 21709192]
5. Goh BKP, Chow PKH, Tan YM, et al. Validation of five contemporary prognostication systems for primary pancreatic endocrine neoplasms: results from a single institution experience with 61 surgically treated cases. *ANZ J Surg*. 2011; 81:79–85. [PubMed: 21299804]
6. Hurtuk MG, Godambe AS, Shoup M, et al. Support for a postresection prognostic score for pancreatic endocrine tumors. *Am J Surg*. 2011; 201:406–410. [PubMed: 21367388]
7. Oberstein PE, Saif MW. Novel agents in the treatment of unresectable neuroendocrine tumors. Highlights from the “2011 ASCO Annual Meeting.” Chicago, IL, USA; June 3–7, 2011. *JOP*. 2011; 12:358–361. [PubMed: 21737896]
8. Kulke MH, Bendell J, Kvols L, et al. Evolving diagnostic and treatment strategies for pancreatic neuroendocrine tumors. *J Hematol Oncol*. 2011; 4:29. [PubMed: 21672194]

9. Scarpa A, Mantovani W, Capelli P, et al. Pancreatic endocrine tumors: improved TNM staging and histopathological grading permit a clinically efficient prognostic stratification of patients. *Mod Pathol.* 2010; 23:824–833. [PubMed: 20305616]
10. Saif MW. Pancreatic neoplasm in 2011: an update. *JOP.* 2011; 12:316–321. [PubMed: 21737886]
11. Williams ED, Sandler M. The classification of carcinoid tumors. *Lancet.* 1963; 1:238–239. [PubMed: 14000847]
12. Klöppel G. Classification and pathology of gastroenteropancreatic neuroendocrine neoplasms. *Endocr Relat Cancer.* 2011; 18:S1–S16. [PubMed: 22005112]
13. Capella C, Heitz PU, Höfler H, et al. Revised classification of neuroendocrine tumours of the lung, pancreas and gut. *Virchows Arch.* 1995; 425:547–560. [PubMed: 7697211]
14. Klimstra DS, Modlin IR, Coppola D, et al. The pathologic classification of neuroendocrine tumors: a review of nomenclature, grading, and staging systems. *Pancreas.* 2010; 39:707–712. [PubMed: 20664470]
15. Sellner F, Thalhammer S, Stattner S, et al. TNM stage and grade in predicting the prognosis of operated, non-functioning neuroendocrine carcinoma of the pancreas: a single-institution experience. *J Surg Oncol.* 2011; 104:17–21. [PubMed: 21360536]
16. Rindi G. The ENETS guidelines: the new TNM classification system. *Tumori.* 2010; 96:806–809. [PubMed: 21302636]
17. Klöppel G, Couvelard A, Perren A, et al. ENETS consensus guidelines for the standards of care in neuroendocrine tumors: towards a standardized approach to the diagnosis of gastroenteropancreatic neuroendocrine tumors and their prognostic stratification. *Neuroendocrinology.* 2009; 90:162–166. [PubMed: 19060454]
18. Liszka L, Pajak J, Mrowiec S, et al. Discrepancies between two alternative staging systems (European Neuroendocrine Tumor Society 2006 and American Joint Committee on Cancer/Union for International Cancer Control 2010) of neuroendocrine neoplasms of the pancreas. A study of 50 cases. *Pathol Res Pract.* 2011; 207:220–224. [PubMed: 21354717]
19. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; <http://www.R-project.org/>.
20. Therneau T. A package for survival analysis in S. R package version 2.37–4. <http://CRAN.R-project.org/package=survival>.
21. Harrell FE. Design package. R package version 2.3-0. <http://cran.r-project.org/src/contrib/Archive/Design>.
22. Dong X, Li Y, Hess KR, et al. DNA mismatch repair gene polymorphisms affect survival in pancreatic cancer. *Oncologist.* 2011; 16:61–70. [PubMed: 21212431]
23. Yi M, Mittendorf EA, Cormier JN, et al. Novel staging system for predicting disease-specific survival in patients with breast cancer treated with surgery as the first intervention: time to modify the current American Joint Committee on Cancer staging system. *J Clin Oncol.* 2011; 29:4654–4661. [PubMed: 22084362]
24. Jamali M, Chetty R. Predicting prognosis in gastroentero-pancreatic neuroendocrine tumors: an overview of the value of Ki-67 immunostaining. *Endocr Pathol.* 2008; 19:282–288. [PubMed: 18931958]
25. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. *BMJ.* 2009; 338:b605. [PubMed: 19477892]
26. Strosberg JR, Cheema A, Weber JM, et al. Relapse-free survival in patients with nonmetastatic, surgically resected pancreatic neuroendocrine tumors: an analysis of the AJCC and ENETS staging classifications. *Ann Surg.* 2012; 256:321–325. [PubMed: 22415420]
27. Oberstein PE, Saif MW. Update on novel therapies for pancreatic neuroendocrine tumors. *JOP.* 2012; 13:372–375. [PubMed: 22797392]
28. Jiao Y, Shi C, Edil BH, et al. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science.* 2011; 331:1199–1203. [PubMed: 21252315]

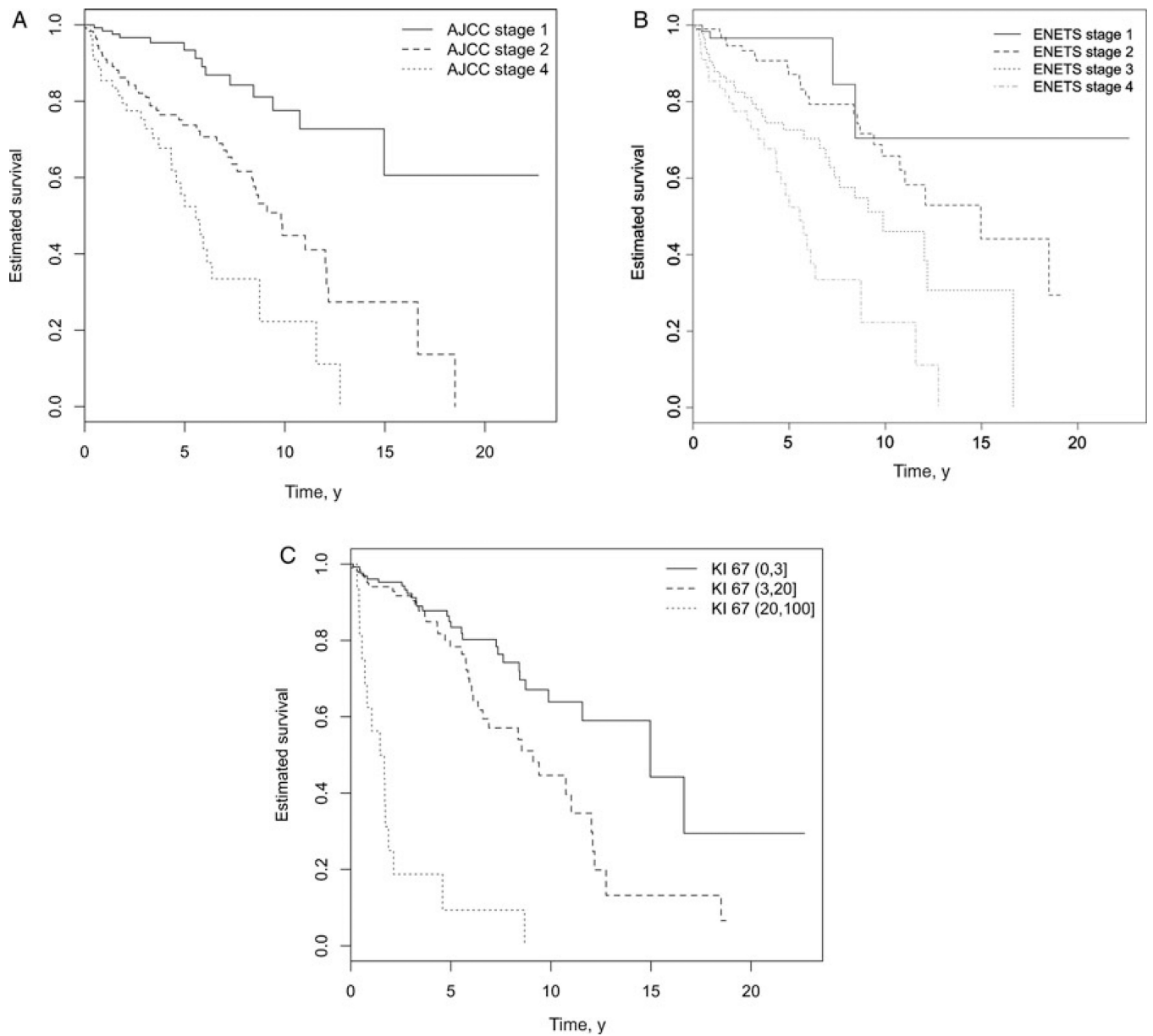


FIGURE 1. A, Kaplan-Meier Survival Curves for 2010 AJCC Staging System. B, Kaplan-Meier Curve for 2006 ENETS Staging System. C, Kaplan-Meier Curve for WHO Ki-67 Grading System.

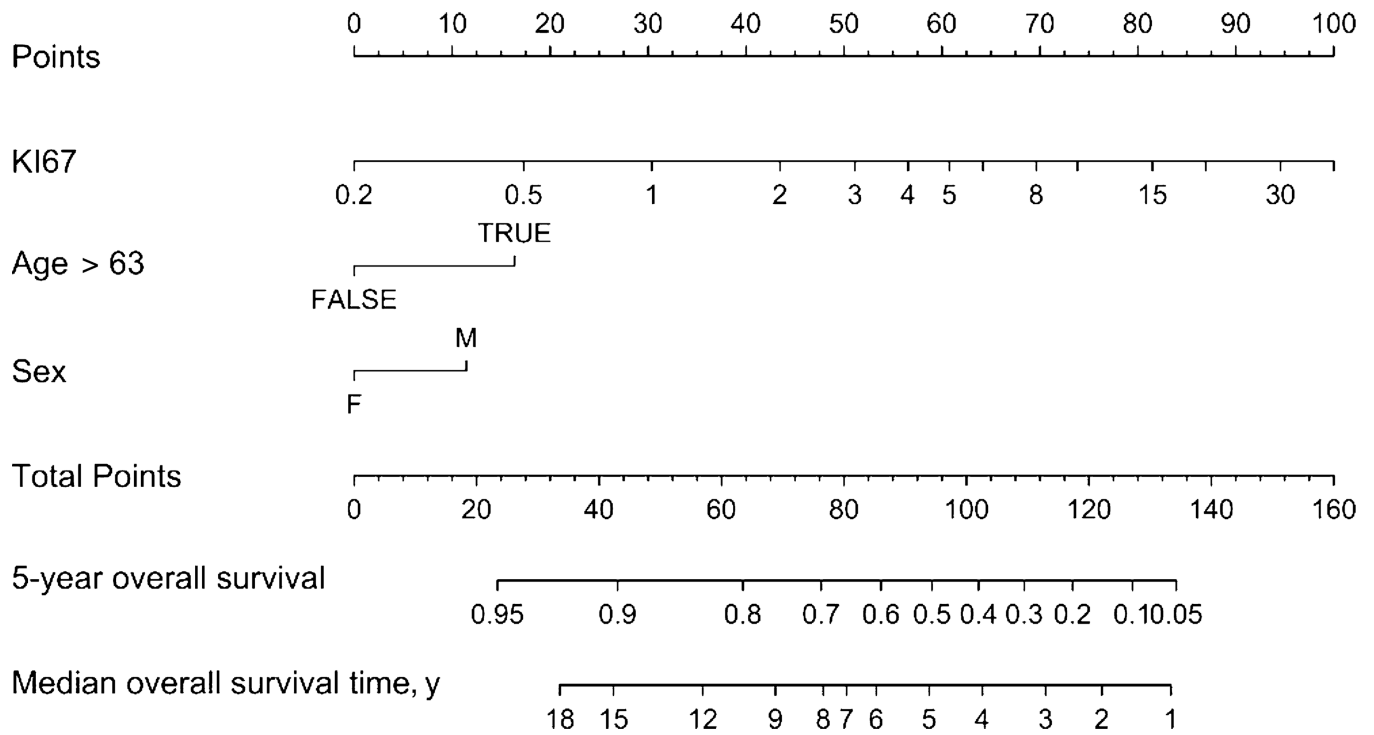


FIGURE 2.

Nomogram for predicted survival of those with nonsyndromic, nonfunctional PanNETs based on our proposed model (grade, age, and sex). *Sum together 'Points' from Ki-67, age at surgery (< 63, > 63) and sex collected from the first line and then find the corresponding result on the 'Total Points' line. From the Total Points' line, find the corresponding 5-year overall survival and median survival time. For example, a female patient ('Points' = 0) who is 75-years old at surgery ('Points' = 18) and has a PanNET surgical specimen with a Ki-67 index of 3% ('Points' = 50) has a 'Total Points' of 68 which corresponds to a 5-year overall survival of approximately 76% with a median overall survival time of 9 years.

TABLE 1

Patient Information

Patient Characteristic	Number	Percentage
Age	Median: 57	Range: 23–93
Sex		
Female	154	47%
Male	172	53%
Race		
White	286	88%
Other*	40	12%
Tumor grade (Ki-67 index) [†]		
Low (<3%)	150	46%
Intermediate (3%–20%)	108	33%
High (>20%)	18	5.5%
Tumors not Ki-67 labeled	50	15%
AJCC stage		
Ia	86	26%
Ib	61	19%
IIa	32	10%
IIb	87	27%
III	0	0%
IV	59	18%
ENETS		
I	74	23%
IIa	54	16%
IIb	48	15%
IIIa	3	1%
IIIb	87	27%
IV	59	18%
Patient developing a recurrence	65	20%
Patients undergoing reoperation	44	17%
Patients currently surviving	235	72%

* 22 Black, 2 Asian, 3 Hispanic ethnicity, 12 Other, and 1 Unknown.

[†] According to ENETS and WHO designations.¹⁶

TABLE 2

Cross Tabulation of AJCC/ENETS and T Stage/Overall Stage Discrepancies

Full-Stage Systems						
Staging	ENETS I	ENETS IIa	ENETS IIb	ENETS IIIa	ENETS IIIb	ENETS IV
AJCC Ia	74	12	0	0	0	0
AJCC Ib	0	41	20	0	0	0
AJCC IIa	0	1	28	3	0	0
AJCC IIb	0	0	0	0	87	0
AJCC IV	0	0	0	0	0	59

Four-Stage Systems				
Staging	ENETS I	ENETS II	ENETS III	ENETS IV
AJCC I	74	73	0	0
AJCC II	0	29	90	0
AJCC IV	0	0	0	59

AJCC/ENETS Discrepancies		
	N	%
T category discrepancies	88	27% (88/326)
T2/T3	58	66% (58/88)
T1/T2	15	17% (15/88)
T3/T4	13	15% (13/88)
T1/T3	2	2% (2/88)
ENETS giving the higher T category	55	62.5% (55/88)
Overall stage discrepancies	51	16% (51/326)
AJCC IIb vs ENETS IIIb	28	55% (28/51)
AJCC Ib vs ENETS IIb	20	39% (20/51)
AJCC IIa vs ENETS IIIa	3	6% (3/51)
ENETS giving higher stage	51	100% (51/51)

TABLE 3

Models of PanNET Survival

Independent Variables	Hazard Ratio	95% Confidence Interval	P
Univariate analysis			
1. Independent variables			
Size, cm			
(0,2] vs (2,4]	2.74	1.07–7.00	0.035*
(2,4] vs > 4	1.52	0.98–2.36	0.060
(0,2] vs > 4	4.17	1.65–10.53	0.003*
Tumor location [†]			
Head	1.34	0.88–2.05	0.170
Body	0.87	0.48–1.58	0.653
Tail	1.08	0.71–1.65	0.711
Uncinate process	0.63	0.09–4.52	0.644
Focality [†]			
Extent of spread [†]	2.84	1.82–4.44	<0.001*
Grade [†]			
Grade II (vs I)	1.91	1.17–3.12	0.010*
Grade III (vs I)	15.67	8.07–30.44	<0.001*
Large vessel invasion [†]			
Small vessel invasion [†]	3.14	2.02–4.90	<0.001*
Race [†]			
Positive margins	2.53	1.59–4.02	<0.001*
Positive lymph nodes	2.17	1.40–3.37	0.001*
Distant metastases	3.18	2.00–5.03	<0.001*
Necrosis present	2.65	1.72–4.09	<0.001*
Perineural invasion	1.99	1.30–3.04	0.002*
Age [‡]			
Sex (male)	1.68	1.09–2.58	0.019*
2. AJCC staging system, 2010 [§] (HCI = 0.68)			
AJCC Ib (vs Ia)	1.17	0.40–3.38	0.775
AJCC IIa (vs Ia)	4.42	1.73–11.25	0.002*
AJCC IIb (vs Ia)	3.91	1.61–9.50	0.003*
AJCC IV (vs Ia)	8.41	3.45–20.51	<0.001*
AJCC IIa (vs Ib)	3.78	1.63–8.78	0.002*
AJCC IIb (vs Ib)	3.35	1.53–7.36	0.003*
AJCC IV (vs Ib)	7.20	3.24–16.00	<0.001*
AJCC IIb (vs IIa)	0.89	0.48–1.62	0.696

Independent Variables	Hazard Ratio	95% Confidence Interval	P
AJCC IV (vs IIa)	1.91	1.03–3.53	0.041*
AJCC IV (vs IIb)	2.15	1.27–3.63	0.004*
3. ENETS staging system, 2006 [§] (HCI = 0.68)			
ENETS II (vs I)	2.14	0.74–6.19	0.161
ENETS III (vs I)	4.45	1.57–12.63	0.005*
ENETS IV (vs I)	9.04	3.16–25.84	<0.001*
ENETS III (vs II)	2.08	1.21–3.56	0.008*
ENETS IV (vs II)	4.22	2.40–7.44	<0.001*
ENETS IV (vs III)	2.03	1.22–3.39	0.007*
Multivariate analyses			
1. Multivariate AJCC reduced model [¶] (HCI = 0.76)			
AJCC II (vs I)	1.68	0.86–3.30	0.13
AJCC IV (vs I)	3.46	1.67–7.16	0.001*
AJCC IV (vs II)	2.06	1.22–3.46	0.007*
Log grade [¶]	1.83	1.47–2.29	<0.001*
Positive margins	1.28	0.76–2.17	0.354
Sex (male)	1.65	1.03–2.63	0.037*
Age [†]	1.03	1.02–1.05	<0.001*
2. Multivariate ENETS reduced model ^{**} (HCI = 0.76)			
ENETS II (vs I)	1.10	0.36–3.34	0.867
ENETS III (vs I)	1.82	0.62–5.30	0.274
ENETS IV (vs I)	3.20	1.07–9.58	0.037*
ENETS III (vs II)	1.65	0.91–3.00	0.097
ENETS IV (vs II)	2.91	1.59–5.32	0.001*
ENETS IV (vs III)	1.76	0.99–3.12	0.053
Log grade [*]	1.94	1.55–2.42	<0.001*
Positive margins	1.33	0.80–2.23	0.273
Sex (male)	1.76	1.10–2.79	0.017*
Age [†]	1.03	1.01–1.05	0.001*
3. Multivariate nonstaged reduced model [†] (HCI = 0.74)			
Size, cm			
(0,2] vs (2,4]	1.14	0.42–3.09	0.795
(0,2] vs >4	1.21	0.43–3.39	0.723
Positive lymph nodes	1.23	0.71–2.13	0.453
Distant metastasis	1.95	1.06–3.58	0.031*
Log grade [¶]	1.88	1.50–2.36	<0.001*
Positive margins	1.28	0.73–2.23	0.396
Sex (male)	1.85	1.12–3.03	0.015*

Independent Variables	Hazard Ratio	95% Confidence Interval	P
Age [†]	1.03	1.01–1.05	<0.001*
4. “Grade Only” model [‡] (HCI = 0.74)			
Log grade [∥]	2.10	1.71–2.57	<0.001*
Sex (male)	1.61	1.02–2.55	0.041*
Age [†]	1.03	1.01–1.05	<0.001*
5. AJCC staging system corrected for age and sex [§] (HCI = 0.72)			
AJCC II (vs I)	3.51	1.91–6.45	<0.001*
AJCC IV (vs I)	8.47	4.39–16.36	<0.001*
AJCC IV (vs II)	2.42	1.47–3.96	<0.001*
Sex (male)	1.51	0.97–2.36	0.067
Age [†]	1.03	1.01–1.04	0.001*
6. ENETS staging system corrected for age and sex [§] (HCI = 0.70)			
ENETS II (vs I)	2.26	0.78–6.56	0.132
ENETS III (vs I)	4.12	1.45–11.75	0.008*
ENETS IV (vs I)	10.22	3.56–29.33	<0.001*
ENETS III (vs II)	1.82	1.06–3.12	0.029*
ENETS IV (vs II)	4.51	2.55–7.98	<0.001*
ENETS IV (vs III)	2.48	1.46–4.21	0.001*
Sex (male)	1.66	1.07–2.57	0.024*
Age [†]	1.02	1.01–1.04	0.004*
7. “Grade Only” plus distant metastases model [‡] (HCI = 0.75)			
Log grade [∥]	1.97	1.59–2.44	<0.001*
Sex (male)	1.73	1.09–2.76	0.020*
Age [†]	1.03	1.01–1.05	<0.001*
Distant metastases	2.32	1.40–3.84	0.001*
8. Proposed model [‡] (HCI = 0.74)			
Log grade [∥]	2.00	1.65–2.42	<0.001*
Sex (male)	1.52	0.96–2.42	0.074
Age > 63	1.82	1.13–2.95	0.014*

* Statistically significant at $\alpha < 0.05$.

[†]Tumor location: head, body, tail, and uncinate process (each category compared to the rest); focality: uni- vs multifocal; extent of spread: contained to the pancreas or extending to peripancreatic soft tissue/adjacent organs; grade: categorical Ki-67 index of (0%–3%], (3, 20%] and >20%; large vessel invasion: celiac axis or superior mesenteric artery invasion; small vessel invasion: any vessel invasion other than large vessel invasion; race: white vs nonwhite.

[‡]Age is compared with the age of the previous year so that a change in age by X years would be associated with a change in the hazard of death by a factor of $(1.02)^X$.

[∥]If $\ln(\text{hazard ratio}) = 0.51$, a change in Ki-67% by a factor of X will be associated with a change in the hazard of death by a factor of $X^{0.51}$.

[§] Complete case analysis, n = 320; number of events = 89; *P* value from log-rank test of equality of survival curves < 0.001.

[¶] Complete case analysis, n = 271; number of events = 81; *P* value from likelihood ratio test of stage terms = 0.002.

^{**} Complete case analysis, n = 271; number of events = 81; *P* value from likelihood ratio test of stage terms = 0.003.

^{††} Complete case analysis, n = 256; number of events = 77.

^{‡‡} Complete case analysis, n = 271; number of events = 81.

^{§§} Complete case analysis, n = 320; number of events = 89; *P* value from likelihood ratio test of stage terms < 0.001.

TABLE 4

Grade Associations With Other Prognostic Factors

	Median Ki-67 for Negative Association (n)	Median Ki-67 for Positive Association (n)
Positive lymph nodes	1.70 (182)	3.59 (135)
Distant metastases	1.87 (281)	6.97 (62)
Extending outside of pancreas	1.57 (215)	4.00 (128)
Necrosis	1.76 (247)	6.02 (65)
Perineural invasion	1.80 (237)	3.86 (105)
Small vessel invasion	1.58 (226)	3.83 (112)
Large vessel invasion	1.90 (302)	3.71 (34)
Positive margins	2.00 (298)	3.29 (45)
Recurrence	1.56 (212)	3.34 (63)

n = number of observations used in each association calculation. T tests of the difference in average log(Ki-67) between the 2 averages were all associated with a $P < 0.013$. T tests were performed without assuming equal variances between groups and sample sizes were never less than 34.