

Developing a multivariable prognostic model for pancreatic endocrine tumors using the clinical data warehouse resources of a single institution

Taxiarchis Botsis^{1,2}, Valsamo K. Anagnostou³, Gunnar Hartvigsen², George Hripcsak¹, Chunhua Weng¹

¹Department of Biomedical Informatics, Columbia University, New York, NY, USA, ²Department of Computer Science, University of Tromsø, Norway, ³Department of Pathology, Yale University School of Medicine, New Haven, CT, USA

Keywords

Data repositories, Data mining, Pancreatic endocrine tumors, Prognosis, Electronic health records

Summary

Objective: Current staging systems are not accurate for classifying pancreatic endocrine tumors (PETs) by risk. Here, we developed a prognostic model for PETs and compared it to the WHO classification system.

Methods: We identified 98 patients diagnosed with PET at NewYork-Presbyterian Hospital/Columbia University Medical Center (1999 to 2009). Tumor and clinical characteristics were retrieved and associations with survival were assessed by univariate Cox analysis. A multivariable model was constructed and a risk score was calculated; the prognostic strength of our model was assessed with the concordance index.

Results: Our cohort had median age of 60 years and consisted of 61.2% women; median follow-up time was 10.4 months (range: 0.1-99.6) with a 5-year survival of 61.5%. The majority of PETs were non-functional and no difference was observed between functional and non-functional tumors with respect to WHO stage, age, pathologic characteristics or survival. Distant metastases, aspartate aminotransferase-AST and surgical resection (HR=3.39, 95% CI: 1.38-8.35, p=0.008, HR=3.73, 95% CI: 1.20-11.57, p=0.023 and HR=0.20, 95% CI: 0.08-0.51, p<0.001 respectively) were the strongest predictors in the univariate analysis. Age, perineural and/or lymphovascular invasion, distant metastases and AST were the independent prognostic factors in the final multivariable model; a risk score was calculated and classified patients into low (n=40), intermediate (n=48) and high risk (n=10) groups. The concordance index of our model was 0.93 compared to 0.72 for the WHO system.

Conclusion: Our prognostic model was highly accurate in stratifying patients by risk; novel approaches as such could thus be incorporated into clinical decisions.

Correspondence to:

Taxiarchis Botsis, Ph.D.
Department of Computer Science,
University of Tromsø
9037 Tromsø, Norway
Tel: + 47 77 64 40 35
Fax: + 47 77 64 45 80
E-mail: Taxiarchis.Botsis@uit.no

Appl Clin Inf 2010; 1: 38–49

doi: 10.4338/ACI-2009-12-RA-0026

received: December 23, 2009

accepted: March 10, 2010

published: March 31, 2010

Citation: Botsis T, Anagnostou VK, Hartvigsen G, Hripcsak G, Weng C. Developing a multivariable prognostic model for pancreatic endocrine tumors using the clinical data warehouse resources of a single institution. *Appl Clin Inf* 2010; 1: 38–49.
<http://dx.doi.org/10.4338/ACI-2009-12-RA-0026>

Introduction

Pancreatic endocrine tumors (PET) are rare neoplasms with an annual incidence of less than 1 per 100,000 [1], however autopsy studies have shown frequencies ranging from 0.8% to 10% in asymptomatic patients undergoing a post mortem examination [2, 3]. PETs account for 1-3% of pancreatic neoplasms and form a heterogeneous group of functional and non-functional tumors varying in aggressiveness; 5-year survival rates range from 41 to 71% [1, 4]. Advanced stage, higher grade, and age have been reported as the strongest predictors of survival and patients with functional tumors typically have a favourable prognosis most possibly due to diagnosis at an earlier stage [1, 4-6].

The current WHO classification system includes pathologic (mitotic index, angioinvasion, perineural invasion and proliferation index) and stage-related (tumor size, presence of any metastasis) predictors [7], however its clinical application may be limited by complex classification criteria and absence of prognostic stratification within a given group. Multiple studies have attempted to develop classification systems to better define prognosis [8-10] yielding variable results; this is mainly attributed to small sample size, methodology differences and lack of robust multivariable statistical analysis. Here we developed a multivariable prognostic model for endocrine pancreatic tumors and compared its prognostic accuracy to existing staging systems.

Methods

Cohort description and data extraction

Cases of PETs diagnosed between January 1999 and January 2009 were retrospectively retrieved from NewYork-Presbyterian Hospital/Columbia University Medical Center (NYP/CUMC) clinical data warehouse (CDW). This warehouse has been in operation since 1994 and has accumulated health data for more than 2.7 million of patients. Since 2002, a comprehensive controlled clinical vocabulary called the Medical Entities Dictionary (MED; <http://med.dmi.columbia.edu/>) has been used to integrate data of various semantic representations from heterogeneous hospital information systems for the CDW. Using the CDW resources we followed a multi-step procedure to identify the cases of PETs and extract the data of interest.

We used the 9th version of International Classification of Diseases, Clinical Modification (ICD-9-CM) and its codes corresponding to the “malignant neoplasm of pancreas” (157.0-157.9) to identify all the patients with ICD-9 diagnoses during the selected period. The pathology reports, radiology reports, clinical notes, laboratory tests, discharge summaries, as well as the drug registry and administrative files were also extracted for case validation. Subsequently, an SQL query including a combination of clinical terms (endocrine or neuroendocrine pancreatic tumor and/or carcinoma, pancreatic islet tumor, gastrinoma, glucagonoma, insulinoma, VIPoma, somatostatinoma, functional and non-functional tumor) was applied on the pathology reports for case identification and yielded 115 records over 3,068 cases of pancreatic cancer; the extracted records were manually reviewed to confirm their appropriateness for this study.

Patient demographics (age at diagnosis, race, gender), tumor characteristics [functional status, localization, size, necrosis, differentiation, mitotic index (low mitotic index with 0-1 mitoses per 50 high power fields vs. intermediate mitotic index with 2-50 mitoses per 50 high power fields vs. high mitotic index with ≥ 50 mitoses per 50 high power fields), perineural and/or lymphovascular invasion, gross local invasion, presence of lymph node and distant metastases], personal (history of other cancer, chronic pancreatitis, cholelithiasis) and family cancer history (cancer in first degree relatives, multiple endocrine neoplasia syndrome 1-MEN 1, von Hippel Lindau-VHL syndrome), laboratory tests at diagnosis (aspartate aminotransferase-AST, alanine aminotransferase-ALT, alkaline phosphatase-ALP, albumin, total bilirubin, CA19-9, carcinoembryonic antigen-CEA) and type of surgical resection (enucleation, partial, central and distal- pancreatectomy, Whipple resection, pancreatoduodenectomy, pancreatosplenectomy) were manually abstracted and automatically extracted from the aforementioned CDW elements; the cohort characteristics are shown in Table I. Tumors arising in the duodenum or the peri-ampullary region (n=3) and outside consultations not

treated at NYP/CUMC (n=6) were excluded. Benign insulinomas (n=8) were also excluded from the study because their remarkably longer survival compared to all other PETs [11] has been reported to skew the outcome in series where patients with insulinomas are grouped with patients having other functional or non-functional tumors [1]. Patients were classified according to the World Health Organization (WHO) classification system for PETs [7] and 17 (17.3%) stage 1.1, 29 (29.6%) stage 1.2, 25 (25.5%) stage 2 and 9 (9.2%) stage 3 tumors were identified; stage was missing for 18 (18.4%) patients (►Table 1).

Statistical analysis

Differences between functional status and other clinical and pathological characteristics factors were investigated using the chi-square and Kruskal-Wallis statistic for categorical and the t-test statistic for continuous variables. Disease-specific survival was defined as the time from diagnosis (evidenced in the pathology reports) to either death caused by disease or last follow-up. Univariate associations between demographics, clinical characteristics, laboratory values, surgical treatment and survival were assessed by Cox proportional hazards regression analysis; all variables that were significant at the 0.10 level were further analyzed in a multivariable Cox proportional hazards model. Age at diagnosis and size were analyzed as binary variables split by the median value and the 2cm cut off point indicated by the WHO classification respectively. Laboratory values were considered either as normal (within reference range) or elevated (2.5 times the upper limit for AST, ALT, ALP and total bilirubin; below lower limit for albumin; and above reference value for CA19-9 and CEA). Clinically related variables were examined for interactions prior to further analysis and the data set including the candidate variables identified in univariate analysis was imputed by applying the Multivariate Imputation by Chained Equations (MICE) method assuming that data were missing at random (MAR) [12]. Subsequently, 1000 bootstrap samples were generated based on the imputed set and a backward elimination multivariable Cox proportional hazards model was developed for each bootstrap sample. The Akaike Information Criterion (AIC) was used as the criterion for selection of the best prognostic model for a level of significance of 0.05 [13]. The regression coefficients from the multivariable model were divided with the smallest coefficient in order to calculate a score (equal to the quotient of the division) for each variable, which was then weighted by its coefficient, with zero points assigned to the reference category. Subsequently, the scores were summed up into a raw prognostic score and patients were stratified into low, intermediate and high risk groups using tertiles as cut off points for risk classification. Survival curves for the risk groups were constructed using the Kaplan-Meier method and survival differences were analyzed by the log rank test. The reduced multivariable prognostic model was compared to the WHO staging system using the concordance index with 95% confidence intervals [14]. P values were based on two-sided testing and differences were considered significant at $p < 0.05$. All statistical analyses were done in R-statistics software (version 2.9.0); the Kaplan-Meier curves were constructed in SPSS (version 15.0 for Windows, Chicago, IL).

Institutional Review/Approval

This study was approved by the NYP/CUMC Institutional Review Board (Study No.: #AAAD7480) and was conducted according to the ethical guidelines mandated by the Declaration of Helsinki.

Results

Patient characteristics

Ninety-eight patients diagnosed with PET were identified; our cohort consisted of 38 men (38.8%) and 60 women (61.2%) with a median age of 61.5 (range: 30-88, mean \pm SE: 61.4 \pm 2.0) and 59.5 (32-88, mean \pm SE: 58.3 \pm 1.8) respectively. Seventy four (75.5%) patients were white, 8 (8.2%) black and 9 (9.2%) and 6 (6.1%) were of Hispanic and Asian origin respectively; race for one patient had not been recorded. The majority of PETs were non-functional (n=79, 80.6%), as for functional tumors there

were 8 (8.2%) malignant insulinomas, 5 (5.1%) gastrinomas, 4 (4.1%) glucagonomas, 1 (1%) somatostatinoma and 1 (1%) mixed somatostatinoma-glucagonoma. One non-functional patient was diagnosed with multiple endocrine neoplasia type 1 (MEN-1) and two patients with non-functional PETs had history of VHL disease. One patient with VHL disease was also diagnosed with concurrent renal cell carcinoma. Eighty three (84.7%) patients underwent surgical treatment whereas 15 (15.3%) did not; treatment of the primary tumor involved enucleation for 1 (1.0%), pancreatectomy for 21 (21.4%), pancreatosplenectomy for 30 (30.6%), pancreatoduodenectomy for 2 (2.0%) and Whipple resection for 29 (29.6%) patients. The median length of follow up for all patients was 10.4 months (range: 0.1-99.6, mean±SE: 22.6±2.6) with a 5-year disease specific survival of 61.5%.

Presence of nodal metastases was identified in 23 (23.5%) cases and 18 (18.4%) and 6 (6.1%) patients presented with distant and both nodal and distant metastases at the time of diagnosis respectively; all distant metastatic lesions were localized in the liver. There was a trend towards worse prognosis for patients with distant metastases compared to patients with nodal metastases (log rank $p=0.06$). There was no difference between patients with functional and non-functional tumors with respect to stage ($p=0.896$), age (median age at diagnosis 61.4 years, range: 46-86, mean±SE: 63.8±2.5 and 60 years, range: 30-88, mean±SE: 58.4±1.6 respectively, $p=0.102$), tumor size ($p=0.470$) and localization ($p=0.436$), mitotic index ($p=0.290$), necrosis ($p=0.465$) or survival (log rank $p=0.87$).

Identification of predictors of survival

Perineural and/or lymphovascular invasion (HR: 8.20, 95% CI: 1.06-63.56, $p=0.044$), tumor size (HR: 4.55, 95% CI: 1.05-19.80, $p=0.044$), distant metastases (HR: 3.39, 95% CI: 1.38-8.35, $p=0.008$) and elevated AST (HR: 3.73, 95% CI: 1.20-11.57, $p=0.023$; Table II) levels were the only variables associated with survival with distant metastases and AST being the strongest predictors. Elevated AST levels did not correlate either with the presence of liver metastasis ($p=0.506$) or the WHO stages ($p=0.728$). The association of chronic pancreatitis with survival was borderline significant (HR: 3.33, 95% CI: 0.99-11.13, $p=0.051$; ▶ Table 2) and resection of any type was strongly associated with survival (HR: 0.20, 95% CI: 0.08-0.51, $p<0.001$; ▶ Table 2). The WHO classification system was associated with survival such that patients with 1.0 and 2.0 disease had a reduced risk compared to stage 3.0 patients (HR=0.18, 95% CI: 0.05-0.66, $p=0.010$ and HR=0.25, 95% CI: 0.07-0.86, $p=0.027$ respectively; ▶ Table 2), however survival rates were similar among patients with 1.1 and 1.2 disease (log rank $p=0.098$).

Development of a multivariable prognostic model

Variables significantly correlated with survival in univariate analysis at a level of significance 0.10 were further incorporated in a multivariate Cox proportional hazards regression analysis model using a stepwise selection/backward elimination process on each of the 1000 bootstrap samples. Age at diagnosis (HR=4.75, 95% CI: 1.58-13.25, $p=0.005$), perineural and/or lymphovascular invasion (HR=8.62, 95% CI: 1.10-67.42, $p=0.040$), distant metastases (HR=2.94, 95% CI: 1.07-8.07, $p=0.036$) and AST (HR=3.53, 95% CI: 1.03-12.03, $p=0.044$) were the independent prognostic factors that were included in the final model (▶ Table 3). Based on the factor coefficients we developed a scoring system assigning 1 point to presence of distant metastases, 1.25 points to elevated AST, 1.5 points to age at diagnosis over 60 and 2 points to perineural and/or lymphovascular invasion. Patients were classified into three groups using tertiles as cut off points: scores ranged from 0 – 1.5, 2 – 3.5 and >3.5 for the low ($n=40$), intermediate ($n=48$) and high risk ($n=10$) groups respectively. Patients in the low risk group had an exceptionally favourable prognosis compared to the intermediate and high risk groups (log rank $p<0.010$) and log rank $p<0.001$, respectively; ▶ Figure 1A. Interestingly 3 and 19 patients with 1.1 and 1.2 WHO disease were classified in the intermediate risk group respectively whereas 3 stage 1.2 patients were classified in the high risk group, underlying the weakness of the WHO system to accurately stratify patients by risk (survival curves for WHO stages are shown in ▶ Figure 1B). The concordance index was calculated to assess the accuracy of our prognostic model and was equal to 0.93 (95% CI: 0.53-0.99), whereas the concordance index of the WHO system was equal to 0.72 (95% CI: 0.34-0.93).

Discussion

The current WHO classification system seems to have reached a plateau in efficacy for predicting survival for patients with pancreatic endocrine tumors; tumors characterized as well-differentiated may recur whereas others with an “uncertain” behaviour may not [15]. Several classification systems have been developed incorporating staging and tumor characteristics with various reproducibility among cohorts [9, 10, 15]. We developed an accurate multivariable model for stratifying PETs by risk and compared its prognostic accuracy with the WHO staging system. Our prognostic model classified PETs with higher accuracy compared to the WHO system; concordance indexes were equal to 0.93 and 0.72 respectively.

The WHO classification system does not distinguish between nodal and distant metastases and gross invasion [7]; when those parameters were grouped together in our study their association with survival did not reach significance in univariate analysis (HR=2.22, 95% CI: 0.84-5.86, $p=0.106$). Interestingly enough when distant metastases were analyzed individually, a significant correlation with survival was observed for both univariate (HR=3.39, 95% CI: 1.38-8.35, $p=0.008$) and multivariable analysis (HR=2.94, 95% CI: 1.07-8.07, $p=0.036$). These findings are consistent with previous studies highlighting the prognostic importance of liver metastases either for functional or non-functional PETs [10, 16]. The impact of nodal metastases on survival remains controversial with several studies underlying their prognostic value [17, 18] whereas others demonstrate that there is no association with survival [19-21]; in our cohort nodal status did not predict survival (HR=1.28, 95% CI: 0.41-3.94, $p=0.670$). Interestingly, the mitotic index was not correlated with survival in our analysis (HR=2.74, 95% CI: 0.89-8.43, $p=0.079$).

No difference was observed between patients with WHO 1.1 and WHO 1.2 disease in our cohort and this is consistent with previous reports [8, 15, 22]; based on that we decided to treat stage 1 patients as one group. This has also been the choice for other researchers as well as one of the cases where WHO classification appeared to have a distinct outcome [23]. The impact of functional status on survival is controversial with several studies demonstrating a significant correlation [5, 6] and others not [16, 17, 24]. In our cohort there was no survival benefit for patients with functional tumors therefore our model could be universally applied to PETs independent of their functional status.

We found that elevated AST is an independent prognostic factor for PETs and this has not been previously reported to the best of our knowledge. AST levels were not associated with liver metastases and a subgroup analysis for PETs with no liver metastasis ($n=79$), showed that elevated AST retained its prognostic value (HR: 6.06, 95% CI: 1.08-34.14, $p=0.041$); these findings are similar to the observations of Clancy et al who reported that elevated alkaline phosphatase levels were predictive of survival in patients with preserved liver function [25]. Our observation that inclusion of biomarkers may increase the accuracy of prognostic classification for PETs is consistent with the findings of both Schmitt et al and Ali et al who showed that CK19 staining may improve the prognostic power of WHO classification system [26, 27]. Our final prognostic model also included age at diagnosis, which has been shown to be a strong predictor of survival with patients over 60 years having the worst prognosis among all [4, 6]. Furthermore, perineural and/or lymphovascular invasion that is a well documented predictor of survival [7] was also included in our final model.

We compared the prognostic strength of our model to the TNM staging system proposed recently by the European Neuroendocrine Tumor Society [9] and the grading system developed by Hochwald et al [8]. The TNM staging system classifies patients in four stages (I, IIa and IIb, IIIa and IIIb and IV) by evaluating primary tumor characteristics, regional lymphnodes and distant metastases whereas the Hochwald system classifies tumors as low grade (no necrosis and <2 mitoses per 50 high power fields) and intermediate grade (necrosis or ≥ 2 mitoses per 50 high power fields). Applying the criteria of those systems to the patients of our cohort (imputed data set) we found that the concordance index was 0.60 (95% CI: 0.21-0.89) and 0.59 (95% CI: 0.21-0.89) for Hochwald (low versus intermediate grade) and TNM (I & II versus III & IV stages) system respectively; corresponding survival curves are shown in ►Figure 2. Both systems suffer from inherent limitations and lack of independent validation; more specifically the TNM early stages (I & II) have marginal or no

survival differences and only stage IV has a clearly distinct outcome [19, 23, 28] whereas the Hochwald approach seems to oversimplify the classification.

Potential limitations of the current study are the retrospective nature of data collection, the short follow-up and the relatively small sample size of our cohort. PETs account for 1-3% of all pancreatic neoplasms and the small sample size reflects the low incidence of these neoplasms (115 PETs over 3,068 pancreatic neoplasms in our study). Validation of our preliminary findings on prospectively collected cohorts is required to prove the robustness of our approach.

We employed the MICE imputation method that is indicated for handling of categorical variables and allows for selection of specific variables acting as predictors in the imputation process [12]. Data imputation has been proven to be superior to the complete case analysis and the missing-indicator method in multivariable diagnostic research [29] and has been suggested as the ideal approach to address missingness in retrospective analyses [30-32].

The set that was used in this study contained a huge amount of longitudinal data over a 10-year period. In this context, it could be argued that the possible changes over time, such as changes of ICD-9-CM codes or changes of institutional operations, might have affected the consistency of data presentations and have introduced a considerable bias in our study. It should be mentioned though that the NYP/CUMC's data warehouse team has been working on these issues to assure the quality of the stored data. Also, the Medical Informatics Dictionary was designed to resolve semantic equivalence of different data sources and support the transitions through multiple versions of clinical terminologies. We are confident that data integrity is ensured with the help of MED regardless of the evolution of clinical terminologies.

From the applied clinical informatics point of view, this study demonstrated the feasibility and value of the secondary use of electronic health record (EHR) for answering clinical research questions. Issues like data incompleteness, which has been observed in our study, can be resolved by applying specific methods, such as imputation techniques. A specific mining strategy is needed to resolve other problems related to EHR data quality, e.g. inaccuracy and inconsistency; for example, manual data abstraction and automated data extraction were combined in our work. Generally, all retrospective studies that are based on the processing of clinical information for predictive analysis, should examine and handle all the data quality issues through the appropriate use of dedicated tools and methods.

Conclusion

We developed an accurate and easily-applied prognostic stratification system for PETs, incorporating age at diagnosis, presence of distant metastases, perineural and lymphovascular invasion and AST levels. Our multivariable model if validated in prospectively collected cohorts may be useful for a more accurate prognostic stratification of patients with pancreatic endocrine tumors and could be incorporated into clinical decision-making.

This work demonstrates the potential of clinical systems to provide researchers with accurate data using a longitudinal clinical data warehouse. Also, certain data weaknesses (e.g. missingness) may be corrected using the appropriate methods (e.g. imputation of missing values). Subsequently, this may result in high quality models that could be incorporated into the clinical decision-making.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

We thank Alla Babina for retrieving patient data from the NYP/CUMC Clinical Data Warehouse and Dr. Dimitris Rizopoulos for his clarifications regarding 'bootStepAIC' package in R-statistics. This study was partly funded by the Research Council of Norway (Project No: 174934) and was also supported by NIH LM009886-01A1 and NLM R01 LM006910 US grants.

Table 1 Cohort characteristics

(*low: 0-1 mitoses per 50 high power fields; intermediate: 2-50 mitoses per 50 high power fields; and high: ≥ 50 mitoses per 50 high power fields)

Variable	Variable Categories (<i>Number of patients</i>)				Missing
Age At Diagnosis	≥ 60 years (49)	>60 years (49)			0
Race	White (74)	Black (8)	Hispanic (9)	Asian (6)	1
Gender	Female (60)	Male (38)			0
Functional Status	Functional (19)	Non-functional (79)			0
Localization	Head/Neck (49)	Body/Tail (47)			2
Necrosis	No (55)	Yes (23)			20
Mitotic Index*	No/Low (51)	Intermediate/High (26)			21
Differentiation	Well-differentiated (53)	Poorly-differentiated (9)			36
Lymph node metastasis	No (46)	Yes (23)			29
Perineural and/or lymphovascular invasion	No (41)	Yes (42)			15
Metastasis and/or Invasion to adjacent organs	No (59)	Yes (39)			0
Distant metastasis	No (79)	Yes (18)			1
Size	≤ 2 cm (32)	> 2 cm (60)			6
WHO classification system	Stage 1.1 (17)	Stage 1.2 (29)	Stage 2 (25)	Stage 3 (9)	18
Chronic Pancreatitis	No (55)	Yes (29)			14
History of Other Cancer	No (40)	Yes (24)			34
Cholelithiasis	No (67)	Yes (15)			16
Cancer in first Degree Relatives	No cancer (31)	Cancer (29)			38
AST	Normal (81)	Elevated (11)			6
ALT	Normal (72)	Elevated (16)			10
Alkaline Phosphatase	Normal (84)	Elevated (8)			6
Albumin	Normal (34)	Elevated (58)			6
Total bilirubin	Normal (87)	Elevated (5)			6
CA 19-9	Normal (46)	Elevated (7)			45
CEA	Normal (45)	Elevated (9)			44
Surgical Resection	No (15)	Any (83)			0

Table 2 Results of the univariate Cox proportional hazards regression analysis

Variable	HR (95% CI)	p-value
Age At Diagnosis		0.065
>60 years vs. ≤60 years	2.47 (0.94-6.48)	
Race		0.693
White	1.00 (Reference)	
Black	0.36 (0.05-2.77)	0.329
Hispanic	0.55 (0.13-2.44)	0.435
Asian	0.79 (0.10-6.02)	0.82
Gender		0.5
Male vs. Female	0.73 (0.30-1.81)	
Functional Status		0.824
Non-functional vs. Functional	1.13 (0.38-3.42)	
Localization		0.131
Body/Tail vs. Head/Neck	2.13 (0.80-5.69)	
Necrosis		0.335
Yes vs. No	1.86 (0.53-6.52)	
Mitotic Index		0.079
Intermediate/High vs. No/Low	2.74 (0.89-8.43)	
Differentiation		0.111
Poorly-differentiated vs. Well-differentiated	2.51 (0.81-7.77)	
Lymph node metastasis		0.67
Yes vs. No	1.28 (0.41-3.94)	
Perineural and/or lymphovascular invasion		0.044
Yes vs. No	8.20 (1.06-63.56)	
Metastasis and/or Invasion to adjacent organs		0.106
Yes vs. No	2.22 (0.84-5.86)	
Distant metastasis		0.008
Yes vs. No	3.39 (1.38-8.35)	
Tumor Size		0.044
>2 cm vs. ≤2 cm	4.55 (1.05-19.8)	
WHO classification system		0.018
Stage 3	1.00 (Reference)	
Stage 2	0.25 (0.07-0.86)	0.027
Stage 1	0.18 (0.05-0.66)	0.01
Chronic Pancreatitis		0.051
Yes vs. No	3.33 (0.99-11.13)	

Variable	HR (95% CI)	p-value
History of Other Cancer		0.908
Yes vs. No	1.07 (0.32-3.57)	
Cholelithiasis		0.799
Yes vs. No	0.82 (0.18-3.81)	
Cancer in first Degree Relatives		0.687
Cancer vs. No cancer	0.80 (0.27-2.39)	
AST		0.023
Elevated vs. Normal	3.73 (1.20-11.57)	
ALT		0.582
Elevated vs. Normal	1.55 (0.33-7.39)	
Alkaline Phosphatase		0.208
Elevated vs. Normal	2.07 (0.671-6.42)	
Albumin		0.399
Elevated vs. Normal	0.67 (0.27-1.68)	
Total bilirubin		0.72
Elevated vs. Normal	1.32 (0.29-5.93)	
CA 19-9		0.598
Elevated vs. Normal	1.81 (0.20-16.24)	
CEA		0.48
Elevated vs. Normal	2.27 (0.23-22.23)	
Surgical Resection		<0.001
Any vs. No	0.20 (0.08-0.51)	

Table 3 Multivariable analysis results of Cox proportional hazards model

Prognostic model	Variable	Coefficient	HR (95% CI)	p-value
<i>(AIC= 126.3288)</i>	Age At Diagnosis	1.52	4.75 (1.58-13.25)	0.005
	Perineural and/or lymphovascular invasion	2.154	8.62 (1.10-67.42)	0.04
	Distant metastasis	1.078	2.94 (1.07-8.07)	0.036
	AST	1.26	3.53 (1.03-12.03)	0.044

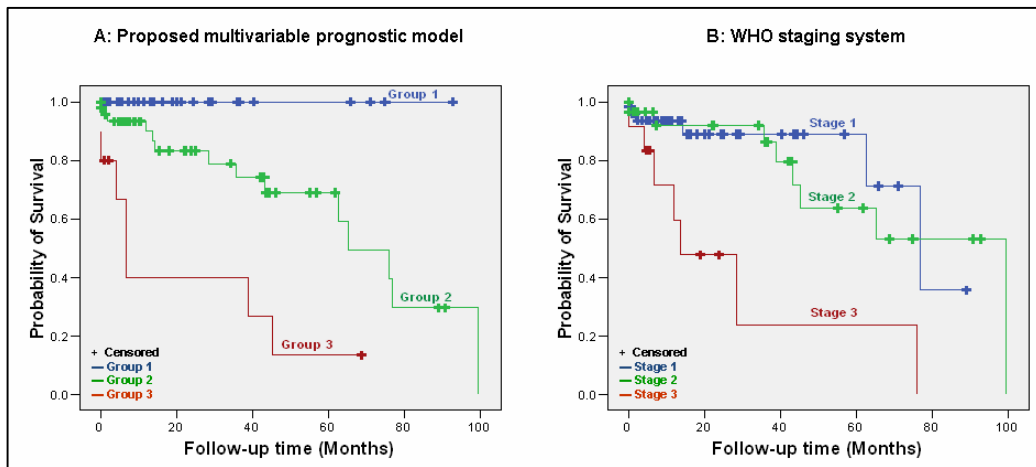


Figure 1 Disease outcome by the (A) proposed multivariable model with low risk patients showing an exceptional benefit towards survival compared to intermediate (log rank $p < 0.001$) and high risk patients (log rank $p < 0.001$); intermediate vs high risk patients also have better outcome (log rank $p = 0.002$) and (B) WHO staging system does not classify patients accurately.

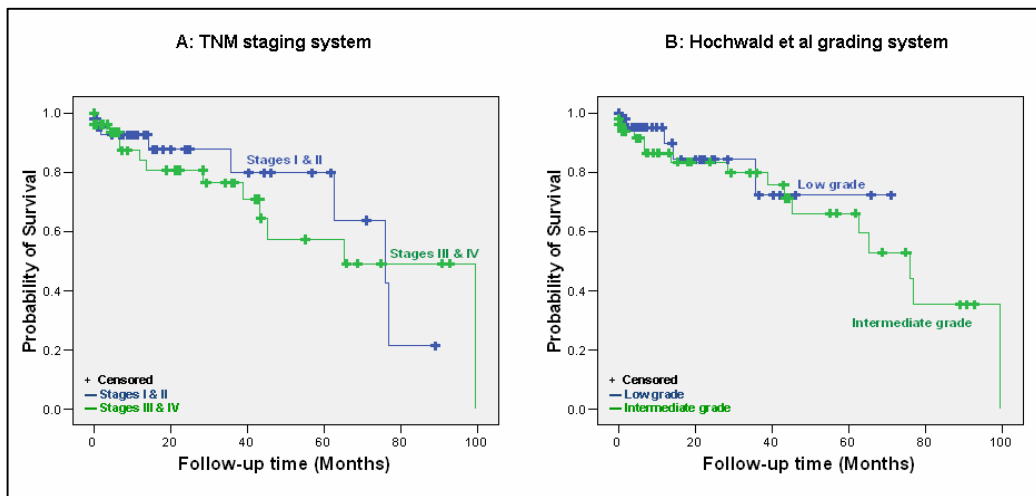


Figure 2 Disease outcome by (A) TNM staging with early stages not showing any survival benefit towards late stages (log rank $p = 0.679$) and (B) Hochwald et al grading system with no significant survival difference (log rank $p = 0.561$) between the patients of the two groups (low grade vs. intermediate grade).

References

1. Halfdanarson TR, Rubin J, Farnell MB, Grant CS, Petersen GM. Pancreatic endocrine neoplasms: epidemiology and prognosis of pancreatic endocrine tumors. *Endocr Relat Cancer* 2008;15:409.
2. Kimura W, Kuroda A, Morioka Y. Clinical pathology of endocrine tumors of the pancreas. Analysis of autopsy cases. *Dig Dis Sci* 1991;36:933.
3. Grimelius L, Hultquist GT, Stenkvist B. Cytological differentiation of asymptomatic pancreatic islet cell tumours in autopsy material. *Virchows Arch A Pathol Anat Histol* 1975;365:275.
4. Modlin IM, Sandor A. An analysis of 8305 cases of carcinoid tumors. *Cancer* 1997;79:813.
5. Phan GQ, Yeo CJ, Hruban RH, Litemoe KD, Pitt HA, Cameron JL. Surgical experience with pancreatic and peripancreatic neuroendocrine tumors: Review of 125 patients. *J Gastrointest Surg* 1998;2:473.
6. Halfdanarson TR, Rabe KG, Rubin J, Petersen GM. Pancreatic neuroendocrine tumors (PNETs): incidence, prognosis and recent trend toward improved survival. *Ann Oncol* 2008;19:1727.
7. Kloppel G, Perren A, Heitz PU. The gastroenteropancreatic neuroendocrine cell system and its tumors: the WHO classification. *Ann N Y Acad Sci* 2004;1014:13.
8. Hochwald SN, Zee S, Conlon KC, Colleoni R, Louie O, Brennan MF, Klimstra DS. Prognostic factors in pancreatic endocrine neoplasms: an analysis of 136 cases with a proposal for low-grade and intermediate-grade groups. *J Clin Oncol* 2002;20:2633.
9. Rindi G, Kloppel G, Alhman H, Caplin M, Couvelard A, de Herder WW, Eriksson B, Falchetti A, Falconi M, Komminoth P, Korner M, Lopes JM, McNicol AM, Nilsson O, Perren A, Scarpa A, Scoazec JY, Wiedenmann B. TNM staging of foregut (neuro)endocrine tumors: a consensus proposal including a grading system. *Virchows Arch* 2006;449:395.
10. Bilimoria KY, Talamonti MS, Tomlinson JS, Stewart AK, Winchester DP, Ko CY, Bentrem DJ. Prognostic score predicting survival after resection of pancreatic neuroendocrine tumors: analysis of 3851 patients. *Ann Surg* 2008;247:490.
11. Service FJ, McMahon MM, O'Brien PC, Ballard DJ. Functioning insulinoma--incidence, recurrence, and long-term survival of patients: a 60-year study. *Mayo Clin Proc* 1991;66:711.
12. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;18:681.
13. Austin PC, Tu JV. Bootstrap methods for developing predictive models. *The American Statistician* 2004;52:131.
14. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982;247:2543.
15. Ferrone CR, Tang LH, Tomlinson J, Gonen M, Hochwald SN, Brennan MF, Klimstra DS, Allen PJ. Determining prognosis in patients with pancreatic endocrine neoplasms: can the WHO classification system be simplified? *J Clin Oncol* 2007;25:5609.
16. Chu QD, Hill HC, Douglass HO, Jr., Driscoll D, Smith JL, Nava HR, Gibbs JF. Predictive factors associated with long-term survival in patients with neuroendocrine tumors of the pancreas. *Ann Surg Oncol* 2002;9:855.
17. Tomassetti P, Campana D, Piscitelli L, Casadei R, Santini D, Nori F, Morselli-Labate AM, Pezzilli R, Corinaldesi R. Endocrine pancreatic tumors: factors correlated with survival. *Ann Oncol* 2005;16:1806.
18. Tsuchiya A, Koizumi M, Ohtani H. World Health Organization Classification (2004)-based re-evaluation of 95 nonfunctioning "malignant" pancreatic endocrine tumors reported in Japan. *Surg Today* 2009;39:500.
19. Fischer L, Kleeff J, Esposito I, Hinz U, Zimmermann A, Friess H, Buchler MW. Clinical outcome and long-term survival in 118 consecutive patients with neuroendocrine tumours of the pancreas. *Br J Surg* 2008;95:627.
20. Chung JC, Choi DW, Jo SH, Heo JS, Choi SH, Kim YI. Malignant nonfunctioning endocrine tumors of the pancreas: predictive factors for survival after surgical treatment. *World J Surg* 2007;31:579.
21. Kazanjian KK, Reber HA, Hines OJ. Resection of pancreatic neuroendocrine tumors: results of 70 cases. *Arch Surg* 2006;141:765.
22. Schindl M, Kaczirek K, Kaserer K, Niederle B. Is the new classification of neuroendocrine pancreatic tumors of clinical help? *World J Surg* 2000;24:1312.
23. Ekeblad S, Skogseid B, Dunder K, Oberg K, Eriksson B. Prognostic factors and survival in 324 patients with pancreatic endocrine tumor treated at a single institution. *Clin Cancer Res* 2008;14:7798.
24. Vagefi PA, Razo O, Deshpande V, McGrath DJ, Lauwers GY, Thayer SP, Warshaw AL, Fernandez-Del Castillo C. Evolving patterns in the detection and outcomes of pancreatic neuroendocrine neoplasms: the Massachusetts General Hospital experience from 1977 to 2005. *Arch Surg* 2007;142:347.

25. Clancy TE, Sengupta TP, Paulus J, Ahmed F, Duh MS, Kulke MH. Alkaline phosphatase predicts survival in patients with metastatic neuroendocrine tumors. *Dig Dis Sci* 2006;51:877.
26. Ali A, Serra S, Asa SL, Chetty R. The predictive value of CK19 and CD99 in pancreatic endocrine tumors. *Am J Surg Pathol* 2006;30:1588.
27. Schmitt AM, Anlauf M, Rousson V, Schmid S, Kofler A, Riniker F, Bauersfeld J, Barghorn A, Probst-Hensch NM, Moch H, Heitz PU, Kloepfel G, Komminoth P, Perren A. WHO 2004 criteria and CK19 are reliable prognostic markers in pancreatic endocrine tumors. *Am J Surg Pathol* 2007;31:1677.
28. La Rosa S, Klersy C, Uccella S, Dainese L, Albarello L, Sonzogni A, Doglioni C, Capella C, Solcia E. Improved histologic and clinicopathologic criteria for prognostic evaluation of pancreatic endocrine tumors. *Hum Pathol* 2009;40:30.
29. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol* 2006;59:1102.
30. Arnold AM, Kronmal RA. Multiple imputation of baseline data in the cardiovascular health study. *Am J Epidemiol* 2003;157:74.
31. Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *J Clin Epidemiol* 2003;56:28.
32. Clark TG, Stewart ME, Altman DG, Gabra H, Smyth JF. A prognostic model for ovarian cancer. *Br J Cancer* 2001;85:944.