Exploring and validating the clinical risk factors for pancreatic cancer in chronic pancreatitis patients using electronic medical records datasets: three cohorts comprising 2,960 patients

Xin Zhao[1], Ren Lang[1], Zhigang Zhang[2], Weiling Zhao[3], Zhiwei Ji[3], Hua Tan[3], Xiaobo Zhou[3]

[1]Department of Hepatobiliary Surgery, Beijing Chaoyang Hospital Affiliated with Capital Medical University, Beijing 100020, China; [2]School of Information Management and Statistics, Hubei University of Economics, Wuhan 430205, China; [3]Center for Computational Systems Medicine at School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA

Contributions: (I) Conception and design: X Zhao; (II) Administrative support: X Zhou; (III) Provision of study materials or patients: W Zhao, H Tan; (IV) Collection and assembly of data: Z Zhang; (V) Data analysis and interpretation: X Zhao, Z Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xiaobo Zhou. Center for Computational Systems Medicine at School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA. Email: Xiaobo.Zhou@uth.tmc.edu.

Background: Patients with chronic pancreatitis (CP) have an increased risk of developing pancreatic cancer (PC). The purpose of this study was to identify predictors of PC in CP patients.

Methods: Electronic medical records (EMRs) of CP patients from two cohorts were collected, and a logistic regression analysis was performed to investigate the risk factors for PC. Subsequently, we validated the value of the risk prediction model with the EMRs of a third cohort.

Results: The derivation cohort consisted of 2,545 CP patients, and among them, 14 patients developed PC 7 years after CP diagnosis. Cyst of the pancreas [COP; odds ratio (OR): 4.37, 95% confidence interval (CI): 1.11 to 18.40, P=0.033], loss of weight (LW; OR: 3.21, 95% CI: 0.76 to 12.91, P=0.096) and high platelet (PLT) count (OR: 1.01 per 1 increment, 95% CI: 1.00 to 1.01, P=0.042) were independent risk factors for PC among CP patients. A risk prediction equation was constructed as follows: $\ln[p/(1-p)] = -6.68 + 1.55\text{COP} + 1.23\text{LW} + 0.0046\text{PLT}$. The areas under the receiver operating characteristic (ROC) curve of our risk score were 0.83 and 0.72 in the derivation and validation cohorts, respectively. A score >0.0128 and >0.0122 had the best balance between sensitivity and specificity in the derivation and validation cohorts, respectively.

Conclusions: In CP patients, LW, COP and high PLT count were identified as novel predictors of PC. A risk prediction model based on these factors exhibited moderate predictive value for CP patients.

Keywords: Pancreatic cancer (PC); chronic pancreatitis (CP); electronic medical records (EMRs); risk factor

Introduction

Pancreatic cancer (PC) is the fourth most fatal malignant tumor, and due to the rising obesity rate, aging population and other unknown factors, the mortality associated with PC has increased in men and has maintained stable in women (1,2). Although some therapeutic strategies were developed in the past decades, the overall 5-year survival rate of PC is less than 5% (3). Due to a lack of early specific symptoms, most patients are diagnosed in the advanced stage and cannot be treated with surgery. Therefore, identifying the patients in the early stage of tumorigenesis contributes to improving the dismal outcome of PC patients.

Chronic and progressive inflammation and fibrillation of the pancreas, termed chronic pancreatitis (CP), may lead to the damage of exocrine and endocrine function of
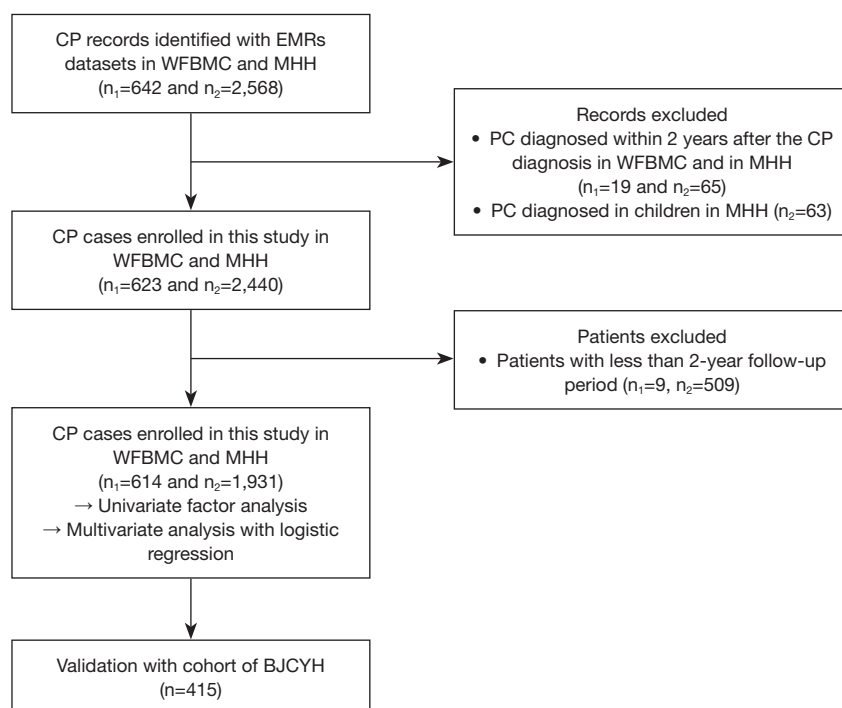
630

Zhao et al. Exploring risk factors for PC in CP patients

Figure 1 Flowchart of patient enrolment and the study design. ₁, WFUBMC; ₂, MHH. CP, chronic pancreatitis; EMRs, electronic medical records; WFUBMC, Wake Forest University Baptist Medical Center; MHH, Memorial Hermann Hospital; PC, pancreatic cancer; BJCYH, Beijing Chaoyang Hospital.

the pancreas (4). Compared with the normal population, the incidence of PC among CP patients was several- to dozens-fold higher if the patients also had diabetes (5). In epidemiological studies, CP and PC have been shown to share several typical risk factors, such as cigarette smoking and alcohol consumption (6,7). Theoretically, prolonged chronic inflammation was one of the primary carcinogenetic factors; thus, CP has been regarded as a risk factor for PC (8,9). However, very little research has focused on the predictors of PC development among CP patients. Recently, electronic medical records (EMRs) have provided a method to explore the alterations of clinical factors in a long-term period; thus, this dataset made it possible to analyze the associations between specific diseases and clinical characteristics (10). The purpose of the current study was to identify the risk factors for PC in CP patients based on EMRs.

Methods

Study design and patient population

In the derivation datasets, two EMRs cohorts from Wake

Forest University Baptist Medical Center (WFUBMC) and Memorial Hermann Hospital (MHH) were utilized. Univariate and multivariate analyses were performed to identify the predictors of PC in CP patients, and a logistic regression equation was constructed based on the two datasets. Then, we validated the predictive value of the prediction model with the cohort from the Beijing Chaoyang Hospital (BJCYH) (*Figure 1*).

The data source from WFUBMC was the i2b2 framework-based treatment delivery workflow (TDW), and we were supported by the study of "Linking Electronic Medical Records to Medicare Claims Data" (IRB00030804) for the access to these data. During the process of data acquisition, due to a limitation from the institutional review board (IRB), we accessed the data of only patients more than 60 years of age in WFUBMC. The EMRs dataset of MHH was organized by the PostgreSQL Server database 11.0.6251. The clinical data of BJCYH were collected from the computer-based patient record systems, and this study was permitted by the research ethics committee (2019-243).

The following clinical data were extracted and stratified in detail: demographic data including age, sex, race etc.;

history of alcohol consumption and smoking; family history of malignancy; accompanying disease including hypertension, type II diabetes mellitus (DM), coronary heart disease etc.; symptoms such as abdominal pain, diarrhea, loss of weight (LW) etc.; laboratory findings including routine blood examinations; and serum biochemical indexes. These data were examined and presented in *Table 1*.

Patients diagnosed with CP were enrolled in the study. The follow-up period was from the first date of CP diagnosis to the last date included in the database or the date of death. The exclusion criteria were as follows: groove pancreatitis, autoimmune pancreatitis, cases of PC diagnosed within 2 years after the diagnosis of CP and a follow-up period less than 2 years (6). Since the onset of CP in children was strongly associated with gene mutations (11,12), we excluded patients ≤19 years old from the data of MHH.

Definitions

On the basis of the *International Classification of Disease* codes (9th and 10th revision), we reviewed the patients with CP (577.1, 577.8, K86.0 and K86.1) and PC (157.0, 157.1, 157.2, 157.3, 157.4, 157.8, 157.9, C25.0, C25.1, C25.2, C25.3, C25.4, C25.7, C25.8, C25.9) based on the EMRs data of the above three cohorts (http://www.icd9data.com/ and http://www.icd10data.com). The diagnosis of accompanying diseases and symptoms were identified as follows: hypertension (401.9 and I10), coronary heart disease (414.01 and I25.10), hyperlipidemia (272.4 and E78.5), obstruction of bile duct (576.2), cyst of the pancreas (COP; 577.2 and K86.2), acute pancreatitis (AP) (577.0), type II DM (250.0 and E11.9), long-term or current use of insulin (Z79.4 and V58.67), hypothyroidism (244.9), abdominal pain (789.0), diarrhea (787.91), and LW (783.21).

Statistical analysis

The continuous variables are summarized as the mean (standard derivation) and were compared with an unpaired, 1-tailed *t*-test. The categorical variables are described using proportions and were compared with the chi-square test. P<0.05 was set as the level of statistical significance. According to the results of univariate analysis, the significant variables were assessed using a multivariable logistic regression model to identify the independent risk factors with entry and stay criteria of 0.05 and 0.1, respectively. All analyses were performed using R (version 3.3.1).

Results

General characteristics of datasets

From January 13, 2000 to October 12, 2015, a cohort of 642 CP patients was entered into the WFBMC cohort. There were 19 patients who were excluded because they were diagnosed with PC within 2 years after CP confirmation, and 9 patients were excluded due to a follow-up period of less than 2 years. Thus, 614 patients were finally enrolled in our study. The average follow-up duration was 8 [2–18] years, and 6 cases (0.98%) were diagnosed with PC after CP. Between July 19, 2000 and August 24, 2015, 2,568 patients with CP were recorded in the MHH database. Sixty-three children with CP were excluded. Among the other 65 patients who were excluded, PC was confirmed either at the same time or less than 2 years after CP diagnosis. After the exclusion of 509 patients with a follow-up period of less than 2 years, 1,931 patients were enrolled in this cohort. The median follow-up period in the MHH dataset was 7 years, ranging from 3 to 12 years, and PC was diagnosed in 8 patients (0.414%). A total of 2,545 patients with CP entered the derivation cohorts. Similarly, 415 CP patients entered into the validation cohort between February 2010 and October 2015, and among them, 7 patients developed PC. The clinical characteristics were compared between the derivation and validation cohorts, showing that there were differences in history of alcoholism, hypertension, LW, glucose level, total bilirubin, white blood cell count, albumin, platelet (PLT) count, and alanine transaminase (*Table 2*).

Predictors of PC development in CP patients

The clinical and laboratory parameters of the derivation datasets were shown in *Table 1*. There were no significant differences between the PC and CP groups with respect to epidemiological distribution, such as age, sex, race, and body mass index (BMI). The prevalence of smoking and alcohol abuse were also comparable between the two groups. Patients who developed PC were more likely to have hyperlipidemia (57.1% *vs.* 31.9%, P=0.04), COP (57.1% *vs.* 16.9%, P<0.01), and DM (78.6% *vs.* 38.5%, P<0.01). There were no differences between the two groups in regard to other accompanying diseases, such as AP, long-term or current use of insulin, hypothyroidism, hypertension, and atherosclerotic heart disease. The incidence of symptomatic diagnosis, including diarrhea (57.1% *vs.* 22.0%, P<0.01) and LW (35.7% *vs.* 11.5%,

Table 1 Overall clinical characteristics of 2,545 patients with CP in MHH and WFUBMC

Clinical parameters	PC group (n=14)	CP group (n=2,531)	P value
Demographics			
Age at diagnosis of CP, mean (± SD) years	59 (±16)	56 (±18)	0.26
Male sex, n (%)	11 (78.6)	1,657 (65.5)	0.39
Family history of malignancy[#], n (%)	1 (7.1)	59 (2.3)	0.22
African-American race, n (%)	7 (50.0)	737 (29.1)	0.13
BMI, mean (± SD)	26.4 (±6.0)	28.3 (±10.5)	0.30
Personal history, n (%)			
Smoking	5 (35.7)	408 (16.1)	0.06
Alcoholism	0	36 (1.4)	1.00
Past history, n (%)			
Hypertension	11 (78.6)	1,400 (55.3)	0.11
Coronary heart disease	3 (21.4)	242 (9.6)	0.14
Hyperlipidemia	8 (57.1)	808 (31.9)	0.04
Obstruction of bile duct	2 (14.3)	176 (7.0)	0.24
Diarrhea	8 (57.1)	558 (22.0)	<0.01
LW	5 (35.7)	290 (11.5)	0.02
Abdominal pain	11 (78.6)	1,665 (65.8)	0.39
COP	8 (57.1)	427 (16.9)	<0.01
Acute pancreatitis	12 (85.7)	1,534 (60.6)	0.08
DM	11 (78.6)	974 (38.5)	<0.01
Long-term or current use of insulin	5 (35.7)	252 (10.0)	<0.01
Hypothyroidism	3 (21.4)	241 (9.5)	0.14
Laboratory examinations, mean (± SD)			
Glucose, mg/dL	166.83 (±87.66)	130.11 (±68.05)	<0.01
White blood cell count, ×10^9/L	7.87 (±4.49)	8.57 (±4.93)	0.37
Total cholesterol[#]	168 (±77.92)	184.39 (±50.99)	0.39
Creatinine[#], mg/dL	0.96 (±0.51)	1.24 (±0.63)	0.34
Alkaline phosphatase, units/L	108.32 (±60.00)	106.54 (±69.65)	0.49
Calcium, mg/dL	8.93 (±4.47)	8.99 (±4.28)	0.40
Total bilirubin, mg/dL	0.73 (±0.43)	0.85 (±0.55)	0.33
Albumin, g/dL	3.67 (±1.87)	3.61 (±1.75)	0.42
Hemoglobin, g/dL	11.87 (±6.18)	12.04 (±5.45)	0.39
PLT count, ×10^9/L	289.10 (±173.96)	232.78 (±120.60)	0.03
Aspartate transaminase, units/L	34.26 (±22.75)	33.87 (±24.21)	0.49
Alanine transaminase, units/L	31.85 (±21.55)	30.04 (±24.64)	0.45
Amylase, units/mL	71.94 (±45.72)	230.81 (±486.97)	0.37

[#], The data were only from MHH. PC, pancreatic cancer; CP, chronic pancreatitis; MHH, Memorial Hermann Hospital; WFUBMC, Wake Forest University Baptist Medical Center; SD, standard deviation; BMI, body mass index; LW, loss of weight; COP, cyst of the pancreas; DM, diabetes mellitus; PLT, platelet.

Table 2 Comparison of the clinical data between the derivation and validation cohorts

Clinical parameters	Derivation cohort (n=2,545)	Validation cohort (n=415)	P value
PC, n (%)	14 (0.55)	7 (1.68)	0.02
Demographics			
Age at diagnosis of CP, mean (± SD) years	56 (±18)	57 (±16)	0.34
Male sex, n (%)	1,668 (65.5)	284 (68.4)	0.28
Personal history (no/yes), n (%)			
Smoking	413 (16.2)	71 (17.1)	0.44
Alcoholism	36 (1.4)	19 (4.6)	<0.01
Past history, n (%)			
Hypertension	1,411 (55.4)	162 (39.0)	<0.01
Coronary heart disease	245 (9.6)	37 (8.9)	0.69
Hyperlipidemia	816 (32.1)	151 (36.4)	0.08
Obstruction of bile duct	178 (7.0)	37 (8.9)	0.18
Diarrhea	566 (22.2)	88 (21.2)	0.66
LW	295 (11.6)	84 (20.2)	<0.01
Abdominal pain	1,676 (65.9)	253 (61.0)	0.07
COP	435 (17.1)	74 (17.8)	0.73
Acute pancreatitis	1,546 (60.7)	233 (56.1)	0.08
DM	985 (38.7)	155 (37.3)	0.63
Long-term or current use of insulin	257 (10.1)	43 (10.4)	0.93
Laboratory examinations, mean (± SD)			
Glucose, mg/dL	130.52 (±48.53)	141.75 (±100.93)	<0.01
White blood cell count, ×10^9/L	8.56 (±5.76)	7.93 (±4.46)	0.04
Alkaline phosphatase, units/L	106.29 (±83.36)	106.56 (±95.24)	0.96
Total bilirubin, mg/dL	0.85 (±0.74)	1.07 (±0.77)	<0.01
Albumin, g/dL	3.61 (±0.77)	3.94 (±3.10)	0.04
Hemoglobin, g/dL	12.04 (±1.78)	12.29 (±4.00)	0.22
PLT count, ×10^9/L	233.38 (±90.12)	183.87 (±75.19)	<0.01
Aspartate transaminase, units/L	33.88 (±46.00)	32.79 (±22.80)	0.61
Alanine transaminase, units/L	30.07 (±41.94)	48.90 (±84.30)	<0.01
Amylase, units/mL	229.26 (±1,064.99)	307.72 (±603.35)	0.19

PC, pancreatic cancer; CP, chronic pancreatitis; SD, standard deviation; LW, loss of weight; COP, cyst of the pancreas; DM, diabetes mellitus; PLT, platelet.

P<0.01), was higher in the PC group than in the CP group. Based on laboratory examinations, the PC group also had a higher glucose level (166.83 *vs.* 130.11 mg/dL, P<0.01) and PLT count (289.10 ×10^9/L *vs.* 232.78 ×10^9/L, P=0.03).

The white blood cell count, total cholesterol (recorded only in MHH), creatinine (recorded only in MHH), alkaline phosphatase, calcium, total bilirubin, albumin, hemoglobin, aspartate transaminase, alanine transaminase, and amylase

634

Zhao et al. Exploring risk factors for PC in CP patients

Table 3 The results of multivariate analysis using a logistic regression model

Clinical parameters	Estimate	Standard error	z value	OR	95% CI		P value
DM	0.79	1.04	0.76	2.20	0.28	20.71	0.449
COP	1.48	0.69	2.13	4.37	1.11	18.40	0.033
Hyperlipidemia	0.09	0.76	0.12	1.10	0.26	5.62	0.901
LW	1.17	0.70	1.66	3.21	0.76	12.91	0.096
Insulin use	0.37	0.81	0.45	1.44	0.29	7.87	0.651
Diarrhea	−0.01	0.72	−0.01	0.99	0.23	4.09	0.992
Glucose value	0.01	0.01	0.78	1.01	0.99	1.02	0.436
PLT value	0.01	0.00	2.04	1.01	1.00	1.01	0.042

OR, odds ratio; CI, confidence interval; DM, diabetes mellitus; COP, cyst of the pancreas; LW, loss of weight; PLT, platelet.

were not different between the two groups. Next, we accounted for the candidate variables in the multivariable logistic regression model and found that the independent predictors of PC included COP [odds ratio (OR): 4.37, 95% confidence interval (CI): 1.11 to 18.40, P=0.033] and PLT count (OR: 1.01 per 1 increment, 95% CI: 1.00 to 1.01, P=0.042) (*Table 3*). In addition, LW as a general predictor for tumors, exhibited a certain influence on PC (OR: 3.21, 95% CI: 0.76 to 12.91, P=0.096); therefore, we included these three factors to construct a logistic regressive model to predict the likelihood of PC progression in CP patients. An equation was used to assess the probability of PC (*Figure 2*). Based on this equation, a diagnostic receiver operating characteristic (ROC) curve was plotted in *Figure 3A* with a cut-off point of 0.0128, a sensitivity of 0.78 (CI: 0.67 to 1.00), a specificity of 0.76 (CI: 0.44 to 0.93), and an area under the curve (AUC) of 0.82 (CI: 0.71 to 0.93).

Validation of the logistic risk assessment model

To validate the value of the prediction model for predicting the probability of PC development, we applied a third cohort from BJCYH that included 415 CP patients. Youden's J statistic was adopted as a cut-off point value for predicting PC among CP patients. The ROC curve was constructed showing a sensitivity of 0.71 (CI: 0.43 to 1.00), a specificity of 0.85 (CI: 0.45 to 0.94), and AUC of 0.73 (CI: 0.48 to 0.97) (*Figure 3B*).

Discussion

Patients with CP tend to have a high risk of PC

$$\ln\left[\frac{p}{1-p}\right] = -6.68 + 1.55COP + 1.23LW + 0.0046PLT$$

$$\text{probability of developing pancreatic cancer} = \frac{e^{-6.68+1.55COP+1.23LW+0.0046PLT}}{1+e^{-6.68+1.55COP+1.23LW+0.0046PLT}}$$

COP =1 if cyst of the pancreas were diagnosed
COP =0 if cyst of the pancreas were not diagnosed
LW =1 if loss of weight was diagnosed
LW =0 if loss of weight was not diagnosed
PLT was the abbreviation of platelet count value ($\times 10^9$/L)

Figure 2 A logistic equation for predicting PC in CP with a cut-off point value of 0.0128. CP, chronic pancreatitis; PC, pancreatic cancer; COP, cyst of the pancreas; LW, loss of weight; PLT, platelet.

development (13). In this retrospective cohort study, we followed up the EMRs of 2,545 CP patients in two cohorts and found that 14 (0.55%) patients developed PC (average 7 years after CP diagnosis). To our knowledge, this study is the first to include comprehensive clinical factors, including alterations in demographics, comorbidities, and laboratory parameters of CP patients for risk assessment. Our results demonstrated that CP patients with COP, LW, and high PLT counts had a higher probability of PC progression than the general CP population. Then, we constructed a logistic regression predictive equation based on these three risk factors and validated its value with a third cohort showing moderate predictive efficiency.

In this study, we found that cystic lesions of the pancreas were associated with a risk of pancreatic malignancy. In the multiple logistic regression model, the morbidity of PC in CP patients with pancreatic cyst lesions was 4.4-times greater than that in those without this disease. Cystic alterations of the pancreas commonly occurred in individuals
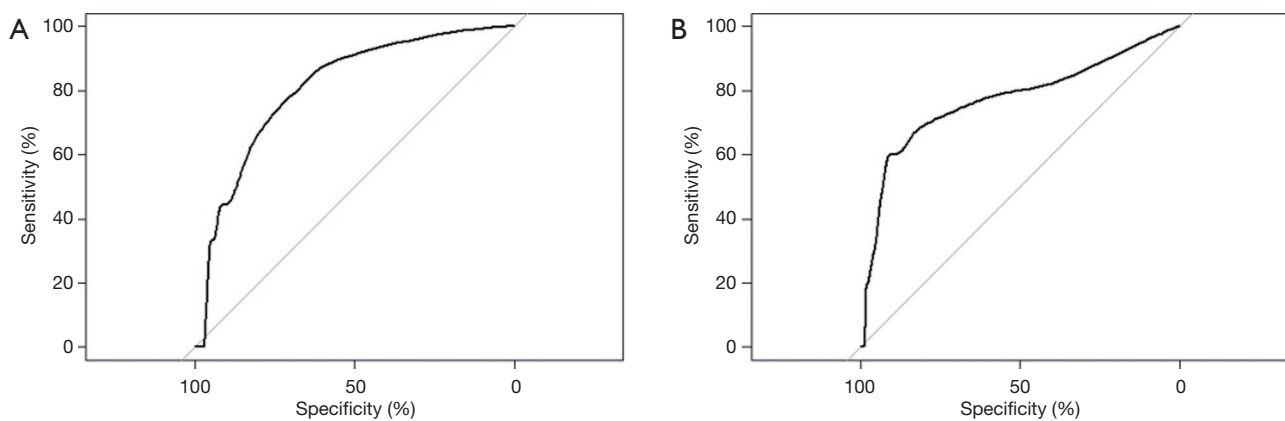
Figure 3 Evaluation for the logistic regressive model in predicting PC. (A) A ROC curve was constructed, showing a sensitivity of 0.78 (CI: 0.67 to 1.00%), a specificity of 0.76 (CI: 0.44 to 0.93), and an AUC of 0.82 (CI: 0.71 to 0.93) in the derivation cohort; (B) in the validation cohort, the sensitivity and specificity of the prediction model were 0.71 (CI: 0.43 to 1.00) and 0.85 (CI: 0.45 to 0.94), respectively. The AUC was 0.73 (CI: 0.48 to 0.97). PC, pancreatic cancer; ROC, receiver operating characteristic; CI, confidence interval; AUC, area under the curve.

with mucinous cystic neoplasia (MCN) and intraductal papillary mucinous neoplasia (IPMN). The frequency of adenocarcinoma or high-grade dysplasia in MCN patients was 14.9%, and the malignant rate was 5.5% in individuals with branch duct IPMN (14,15). In a recent veteran cohort that included 520,970 individuals, the hazard ratio (HR) of PC in patients with pancreatic cysts was 19.64-times higher than that of those without cysts (P<0.0001) (16). Among the 17 PC patients, twelve had a history of AP or CP combined with cyst lesions (16). Another study revealed that the 3-year risk of developing PC was 1.2% in patients with pancreatic cyst lesions, with an HR of 10.0 (17). The results of our analyses showed that the incidence rate of cyst lesions in individuals with CP was approximately 17.0% in the derivation cohorts. It seemed that the ratio of cysts in the older population (WFBMC) was higher than that in the young group (MHH) (21.7% *vs.* 15.6%, P=0.002). The results of the derivation and validation cohorts strongly indicated that the presence of pancreatic cyst lesions was a high-risk factor predicting PC development in CP patients. Our findings suggested that CP patients with pancreatic cyst lesions should be closely monitored, especially in old populations. At the genome level, it was revealed that individuals with MCNs more frequently had KRAS mutations and a higher risk of malignant progression (18). KRAS/GNAS mutations were also identified in the fluid of pancreatic cysts in 3 (30%) patients with MCNs and 56 (100%) patients with IPMN (19). These findings revealed a potential link between the phenotype of cystic lesions

and the development of PC. Unfortunately, in the EMRs datasets, we were not able to access the genomic data or the diameter of the cyst lesions, which were also reported to be related to malignancy (14). Thus, more imaging characteristics of the pancreas combined with genetic mutation data have greater potential for PC prediction.

In the derivation cohorts, the PLT value was another significant variable associated with PC development in the multiple logistic regression model (P=0.042). The PLT level in the PC group increased significantly compared with that in the CP group (289.10 ×10^9/L *vs.* 232.78 ×10^9/L, P=0.03). An increasing number of studies have shown that elevated PLT levels are associated with tumor cell growth and angiogenesis (20). For example, it was reported that the PLT count in pancreatic ductal adenocarcinoma (PDAC) patients was higher than that in CP patients with an inflammatory mass of the pancreatic head (21). Similarly, the level of PLT in PC patients in an early tumor-node-metastasis (TNM) stage (I, II) was significantly higher than that in healthy controls (22). In addition, a previous study verified that the PLT-lymphocyte ratio was an independent predictor of PC survival (23). This study showed that in the PC group, the PLT level was higher than that in the CP group before PC diagnosis. During tumor progression, some cytokines, such as thrombopoietin, collagen and proteins, are secreted and promoted the production of PLT (24). Additionally, angiogenic factors such as vascular endothelial growth factor, PLT-derived growth factor and angiostatic PLT factor 4 have been reported to be highly expressed

in PLT, which facilitates tumor neovascularity (25). Apart from PLT count, previous studies also found that the mean PLT volume was elevated in the context of cancer (26). Overall, these results indicated that changes in PLT characteristics were associated with PC progression. Therefore, periodic monitoring of the PLT count is recommended for CP patients. As a common symptom of CP, 39–56% of patients presented LW (27). The incidence of LW was significantly different between the PC and CP groups in the univariate analysis. The overall incidence of LW in the derivation cohorts was 11.6%, and it seemed that in the aged population (WFUBMC), the occurrence rate of LW was higher than that in the young population (MHH) (22.1% *vs.* 8.2%, P<0.01). BMI, another clinical factor reflecting individual nutritional status, appeared to be lower in the PC groups; however, this difference was not statistically significant. The dysfunction of pancreatic exocrine cells is aggravated by the progression of inflammation of the pancreas, which mainly reduces digestive function and leads to malnutrition. It was reported that substantial (>5%) LW was present in 71.5% of patients with PDAC, and when the threshold was raised to greater than 10%, LW was associated with worse survival (28). Consequently, the results of our study suggested that we should pay close attention to CP patients with LW.

The age at the onset of CP and smoking history were reported as risk factors for PC in a previous study (6). However, there was no significant difference between the CP and PC groups in the current study. It seemed that in the older population (WFUBMC), the incidence of PC progression was higher than that in the young group (MHH), but the difference was not significant (0.98% *vs.* 0.41%, P=0.1). Smoking and heavy alcohol consumption are also general risk factors for most cancers, including PC (29). However, we did not find an obvious influence on PC development mainly because of the limitation of sample size. In the univariate analysis of the MHH data, 87.5% (7/8) of PC patients were African-American, indicating racial PC susceptibility. The results of the univariate analysis demonstrated that the incidence of type II DM and the level of glucose were higher in the PC group than in the CP group. This result was in agreement with a previous study that revealed that DM was associated with an elevated risk of new-onset cancer, especially PC (30). In addition, the levels of glucose, insulin, and insulin resistance were correlated with PDAC (31). Insulin and/or high levels of insulin-like growth factor-1 may activate the PI3K signaling cascade, which plays an important role in PDAC development (32).

Carbohydrate antigen (CA) 19-9, as a diagnostic and prognostic biomarker of PC, was extensively used in clinical practice (33); therefore, we investigated the value of CA19-9 in the derivation cohorts. In the WFUBMC dataset, a significant increase of CA19-9 was found in the PC group compared with CP group (397.0 *vs.* 326.9 U/mL, P<0.01). However, the number of cases with CA19-9 values was very limited, and the results of CA19-9 were recorded shortly before the PC diagnosis, which meant that the high level of CA19-9 reflected the phenotype of PC. In the remaining two datasets, the number of patients with CA19-9 data was too small to perform statistical calculations. Therefore, it was difficult to include this variable in the multivariate logistic regression analysis.

A limitation of this study is that some factors, such as genomic mutation data, features of pancreatic image and tumor biomarkers, were not available in the datasets for risk assessments. Another limitation is that the PC patients diagnosed within 2 years after CP diagnosis were excluded according to the previously established criteria to avoid initial misclassification. In this sense, some true positive cases were omitted, which might have caused enrollment bias. Finally, the derivation data were from the United States population, and the validation data were from the Chinese population. Consequently, the predictive efficiency was subject to ethnic differences.

Conclusions

In conclusion, our study revealed that COP, LW, and elevated levels of PLT were risk factors for PC development in CP patients. Therefore, closer monitoring should be performed for CP patients with these accompanied diagnoses.

Acknowledgments

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tcr.2019.11.49). The authors have no conflicts

of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. We were supported by the study of "Linking Electronic Medical Records to Medicare Claims Data" (IRB00030804) for the access to these data. The clinical data of BJCYH were collected from the computer-based patient record systems, and this study was permitted by the research ethics committee (2019-243). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Zhu H, Li T, Du Y, et al. Pancreatic cancer: challenges and opportunities. BMC Med 2018;16:214.
2. Cronin KA, Lake AJ, Scott S, et al. Annual report to the nation on the status of cancer, part I: national cancer statistics. Cancer 2018;124:2785-800.
3. Rawla P, Sunkara T, Gaduputi V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. World J Oncol 2019;10:10-27.
4. Kleeff J, Whitcomb DC, Shimosegawa T, et al. Chronic pancreatitis. Nat Rev Dis Primers 2017;3:17060.
5. Shafqet M, Sharzehi K. Diabetes and the pancreatobiliary diseases. Curr Treat Options Gastroenterol 2017;15:508-19.
6. Tsai HJ, Chang JS. Environmental risk factors of pancreatic cancer. J Clin Med 2019. doi: 10.3390/jcm8091427.
7. Rijkers AP, Bakker OJ, Ahmed Ali U, et al. Risk of pancreatic cancer after a primary episode of acute pancreatitis. Pancreas 2017;46:1018-22.
8. Sollie S, Michaud DS, Sarker D, et al. Chronic inflammation markers are associated with risk of pancreatic cancer in the Swedish AMORIS cohort study. BMC Cancer 2019;19:858.
9. Syed A, Babich O, Thakkar P, et al. Defining pancreatitis as a risk factor for pancreatic cancer: the role, incidence, and timeline of development. Pancreas 2019;48:1098-101.
10. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2016;Suppl 1:S48-61.
11. Giefer MJ, Lowe ME, Werlin SL, et al. Early-onset acute recurrent and chronic pancreatitis is associated with PRSS1 or CTRC gene mutations. J Pediatr 2017;186:95-100.
12. Schwarzenberg SJ, Bellin M, Husain SZ, et al. Pediatric chronic pancreatitis is associated with genetic risk factors and substantial disease burden. J Pediatr 2015;166:890-6.e1.
13. Vege SS. Continuing medical education questions: September 2017: chronic pancreatitis and pancreatic cancer risk: a systematic review and meta-analysis. Am J Gastroenterol 2017;112:1373.
14. Pergolini I, Sahora K, Ferrone CR, et al. Long-term risk of pancreatic malignancy in patients with branch duct intraductal papillary mucinous neoplasm in a referral center. Gastroenterology 2017;153:1284-94.e1.
15. Postlewait LM, Ethun CG, McInnis MR, et al. Association of preoperative risk factors with malignancy in pancreatic mucinous cystic neoplasms: a multicenter study. JAMA Surg 2017;152:19-25.
16. Munigala S, Gelrud A, Agarwal B. Risk of pancreatic cancer in patients with pancreatic cyst. Gastrointest Endosc 2016;84:81-6.
17. Ohno E, Hirooka Y, Kawashima H, et al. Natural history of pancreatic cystic lesions: a multicenter prospective observational study for evaluating the risk of pancreatic cancer. J Gastroenterol Hepatol 2018;33:320-8.
18. Shibata H, Ohike N, Norose T, et al. Mucinous cystic neoplasms lined by abundant mucinous epithelium frequently involve KRAS mutations and malignant progression. Anticancer Res 2017;37:7063-8.
19. Singhi AD, McGrath K, Brand RE, et al. Preoperative next-generation sequencing of pancreatic cyst fluid is highly accurate in cyst classification and detection of advanced neoplasia. Gut 2018;67:2131-41.
20. Sharma D, Brummel-Ziedins KE, Bouchard BA, et al. Platelets in tumor progression: a host factor that offers multiple potential targets in the treatment of cancer. J Cell Physiol 2014;229:1005-15.
21. Kakkat S, Rajan R, Sindhu RS, et al. Comparison of platelet-lymphocyte ratio and CA 19-9 in differentiating benign from malignant head masses in patients with chronic pancreatitis. Indian J Gastroenterol 2017;36:263-7.
22. Sabrkhany S, Kuijpers MJE, van Kuijk SMJ, et al. A combination of platelet features allows detection of early-

stage cancer. Eur J Cancer 2017;80:5-13.

23. Song W, Tian C, Wang K, et al. Preoperative platelet lymphocyte ratio as independent predictors of prognosis in pancreatic cancer: a systematic review and meta-analysis. PLoS One 2017;12:e0178762.

24. Lin RJ, Afshar-Kharghan V, Schafer AI. Paraneoplastic thrombocytosis: the secrets of tumor self-promotion. Blood 2014;124:184-7.

25. Yao L, Dong H, Luo Y, et al. Net platelet angiogenic activity (NPAA) correlates with progression and prognosis of non-small cell lung cancer. PLoS One 2014;9:e96206.

26. Pyo JS, Sohn JH, Kang G. Diagnostic and prognostic roles of the mean platelet volume in malignant tumors: a systematic review and meta-analysis. Platelets 2016;27:722-8.

27. Szücs Á, Marjai T, Szentesi A, et al. Chronic pancreatitis: multicentre prospective data collection and analysis by the Hungarian Pancreatic Study Group. PLoS One 2017;12:e0171420.

28. Nemer L, Krishna SG, Shah ZK, et al. Predictors of

pancreatic cancer-associated weight loss and nutritional interventions. Pancreas 2017;46:1152-7.

29. Korc M, Jeon CY, Edderkaoui M, et al. Tobacco and alcohol as risk factors for pancreatic cancer. Best Pract Res Clin Gastroenterol 2017;31:529-36.

30. Biadgo B, Abebe M. Type 2 Diabetes mellitus and its association with the risk of pancreatic carcinogenesis: a review. Korean J Gastroenterol 2016;67:168-77.

31. Carreras-Torres R, Johansson M, Gaborieau V, et al. The role of obesity, type 2 diabetes, and metabolic factors in pancreatic cancer: a mendelian randomization study. J Natl Cancer Inst 2017. doi: 10.1093/jnci/djx012.

32. Pasquale V, Dugnani E, Liberati D, et al. Glucose metabolism during tumorigenesis in the genetic mouse model of pancreatic cancer. Acta Diabetol 2019;56:1013-22.

33. Mirkin KA, Hollenbeak CS, Wong J. Prognostic impact of carbohydrate antigen 19-9 level at diagnosis in resected stage I-III pancreatic adenocarcinoma: a U.S. population study. J Gastrointest Oncol 2017;8:778-88.