Received: 26 October 2022 | Accepted: 15 February 2023

DOI: 10.1111/jebm.12516

ARTICLE

WILEY

Risk prediction of hepatobiliary and pancreatic cancers in elderly Chinese: The Dongfeng-Tongji cohort

Jingjing Lyu[1,2] | Xin Guan[1,3] | Yuhan Zhou[1,3] | Huan Guo[1,3] | Shanshan Cheng[1,2] | Chaolong Wang[1,2]

[1]Ministry of Education Key Laboratory of Environment and Health, State Key Laboratory of Environmental Health (Incubating), School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]Department of Occupational and Environmental Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Correspondence

Chaolong Wang, Ministry of Education Key Laboratory of Environment and Health, State Key Laboratory of Environmental Health (Incubating), School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.
Email: chaolong@hust.edu.cn

Shanshan Cheng, Ministry of Education Key Laboratory of Environment and Health, State Key Laboratory of Environmental Health (Incubating), School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.
Email: sscheng@hust.edu.cn

Funding information

National Key Research and Development Program of China, Grant/Award Number: 2019YFC1315905; Natural Science Foundation of China, Grant/Award Number: 82021005; Health Commission of Hubei Province Scientific Research Project, Grant/Award Number: WJ2021M102; Fundamental Research Funds for the Central Universities, Grant/Award Number: 2019kfyXJJS036

Abstract

Aim: Hepatobiliary and pancreatic (HBP) cancers are among the deadliest malignancies. The objective of the study is to build cost-effective models to identify high-risk individuals for early diagnosis and substantially to reduce the burden of HBP cancers.

Methods: Based on the prospective Dongfeng-Tongji cohort with ~6 years follow-up, we identified 162 incident cases of hepatocellular carcinoma (HCC), 53 of biliary tract cancer (BTC), and 58 of pancreatic cancer (PC). We matched three controls to each case by age, sex, and hospital. We applied conditional logistic regression to identify predictive clinical variables, from which we constructed clinical risk scores (CRSs). We evaluated the utility of CRSs in stratifying high-risk individuals by 10-fold cross-validation.

Results: Among 50 variables we screened, 6 were independent predictors of HCC, with the top ones being hepatitis (OR = 8.51, 95% CI (3.83, 18.9)), plateletcrit (OR = 0.57, 95% CI (0.42, 0.78)), and alanine aminotransferase (OR = 2.06, 95% CI (1.39, 3.06)). Gallstone (OR = 2.70, 95% CI (1.17, 6.24)) and direct bilirubin (OR = 1.58, 95% CI (1.08, 2.31)) were predictive of BTC, while hyperlipidemia (OR = 2.56, 95% CI (1.12, 5.82)) and fasting blood glucose (OR = 2.00, 95% CI (1.26, 3.15)) were predictive of PC. The CRSs achieved AUCs of 0.784 for HCC, 0.648 for BTC, and 0.666 for PC, respectively. When applying to the full cohort with age and sex included as predictors, the AUCs were increased to 0.818, 0.704, and 0.699, respectively.

Conclusions: Disease history and routine clinical variables are predictive of incident HBP cancers in elderly Chinese.

KEYWORDS

hepatobiliary cancer, nested case-control study, pancreatic cancer, prospective cohort, risk stratification

1 | INTRODUCTION

The liver, biliary tracts (including gallbladder and bile ducts), and pancreas are major digestive glands, which are anatomically connected and constitute the hepatobiliary and pancreatic (HBP) system.[1] Liver cancer, including hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (IHC), is the sixth most commonly diagnosed cancer, accounting for 4.7% of cancer incidence in 2018 globally.[2] Biliary tract cancer (BTC) and pancreatic cancer (PC) are rare in the general population, although substantial variation exists in the incidence rates across geographic regions and ethnic groups.[3,4] Importantly, HBP cancers are among the deadliest malignancies, particularly with pancreatic and liver cancers having 5-year relative survival rates as low as 10% and 20%, respectively.[5] Despite advances in the diagnosis and treatment, both the incidence and the mortality of HBP cancers have shown rising trends in recent years.[6] Because HBP cancers often cause few symptoms at their early stage, most patients, have developed to an advanced stage with a high level of tumor invasiveness and distant metastasis.[3,7,8] Compared to those at the advanced stage, patients receiving treatment at the early stage can benefit nearly 90% increase in disease-free survival.[9] Nevertheless, population-based screening for rare cancers is not recommended. Therefore, cost-effective models to identify and prioritize high-risk individuals for early diagnosis are important to reduce the burden of HBP cancers.

Current risk prediction models for HBP cancers are often based on epidemiological risk factors, including medical history, family history of disease, and lifestyle behaviors. Specifically, age, infections with hepatitis B virus (HBV) and hepatitis C virus (HCV), alcohol consumption, smoking, and diabetes are high-risk factors for HCC, which accounts for ~90% of liver cancer.[8,10] Following the current screening guideline of HCC, risk prediction models mainly focus on carriers of HBV or HCV,[11] despite the occurrence of HCC in many non-carriers of these viruses.[12–14] A few prospective studies have shown promising results in predicting HCC incidence in the general population using information of lifestyle and disease history,[11,15] as well as serum biochemistry indices.[16] It has been shown that transaminases have high predictive power of HCC risk, even without HBV or HCV infection status.[16] In contrast to HCC, relatively few prospective studies have been performed to evaluate risk prediction of BTC and PC. Major risk factors for BTC include age, female sex, obesity, history of gallstones and cholangitis, family history, and exposure to specific chemicals and heavy metals.[7,17,18] Currently, screening of BTC has focused on individuals with gallstones and there is no highly suggestive biomarkers from physical exam or blood test.[9] For PC, screening has mostly focused on familial PC for those with a family history or high-risk germline mutations.[19–21] There is no evidence that screening for PC in the general asymptomatic population can reduce morbidity or mortality, primarily due to its low incidence.[22,23] Epidemiological risk factors of PC include newly onset diabetes, preexisting diabetes, age, smoking, obesity, and chronic pancreatitis.[3]

While many studies of HCC have been performed in East Asian populations,[11,15,16] epidemiological studies of BTC and PC are primarily based on Europeans.[24,25] The transferability of findings from Europeans to other population remains underexplored. Here, we aim to identify risk factors and clinical indices associated with the incidence of HCC, BTC, and PC in a prospective elderly Chinese cohort.[26] We evaluate and compare the predictive value of clinical data that are easily accessible from questionnaires and routine physical examinations, with the goal to help develop cost-effective risk stratification models for HBP cancers in the Chinese population.

2 | MATERIALS AND METHODS

2.1 | Cohort description

We designed nested case-control studies of HCC, BTC, and PC based on the Dongfeng-Tongji (DFTJ) cohort. The DFTJ cohort is a prospective cohort of retired employees of the Dongfeng Motor Corporation (DMC) in Shiyan, Hubei, China.[26] A total of 27,009 participants (55.4% female) were enrolled from September 2008 to June 2010 with a mean age of 63.6 years. The first follow-up was performed from April to October in 2013, with a follow-up rate of 96.2%. In addition, 14,120 participants were newly enrolled, totaling 38,295 participants in 2013. All information was collected at the baseline including demographics, socioeconomics, lifestyle, family history, and past medical history were collected through face-to-face interviews using a semistructured questionnaire by trained interviewers. Routine physical examinations were performed at approximately the same time. Details of the DFTJ cohort have been described elsewhere.[26] Disease incidences were followed up till December 31, 2018, based on the healthcare service system of DMC and the electronic medical records. HBP cancers were determined according to the International Classification of Diseases, 10th Revision (ICD-10) with C22.0 for HCC, C22.1 and C23-24 for BTC, and C25 for PC.

We selected 55 variables obtained from questionnaires and routine physical examinations to evaluate the prediction of incident HCC, BTC, and PC. These variables include demographics, lifestyle, disease histories, height, weight, body mass index (BMI), waist circumference, waist-height ratio (WHtR), blood pressure, and 32 laboratory measurements on fasting blood samples, covering lipids, fasting blood glucose, hematologic traits, and hepatic and renal functions (Supplementary Table S1). Cigarette smoking and alcohol drinking status were self-reported as never, former, and current. Smoking refers to one cigarette per day for more than half a year, while drinking refers to once per week for more than half a year. We grouped former and current smokers/drinkers as ever smokers/drinkers. Disease histories were self-reported except for diabetes, which was determined by self-reported diabetes or use of insulin or hypoglycemic drugs during the last two weeks.

2.2 | Quality controls

Quality controls (QC) were performed separately on data from the 2008 and 2013 enrollments (Supplementary Figure S1). We excluded

individuals with a disease history of cancer or diagnosis of cancer other than HBP cancers during follow-up, as well as those missing disease history or more than 10 out of 32 laboratory measurements, leaving 18,471 and 25,274 individuals in the DFTJ 2008 and 2013 data sets, respectively. We then excluded variables with more than 20% missing data, including basophil counts (BASO), eosinophil counts (EOS), γ-glutamyl transferase (GGT), monocyte counts (MONO) in the 2008 data set, and mean corpuscular hemoglobin concentration (MCHC) in the 2013 data set, leaving 50 common variables for downstream analyses (Table 1).

We harmonized the data to remove batch effects between hospitals. Specifically, we regressed out hospital indicators, age, and sex for quantitative variables, and standardized the residuals to mean of 0 and standard deviation (SD) of 1. Missing data were imputed by the predictive mean matching (PMM) method implemented in the R package "mice" (version 3.11.0).[27,28] We performed multiple imputations for 5 times, each with 10 iterations,[29] and took the mean and the mode of the imputed values for quantitative and categorical variables, respectively.

2.3 | Nested case-control sample selection

We excluded incident cases who were diagnosed with HBP cancers within 3 months because these individuals might be existing cases but were undiagnosed at the baseline. We also excluded incident cases diagnosed in more than 70 months (~6 years) from the baseline, because most of these cases were from the 2008 data set who were also included in the 2013 data set (Supplementary Figure S2). For cases in both data sets, we used their 2013 data as the baseline. Finally, we matched three cancer-free controls to each incident case based on age, sex, and hospital. Exact match was applied to sex and hospital, while age was matched to minimize the difference between cases and controls. Combining the 2008 and the 2013 data sets, we selected 162 cases and 486 controls for HCC, 53 cases and 159 controls for BTC, and 58 cases and 174 controls for PC (Supplementary Figure S1). No controls died before cancer diagnosis of their matched cases, while a small proportion of controls died before the end of follow-up (4.5%, 7.5%, and 6% for HCC, BTC, and PC, respectively).

2.4 | Statistical analysis

We compared baseline characteristics between incident cases and their matched controls, by the Wald test from conditional logistic regression (CLR) for continuous variables[30] and by the Cochran-Mantel-Haenszel (CMH) test for categorical variables.[31] For variables with two-sided $p < 0.05$ in the univariable analysis, we first reduced collinearity by excluding one with higher mean correlation with other variables from a pair of highly correlated variables (Spearman's correlation $r_s > 0.8$). The remaining variables were jointly fit with a multivariate CLR model to explore their independent effects, represented by the odds ratio (OR), on HBP cancers.

We defined a clinical risk score (CRS) for each cancer as $CRS_i = \sum_j \beta_j x_{ij}$, where β_j is the coefficient of variable j from the multivariable CLR model and x_{ij} is the value of variable j for individual i. We validated CRS via 10-fold cross-validation (CV) by randomly splitting the nested case-control samples and repeated the CV for 20 times. We used the receiver operating characteristic (ROC) curves and the area under ROC curve (AUC) to evaluate the performance of CRS. Computation of AUC was done using the R package "pROC" (version 1.16.2).[32] Furthermore, we stratified samples into high- and low-risk groups of each cancer based on the median of CRS, and compared their cumulative incidence by the Cox model.

We also evaluated the predictive value of CRS for each cancer by including all controls without matching. There were 18,215 and 25,108 cancer-free controls from the DFTJ 2008 and 2013 data sets, respectively (Supplementary Figure S1). For each cancer, we combined cases and controls from the 2008 and the 2013 data sets to maximize the sample size. As 13,430 control samples were included in both data sets, we built predictive models for each cancer using mixed-effects logistic regression: logit $(Y_i) = X_i\alpha + b_i$, where X_i was a $1 \times p$ row vector of covariates for individual i, α was a $p \times 1$ column vector of fixed effects, and $b_i \sim N(0, \sigma^2)$ was a normally distributed random intercept for individual i. We considered three models with X_i being (1) age, sex, and batch indicator of baseline year (i.e., 2008 vs. 2013), (2) CRS and batch indicator, and (3) age, sex, CRS, and batch indicator, respectively. The logistic mixed models were implemented using the R package "lme4" (version 1.1.31).[33] For each model, we computed the AUC based on 10-fold CV with 20 repeats. We compared model 3 and model 1 by the continuous net reclassification improvement (NRI) and integrated discrimination improvement (IDI) indices. In addition, we evaluated the specificity (SPE), the positive predictive value (PPV), negative predictive value (NPV), and Youden index of model 3 for each cancer given a fixed sensitivity (SEN).

Finally, we evaluated the risk prediction model of HCC by including the HBV infection status. The analysis was based on 25,216 DFTJ participants who received the hepatitis B surface antigen (HBsAg) test in 2013, including 82 incident HCC cases and 25,061 cancer-free controls (excluding 39 BTC and 34 PC cases). Participants tested positive for HBsAg was recorded as HBV positive. We built three LR models to predict incident HCC based on (1) age, sex, and HBV status, (2) age, sex, and CRS, and (3) age, sex, HBV status, and CRS. Models were evaluated by ROC curves and AUC based on 10-fold CV with 20 repeats. Prediction models were implemented using the R package "caret" (version 6.0.86).[34]

All statistical analyses were performed using R (version 4.1.2) and a two-side p value < 0.05 was considered as statistically significant.

3 | RESULTS

From the DFTJ cohort, we identified 162, 53, and 58 incidences of HCC, BTC, and PC, respectively, with diagnosis within 3 to 70 months from the baseline (Supplementary Figures S1 and S2). The mean (SD) age at baseline and male proportion were 67.14 (7.89) years and 75%

TABLE 1 Demographic and clinical features of hepatocellular carcinoma, biliary tract cancer, pancreatic cancer incident cases and their matched controls.

Variables	Hepatocellular carcinoma			Biliary tract cancer			Pancreatic cancer		
	Cases ($n = 162$)	Controls ($n = 486$)	p Value	Cases ($n = 53$)	Controls ($n = 159$)	p Value	Cases ($n = 58$)	Controls ($n = 174$)	p Value
Demography									
Age (years)	67.14 ± 7.89	67.14 ± 7.87	0.491	69.37 ± 8.02	69.35 ± 7.94	0.128	69.22 ± 7.09	69.22 ± 7.03	0.228
Sex (male)	121 (0.75)	363 (0.75)	1.000	23 (0.43)	69 (0.43)	1.000	29 (0.50)	87 (0.50)	1.000
Lifestyle									
Cigarette smoking	77 (0.48)	207 (0.43)	0.262	15 (0.28)	43 (0.27)	1.000	18 (0.31)	51 (0.29)	0.917
Alcohol drinking	68 (0.42)	190 (0.39)	0.553	15 (0.28)	38 (0.24)	0.608	17 (0.29)	56 (0.32)	0.776
Disease history									
Hypertension	64 (0.40)	209 (0.43)	0.477	27 (0.51)	79 (0.50)	1.000	28 (0.48)	80 (0.46)	0.883
Hyperlipidemia	27 (0.17)	111 (0.23)	0.116	12 (0.23)	36 (0.23)	1.000	19 (0.33)	32 (0.18)	0.033
Diabetes	40 (0.25)	74 (0.15)	0.008	12 (0.23)	29 (0.18)	0.666	16 (0.45)	22 (0.13)	0.012
CHD	30 (0.19)	92 (0.19)	0.969	12 (0.23)	33 (0.21)	0.922	19 (0.33)	43 (0.25)	0.293
Stroke	5 (0.03)	31 (0.06)	0.191	2 (0.04)	11 (0.07)	0.622	3 (0.05)	11 (0.06)	1.000
Emphysema	6 (0.04)	14 (0.03)	0.793	2 (0.04)	5 (0.03)	0.827	0 (0.00)	2 (0.01)	1.000
Bronchitis	23 (0.14)	59 (0.12)	0.586	6 (0.11)	21 (0.13)	0.906	4 (0.07)	20 (0.11)	0.460
Asthma	4 (0.03)	22 (0.05)	0.352	4 (0.08)	9 (0.06)	0.869	0 (0.00)	7 (0.04)	0.251
Tuberculosis	6 (0.04)	21 (0.04)	0.908	2 (0.04)	6 (0.04)	1.000	1 (0.02)	8 (0.05)	0.549
Gallstone	27 (0.17)	62 (0.13)	0.205	14 (0.26)	21 (0.13)	0.039	7 (0.12)	24 (0.14)	0.914
Hepatitis	54 (0.34)	22 (0.05)	2.16×10^{-21}	5 (0.09)	8 (0.05)	0.369	3 (0.05)	6 (0.03)	0.841
Nephritis	10 (0.06)	17 (0.04)	0.220	3 (0.06)	7 (0.04)	1.000	3 (0.05)	6 (0.03)	0.847
Metabolic traits									
Height (cm)	163.26 ± 8.16	163.82 ± 7.80	0.254	158.88 ± 8.25	158.86 ± 8.03	0.989	160.92 ± 8.74	160.30 ± 8.27	0.478
Weight (kg)	64.70 ± 11.38	65.88 ± 10.96	0.181	62.24 ± 10.10	61.8 ± 10.07	0.744	62.48 ± 9.17	63.27 ± 10.74	0.565
BMI (kg/m^2)	24.27 ± 3.82	24.50 ± 3.39	0.457	24.60 ± 3.20	24.44 ± 3.24	0.721	24.09 ± 2.64	24.58 ± 3.38	0.270
Waist (cm)	85.40 ± 10.28	85.15 ± 9.43	0.850	85.42 ± 9.16	84.21 ± 8.86	0.346	84.72 ± 8.43	83.37 ± 9.01	0.289
WHtR	0.52 ± 0.07	0.52 ± 0.06	0.509	0.54 ± 0.06	0.53 ± 0.06	0.349	0.53 ± 0.05	0.52 ± 0.05	0.379
SBP (mm Hg)	136.96 ± 21.07	138.07 ± 21.18	0.580	143.28 ± 21.44	143.71 ± 21.75	0.937	140.64 ± 21.98	137.21 ± 21.60	0.251
DBP (mm Hg)	76.86 ± 11.74	79.74 ± 11.52	0.004	80.49 ± 11.36	78.56 ± 12.51	0.287	79.69 ± 11.70	77.82 ± 11.83	0.281
TG (mmol/L)	1.20 ± 0.84	1.40 ± 0.88	0.010	1.42 ± 0.76	1.57 ± 1.46	0.410	1.49 ± 0.69	1.46 ± 1.24	0.843
LDL-C (mmol/L)	2.58 ± 0.82	2.93 ± 0.79	8.41×10^{-7}	2.73 ± 0.86	2.95 ± 0.87	0.128	2.73 ± 0.96	2.84 ± 0.89	0.426
HDL-C (mmol/L)	1.39 ± 0.37	1.37 ± 0.36	0.555	1.42 ± 0.41	1.38 ± 0.38	0.440	1.35 ± 0.35	1.39 ± 0.37	0.460
TC (mmol/L)	4.52 ± 0.95	4.92 ± 0.94	3.35×10^{-6}	4.82 ± 0.96	4.97 ± 1.00	0.357	4.78 ± 1.20	4.88 ± 1.05	0.547
FBG (mmol/L)	6.53 ± 2.29	6.13 ± 1.88	0.034	6.89 ± 2.62	6.37 ± 2.45	0.188	7.03 ± 2.99	5.92 ± 1.09	3.76×10^{-5}

(Continues)

TABLE 1 (Continued)

Variables	Hepatocellular carcinoma			Biliary tract cancer			Pancreatic cancer		
	Cases ($n = 162$)	Controls ($n = 486$)	*p* Value	Cases ($n = 53$)	Controls ($n = 159$)	*p* Value	Cases ($n = 58$)	Controls ($n = 174$)	*p* Value
Hematologic traits									
WBC (10^9/L)	5.61 ± 1.88	5.96 ± 1.56	0.012	5.71 ± 1.53	5.85 ± 1.53	0.553	5.89 ± 1.64	5.90 ± 1.64	0.975
LYM (10^9/L)	1.88 ± 0.87	1.83 ± 0.71	0.392	1.82 ± 0.73	1.79 ± 0.61	0.756	1.84 ± 0.65	1.86 ± 0.67	0.829
NEUT (10^9/L)	3.10 ± 1.31	3.56 ± 1.11	9.24×10^{-6}	3.35 ± 1.02	3.53 ± 1.26	0.318	3.50 ± 1.18	3.43 ± 1.20	0.704
RBC (10^{12}/L)	4.47 ± 0.51	4.64 ± 0.50	9.96×10^{-5}	4.54 ± 0.46	4.44 ± 0.46	0.115	4.52 ± 0.43	4.50 ± 0.45	0.715
MCV (fL)	93.68 ± 7.57	92.50 ± 5.49	0.017	91.58 ± 5.05	92.1 ± 5.98	0.524	91.40 ± 4.21	92.27 ± 5.78	0.346
RDW (%)	14.89 ± 7.86	14.66 ± 7.75	0.381	16.05 ± 8.20	15.25 ± 8.10	0.082	14.24 ± 6.46	14.57 ± 6.26	0.568
MCH (pg)	31.02 ± 2.89	30.60 ± 1.91	0.021	29.93 ± 1.52	30.45 ± 2.16	0.083	30.26 ± 1.72	30.54 ± 2.01	0.328
HGB (g/L)	138.17 ± 16.11	141.50 ± 14.46	0.004	135.77 ± 13.62	134.82 ± 14.92	0.622	136.52 ± 11.88	137.06 ± 13.17	0.740
HCT (%)	41.73 ± 4.52	42.80 ± 4.23	0.002	41.54 ± 4.26	40.74 ± 4.04	0.155	41.30 ± 3.51	41.42 ± 3.84	0.809
PLT (10^9/L)	150.67 ± 57.96	187.28 ± 52.47	1.55×10^{-13}	181.88 ± 59.79	195.41 ± 62.78	0.164	187.42 ± 59.68	189.62 ± 56.03	0.793
MPV (fL)	8.98 ± 2.02	8.69 ± 1.77	0.027	9.04 ± 1.67	8.58 ± 1.86	0.043	8.93 ± 1.75	8.68 ± 1.86	0.336
PDW (fL)	16.88 ± 3.05	16.60 ± 2.74	0.149	16.54 ± 2.84	16.49 ± 3.39	0.986	16.48 ± 3.14	16.48 ± 2.64	0.975
PCT (%)	1.35 ± 0.52	1.63 ± 0.50	5.74×10^{-11}	1.62 ± 0.52	1.65 ± 0.55	0.681	1.65 ± 0.51	1.64 ± 0.52	0.831
Hepatic function									
TBIL (umol/L)	19.10 ± 9.34	15.00 ± 5.26	7.25×10^{-11}	15.58 ± 7.19	14.51 ± 5.52	0.228	13.82 ± 5.46	14.70 ± 11.55	0.527
DBIL (umol/L)	6.72 ± 4.69	4.52 ± 1.85	3.92×10^{-17}	4.98 ± 2.76	4.35 ± 1.94	0.036	4.21 ± 1.59	4.78 ± 7.83	0.502
IDBIL (umol/L)	12.38 ± 6.10	10.48 ± 4.13	1.09×10^{-5}	10.62 ± 5.04	10.15 ± 4.45	0.466	9.61 ± 4.61	9.91 ± 4.64	0.651
AST (U/L)	39.99 ± 31.18	24.73 ± 25.51	3.23×10^{-13}	26.72 ± 20.46	24.53 ± 10.44	0.310	25.81 ± 16.46	24.70 ± 14.97	0.637
ALP (U/L)	103.3 ± 40.12	86.88 ± 26.45	1.13×10^{-8}	103.70 ± 45.28	93.48 ± 28.38	0.063	98.43 ± 30.39	86.79 ± 32.84	0.023
ALT (U/L)	36.66 ± 44.85	22.13 ± 13.62	4.43×10^{-12}	25.60 ± 34.73	22.75 ± 18.21	0.461	25.88 ± 21.25	22.18 ± 15.29	0.151
Renal function									
BUN (mmol/L)	5.32 ± 1.60	5.74 ± 1.72	0.003	5.70 ± 1.61	5.51 ± 1.56	0.393	5.58 ± 1.38	5.69 ± 2.06	0.668
SCR (μmol/L)	80.36 ± 16.75	84.68 ± 25.24	0.008	81.66 ± 25.29	81.47 ± 22.75	0.956	82.69 ± 21.35	90.06 ± 69.04	0.323
SUA (μmol/L)	314.19 ± 77.88	329.13 ± 87.24	0.037	310.91 ± 82.73	323.01 ± 87.45	0.352	309.55 ± 78.38	321.66 ± 82.23	0.312

Continuous variables were summarized as mean ± SD and were compared by conditional logistic regression test. Categorical variables were summarized as counts (percentage) and were compared by Cochran-Mantel-Haenszel test.

CHD: coronary heart disease; BMI: body mass index; WHtR: waist-height ratio; SBP: systolic blood pressure; DBP: diastolic blood pressure; TG: total triglyceride; LDL-C: low-density lipoprotein cholesterol; HDL-C: high-density lipoprotein cholesterol; TC: total cholesterol; FBG: fasting blood glucose; WBC: white blood cell counts; LYM: lymphocyte counts; NEUT: neutrophil counts; RBC: red blood cell counts; MCV: mean corpuscular volume; RDW: red cell volume distribution width; MCH: mean corpuscular hemoglobin; HGB: hemoglobin; HCT: hematocrit; PLT: platelet counts; MPV: mean platelet volume; PDW: platelet distribution width; PCT: plateletcrit; TBIL: total bilirubin; DBIL: direct bilirubin; IDBIL: indirect bilirubin; AST, aspartate aminotransferase; ALP: alkaline phosphatase; ALT: alanine aminotransferase; BUN: blood urea nitrogen; SCR: serum creatinine; SUA: serum uric acid.

for HCC, 69.37 (8.02) years and 43% for BTC, and 69.22 (7.09) years and 50% for PC (Table 1). For all three cancers, the age and sex of controls were well matched to the cases, and there was no difference in cigarette smoking and alcohol drinking between cases and controls ($p > 0.05$). HCC cases were more likely to have disease histories of hepatitis ($p = 2.16\times10^{-21}$) and diabetes ($p = 0.008$), and were found to have significant changes in 5 metabolic traits, 10 hematologic indices, 6 hepatic function indices and 3 renal function indices ($p < 0.05$, Table 1). In contrast, relatively few indices showed significant difference between incident cases and matched controls for BTC and PC. We found that a disease history of gallstone ($p = 0.039$), elevated levels of mean platelet volume (MPV, $p = 0.043$), and direct bilirubin (DBIL, $p = 0.036$) were associated with increased risk of BTC, while disease histories of hyperlipidemia ($p = 0.033$) and diabetes ($p = 0.012$), and higher levels of fasting blood glucose (FBG, $p = 3.76 \times 10^{-5}$) and alkaline phosphatase (ALP, $p = 0.023$) were associated with increased risk of PC (Table 1).

We then applied multivariable CLR, including variables significant in the univariable analysis as predictors, to estimate the independent effect size of each factor. We excluded DBIL, platelet counts (PLT), total bilirubin (TBIL), neutrophil counts (NEUT), hemoglobin (HGB), and low-density lipoprotein cholesterol (LDL-C) from the model for HCC due to their high correlation ($r_s > 0.8$) with at least one variable retained in the model (Supplementary Figure S3). We found six variables independently associated with the risk of HCC (Table 2), among which a disease history of hepatitis (OR = 8.51, 95% CI (3.83, 18.91)), alanine aminotransferase (ALT, OR = 2.06, 95% CI (1.39, 3.06)), ALP (OR = 1.47, 95% CI (1.09, 1.98)), and FBG (OR = 1.31, 95% CI (1.03, 1.66)) were positively associated with HCC, while plateletcrit (PCT, OR = 0.57, 95% CI (0.42, 0.78)) and total triglyceride (TG, OR = 0.66, 95% CI (0.46, 0.94)) were negatively associated with HCC. For BTC, a disease history of gallstone (OR = 2.70, 95% CI (1.17, 6.24)) and DBIL (OR = 1.58, 95% CI (1.08, 2.31)) showed independent risk effects. Finally, hyperlipidemia (OR = 2.56, 95% CI (1.12, 5.82)) and FBG (OR = 2.00, 95% CI (1.26, 3.15)) were independently associated with increased risk of PC.

We constructed a CRS for each cancer based on the estimated effects from multivariable CLR models (Table 2). In the nested case-control samples, the CRS achieved the highest AUC in HCC (0.818, 95% CI (0.776, 0.861)), followed by PC (0.691, 95% CI (0.611, 0.770)) and BTC (0.668, 95% CI (0.584, 0.751)) (Figure 1). Considering potential overfitting, we also performed 10-fold CV, which yielded slightly lower AUCs of 0.784 (95% CI (0.775, 0.793)), 0.666 (95% CI (0.651, 0.682)), and 0.648 (95% CI (0.633, 0.662)) for HCC, PC, and BTC, respectively. There were no significant differences between AUCs obtained from the full data set and 10-fold CV ($p > 0.1$, Delong's tests), suggesting little overfitting. If we stratified participants by the median CRS of each cancer, the high-risk groups were found to have significantly higher incidences than the low-risk groups, as shown by the cumulative incidence curves ($p < 0.001$, Wald test from Cox model, Figure 1).

When we considered the whole DFTJ cohort by including all 43,323 controls without matching, the CRS was a significant predictor of all three HBP cancers in models with and without adjustment of age and sex ($p < 0.05$, Supplementary Table S2). Models including age, sex, and CRS achieved the highest AUCs, which were 0.818 (95% CI (0.814, 0.823)), 0.704 (95% CI (0.695, 0.714)), and 0.699 (95% CI (0.689, 0.708)) for HCC, BTC and PC, respectively (Figure 2). Comparing to the reference models with age and sex, adding CRS significantly improved the prediction accuracy for HCC and PC, as supported by NRIs, which were 0.855 (95% CI (0.638, 1.071)) and 0.278 (95% CI (0.084, 0.473)) for HCC, PC, respectively. Similarly, IDIs were significant for HCC (0.018, 95% CI (0.011, 0.026)), but not for PC (0.0003, 95% CI (−0.0002, 0.0009)). It should be noted that, CRS did not improve the prediction of BTC (NRI = −0.0001, 95% CI (−0.0002, 0.0000); IDI = −0.047, 95% CI (−0.189, 0.095)).

We also evaluated the model performance when prioritizing high-risk groups at different sensitivity levels (Table 3). If we wanted to identify half of the HCC incidences (50% sensitivity), 3533 samples would be labeled as positive, leading to a PPV of 2.29% (1.80–2.79%) and a specificity of 92.03% (91.78–92.29%). Detection of BTC and PC incidences was much more difficult. Fixing sensitivity at 50%, the predictive model could only achieve 0.27% (0.17–0.37%) PPV and 76.65% (76.25–77.05%) specificity for BTC, and even lower for PC (0.21%, 95% CI (0.13%, 0.28%)) and PPV (67.63%,95% CI (67.19%, 68.07%) specificity, Table 3). All predictive models achieved more than 99.8% NPVs (Table 3).

Finally, we evaluated the prediction of HCC incidences when adding HBV infection status to the model. Among 25,216 participants with HBV test results, 1709 were HBV positive (6.8%) and 82 were HCC incident cases (0.33%). Among these 82 HCC incident cases, 42 (51.2%) were tested HBV positive at the baseline. Compared to model with age, sex, and HBV infection status, the model with age, sex, and CRS increased the AUC by 0.028 (0.016–0.039) ($p = 1.22 \times 10^{-6}$, Delong's test). By adding HBV infection status to the CRS model, we could further improve the AUC from 0.843 (0.839–0.847) to 0.873 (0.869–0.878) ($p = 1.07 \times 10^{-14}$, Delong's test, Figure 3).

4 | DISCUSSION

HCC, BTC, and PC are the three major HBP cancers, all of which have poor prognosis and high mortality rates. Despite their common need of cost-effective risk prediction models, few studies have simultaneously investigated and compared incident HCC, BTC, and PC in the same cohort. Here, we leveraged the large prospective DFTJ cohort to evaluate the predictive value of 50 regular clinical variables for the six-year (about 70 months) incidences of HCC, BTC, and PC in elderly Chinese. Based on age- and sex-matched nested case-control samples, we identified predictive clinical variables and proposed a clinical risk score (CRS) for each cancer type. The predictive value of each CRS was evaluated in both the nested case-control samples and the full cohort. We found that 26 out of 50 variables were significantly different between incident cases and controls for HCC, 3 variables for BTC and 4 variables for PC. Consequently, while the CRSs for all three cancers are effective to stratify low- and high-risk groups, CRS for HCC achieved the best predictive performance compared to those for BTC and PC.

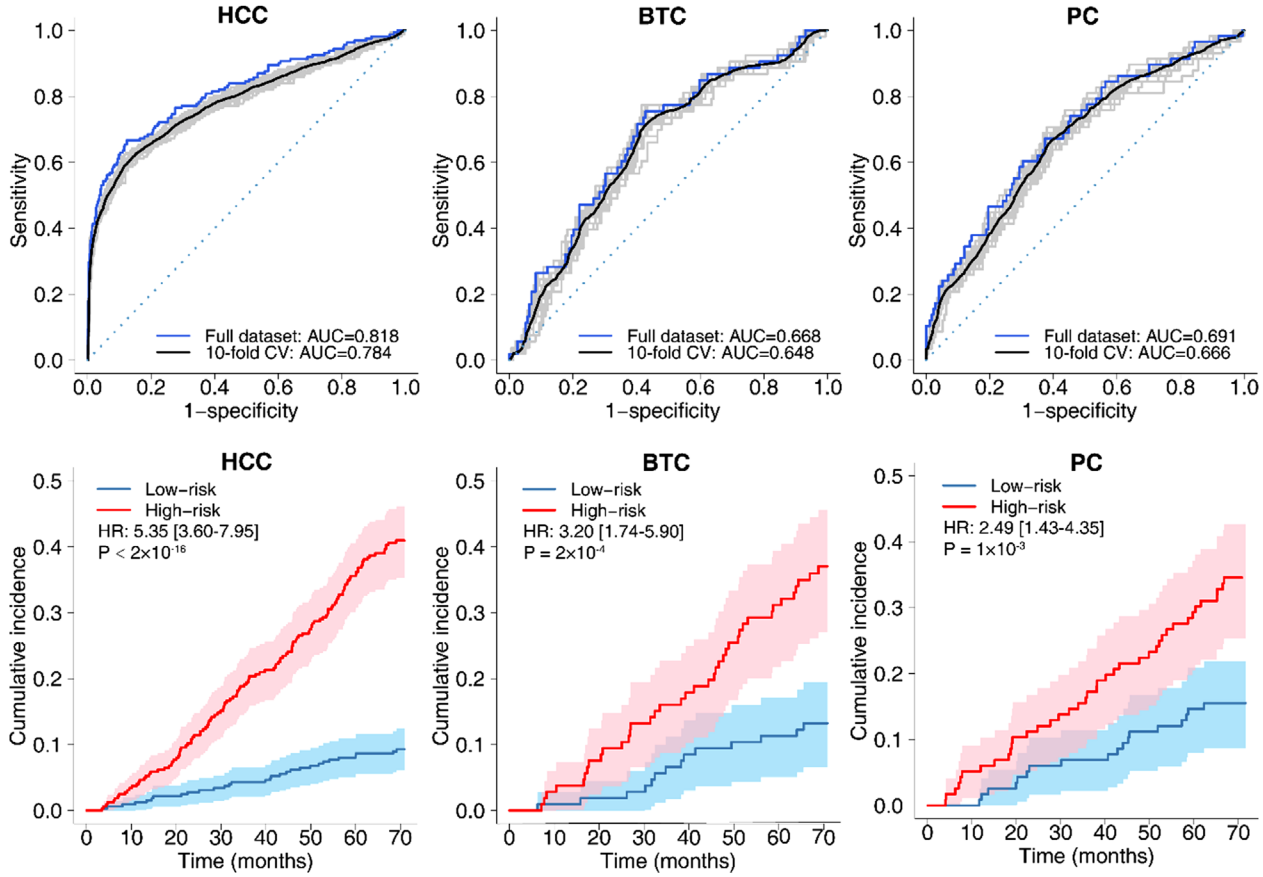

FIGURE 1 Performance of CRS in predicting incident hepatocellular carcinoma, biliary tract cancer, and pancreatic cancer in the nested case-control samples. Top: ROC curves: In each panel, the blue curve was based on clinical risk score derived from the full data set, while gray curves indicated results from 20 repeats of 10-fold cross-validation, of which the mean was indicated by the black curve. Bottom: Cumulative incidence curves: The samples were stratified into high- and low-risk groups of each cancer based on the median CRS. The shaded area reflected 95% confidence interval. HCC: hepatocellular carcinoma; BTC: biliary tract cancer; PC: pancreatic cancer; CRS: clinical risk score.

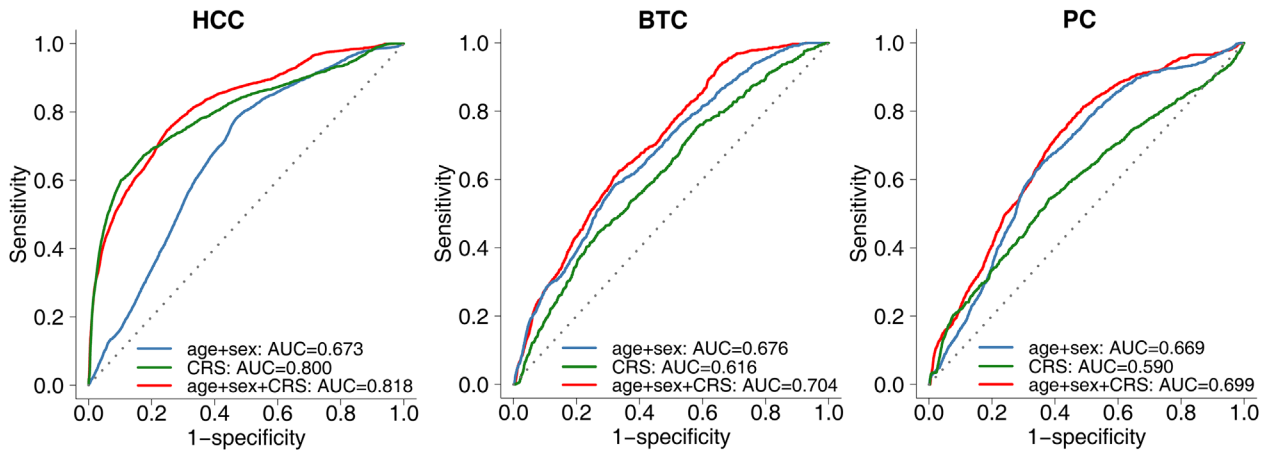

FIGURE 2 ROC curves for predicting incident hepatocellular carcinoma, biliary tract cancer, and pancreatic cancer in the Dongfeng-Tongji cohort. Each model was adjusted for baseline year (2008 versus 2013). HCC: hepatocellular carcinoma; BTC: biliary tract cancer; PC: pancreatic cancer; CRS: clinical risk score.

TABLE 2 Adjusted odds ratios (ORs) of selected variables for hepatocellular carcinoma, biliary tract cancer, and pancreatic cancer.

Variables	β_j	SE	OR	95% CI	*p* Value
Hepatocellular carcinoma					
Diabetes[a]	0.243	0.361	1.275	0.628, 2.587	0.501
Hepatitis[a]	2.141	0.408	8.505	3.825, 18.914	1.52×10^{-7}
DBP	−0.279	0.143	0.757	0.571, 1.002	0.052
TG	−0.422	0.184	0.655	0.457, 0.940	0.022
TC	−0.092	0.144	0.912	0.687, 1.210	0.524
FBG	0.267	0.123	1.306	1.027, 1.662	0.030
WBC	0.152	0.144	1.165	0.877, 1.546	0.291
RBC	0.167	0.481	1.181	0.460, 3.031	0.729
MCV	0.119	0.550	1.127	0.383, 3.314	0.829
MCH	0.105	0.359	1.111	0.550, 2.244	0.770
HCT	−0.396	0.462	0.673	0.272, 1.665	0.392
MPV	0.110	0.108	1.116	0.902, 1.380	0.311
PCT	−0.556	0.153	0.573	0.424, 0.775	2.90×10^{-4}
IDBIL	0.167	0.119	1.182	0.936, 1.493	0.159
AST	−0.065	0.086	0.937	0.792, 1.109	0.450
ALP	0.386	0.152	1.471	1.093, 1.979	0.011
ALT	0.722	0.202	2.059	1.387, 3.059	3.45×10^{-4}
BUN	−0.266	0.154	0.767	0.567, 1.037	0.085
SCR	−0.374	0.265	0.688	0.409, 1.157	0.159
SUA	0.001	0.145	1.001	0.754, 1.329	0.996
Biliary tract cancer					
Gallstone	0.993	0.427	2.700	1.169, 6.237	0.020
MPV	0.265	0.162	1.303	0.949, 1.790	0.102
DBIL	0.458	0.193	1.582	1.083, 2.309	0.018
Pancreatic cancer					
Hyperlipidemia	0.939	0.420	2.556	1.123, 5.820	0.025
Diabetes	0.067	0.488	1.069	0.411, 2.785	0.891
FBG	0.690	0.233	1.995	1.262, 3.151	0.003
ALP	0.240	0.156	1.271	0.937, 1.724	0.123

[a]Binary variables indicating disease histories.
β_j is the regression coefficient of variable *j* in the multivariable CLR model. For quantitative variables, OR was reported in the unit of per SD increment in the exposure.

Liver is the largest and most important digestive gland organ, to which early damage is expected to be reflected in the blood indices. Consistent with this, the most significant biomarkers for HCC include LDL-C, PLT, PCT, and several hepatic function indices. By summing the independent predictive effects across 20 biomarkers, the CRS for HCC achieved an AUC of 0.828 to predict incident HCC in the DFTJ cohort. Among the independent predictors, hepatitis and ALT showed the strongest risk effects, while PCT showed the strongest protective effect. Furthermore, when testing on the subset of samples with information on HBV infection status, the CRS showed better predictive performance than the HBV infection status, while adding the HBV infection status could further increase the AUC from 0.862 to 0.890. These results confirmed the predictive value of transaminases, platelets, histories of hepatitis and HBV infection for HCC in elderly Chinese.[11,15,16] Counterintuitively, we found a protective effect of TG on HCC, despite that TG is a well-known risk factor for many metabolic and cardiovascular diseases. Among previous studies based on European cohorts, Borena et al. found a protective but insignificant effect of TG on primary liver cancer,[35] while Häggström et al. reported an age-dependent effect of TG on HCC in men (protective for those younger than 70 years old).[36] Moreover, there is evidence showing that increased TG storage is protective against fatty-acids-mediated hepatotoxicity.[37] Future replication and experimental studies will be needed to confirm the protective effect of TG on HCC.

The prediction performances of incident BTC and PC were similar. For BTC, we only found three clinical variables to be

TABLE 3 Prediction of hepatocellular carcinoma, biliary tract cancer, and pancreatic cancer incidence in the population at different sensitivity levels.

Variables	SEN%	Cutoff	No. of positive prediction	Youden index	SPE% (95% CI)	PPV% (95% CI)	NPV% (95% CI)
HCC	50	0.0075	3533	0.420	92.03 (91.78–92.29)	2.29 (1.80–2.79)	99.80 (99.75–99.84)
	60	0.0060	6257	0.463	85.78 (85.45–86.11)	1.57 (1.26–1.87)	99.83 (99.79–99.87)
	70	0.0051	8885	0.501	79.75 (79.38–80.13)	1.28 (1.05–1.52)	99.86 (99.82–99.90)
	80	0.0041	13,004	0.505	70.28 (69.85–70.71)	1.00 (0.83–1.17)	99.89 (99.86–99.93)
	90	0.0018	24,839	0.331	43.00 (42.54–43.47)	0.59 (0.49–0.68)	99.91 (99.87–99.96)
BTC	50	0.0015	10,143	0.276	76.65 (76.25–77.05)	0.27 (0.17–0.37)	99.92 (99.89–99.95)
	60	0.0013	12,858	0.308	70.39 (69.97–70.82)	0.25 (0.16–0.34)	99.93 (99.90–99.96)
	70	0.0010	20,220	0.251	53.41 (52.95–53.89)	0.19 (0.13–0.25)	99.94 (99.90–99.97)
	80	0.0008	23,769	0.264	45.23 (44.77–45.70)	0.18 (0.13–0.24)	99.95 (99.92–99.98)
	90	0.0006	29,617	0.242	31.75 (31.31–32.19)	0.17 (0.12–0.21)	99.97 (99.94–100.0)
PC	50	0.0013	14,053	0.1763	67.63 (67.19–68.07)	0.21 (0.13–0.28)	99.90 (99.87–99.94)
	60	0.0012	16,999	0.2119	60.84 (60.38–61.30)	0.21 (0.14–0.27)	99.91 (99.88–99.95)
	70	0.0011	22,113	0.1974	49.05 (48.58–49.52)	0.18 (0.13–0.24)	99.92 (99.88–99.96)
	80	0.0011	27,404	0.1789	36.85 (36.40–37.31)	0.17 (0.12–0.22)	99.93 (99.89–99.97)
	90	0.0010	38,665	0.0225	10.87 (10.58–11.17)	0.14 (0.10–0.17)	99.89 (99.80–99.99)

The prediction was done based on data including all controls from the DFTJ cohort via model with predictive variables being age, sex, and CRS.
HCC: hepatocellular carcinoma; BTC: biliary tract cancer; PC: pancreatic cancer; SEN: sensitivity; SPE: specificity; PPV: positive predictive value; NPV: negative predictive value.

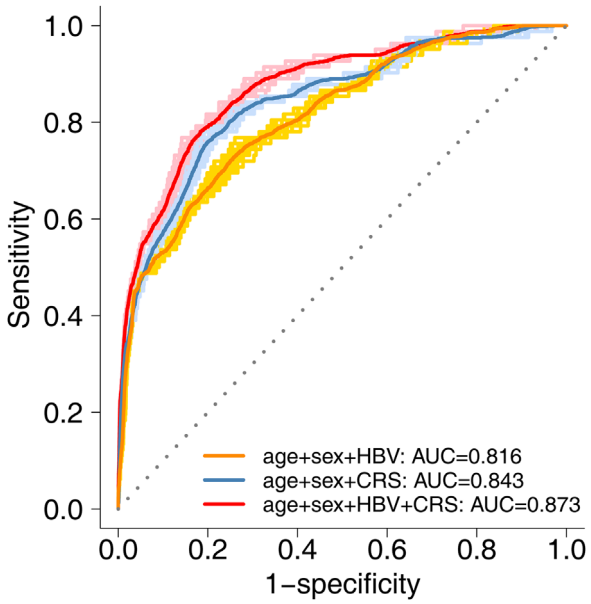

FIGURE 3 ROC curves for predicting incident hepatocellular carcinoma in the samples with HBV infection status from the DFTJ 2013 data set. For each model, the light-colored curves indicate results from 20 repeats of 10-fold cross-validation, of which the mean was indicated by the dark-colored curve. HBV: hepatitis B virus; CRS: clinical risk score.

significantly different between cases and controls. These included gallstone, MPV, and DBIL, among which gallstone and DBIL were independent predictors. A history of gallstone has been recognized as the major risk factor of BTC in previous studies.[7,17] Bilirubin is transformed to DBIL in the liver before being released to the small intestine through the biliary tract.[38] When biliary obstruction occurs, DBIL will be released into the serum instead.[39] Thus, an increased serum level of DBIL is indicative of lesion in the biliary tract, which might lead to the development of BTC. Likewise, we only found four clinical variables (disease histories of hyperlipidemia and diabetes, FBG, and ALP) to be different between incident cases and controls of PC, consistent with the key function of pancreas being to produce enzymes to digest sugars, fats, and starches. Both diabetes history and newly onset diabetes are known risk factors of PC.[3] The CRS, together with age and sex, could achieve ~0.7 AUC for predicting the 6-year incidence of BTC or PC in this elderly Chinese cohort. However, due to their low incidence rates, screening for BTC and PC in the cohort with 50% sensitivity had PPV as low as 0.3%, which would further decrease if screening requires higher sensitivity. Thus, future research to identify novel predictive biomarkers is urgently needed to build cost-effective models for risk stratification of BTC and PC in the general population.

In summary, we have provided a comprehensive evaluation of 50 clinical variables, covering health history and blood test results from regular physical examinations, to predict incident HCC, BTC, and PC in a large prospective cohort of elderly Chinese. We have replicated several key findings from previous studies based on different ethnic or age groups and proposed clinical risk scores, which were effective to identify high-risk groups of HBP cancers. There are several limitations in the present study. First, because of their low incident rates, we have relatively small numbers of cases for BTC and PC, which could lead to low statistical power. Second, due to the small sample size, we did not

distinguish subtypes of BTC, including intrahepatic and extrahepatic cholangiocarcinoma and gallbladder cancer, which might have different clinical features. Third, our model was based on retired employees which were all middle-aged and older adults. Validation in external cohorts is needed to confirm the general applicability of our predictive models in other populations and age groups. Despite of the limitations, we have shown that models based on regular clinical variables are reasonably effective to identify high-risk group of incident HCC, BTC, and PC in the general population. Our findings provide useful information for future studies to refine screening and preventive strategies of these deadly cancers.

ORCID

Chaolong Wang https://orcid.org/0000-0003-3945-1012

REFERENCES

1. Everhart JE, Ruhl CE. Burden of digestive diseases in the United States Part III: Liver, biliary tract, and pancreas. *Gastroenterology*. 2009;136(4):1134-44.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.
3. Mizrahi JD, Surana R, Valle JW, Shroff RT. Pancreatic cancer. *Lancet*. 2020;395(10242):2008-20.
4. Valle JW, Kelley RK, Nervi B, Oh D-Y, Zhu AX. Biliary tract cancer. *Lancet*. 2021;397(10272):428-44.
5. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin*. 2021;71(1):7-33.
6. Islami F, Ward EM, Sung H, Cronin KA, Tangka FKL, Sherman RL, et al. Annual Report to the Nation on the Status of Cancer, Part 1: National Cancer Statistics. *J Natl Cancer Inst*. 2021;113(12):1648-69.
7. Sharma A, Sharma KL, Gupta A, Yadav A, Kumar A. Gallbladder cancer epidemiology, pathogenesis and molecular genetics: recent update. *World J Gastroenterol*. 2017;23(22):3978-98.
8. Llovet JM, Kelley RK, Villanueva A, et al. Hepatocellular carcinoma. *Nat Rev Dis Primers*. 2021;7:6.
9. Bailey A, Shah SA. Screening high risk populations for cancer: hepatobiliary. *J Surg Oncol*. 2019;120(5):847-50.
10. Li TC, Li CI, Liu CS, et al. Risk score system for the prediction of hepatocellular carcinoma in patients with type 2 diabetes: Taiwan Diabetes Study. *Semin Oncol*. 2018;45(5-6):264-74.
11. Yu CX, Song C, Lv J, et al. Prediction and clinical utility of a liver cancer risk model in Chinese adults: a prospective cohort study of 0.5 million people. *Int J Cancer*. 2021;148(12):2924-34.
12. Wong VW-S, Chan SL, Mo F, et al. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers. *J Clin Oncol*. 2010;28(10):1660-5.
13. Chun HS, Kim BK, Park JY, et al. Design and validation of risk prediction model for hepatocellular carcinoma development after sustained virological response in patients with chronic hepatitis C. *Eur J Gastroenterol Hepatol*. 2020;32(3):378-85.
14. Yang HI, Sherman M, Su J, et al. Nomograms for risk of hepatocellular carcinoma in patients with chronic hepatitis B virus infection. *J Clin Oncol*. 2010;28(14):2437-44.
15. Michikawa T, Inoue M, Sawada N, et al. Development of a prediction model for 10-year risk of hepatocellular carcinoma in middle-aged Japanese: the Japan Public Health Center-based Prospective Study Cohort II. *Prev Med*. 2012;55(2):137-43.
16. Wen CP, Lin J, Yang YC, Tsai MK, et al. Hepatocellular carcinoma risk prediction model for the general population: the predictive power of transaminases. *J Natl Cancer I*. 2012;104(20):1599-611.
17. Hsing AW, Bai Y, Andreotti G, et al. Family history of gallstones and the risk of biliary tract cancer and gallstones: a population-based study in Shanghai, China. *Int J Cancer*. 2007;121(4):832-8.
18. Jain K, Sreenivas V, Velpandian T, Kapil U, Garg PK. Risk factors for gallbladder cancer: a case-control study. *Int J Cancer*. 2013;132(7):1660-6.
19. Boursi B, Finkelman B, Giantonio BJ, Haynes K, Rustgi AK, Rhim AD, et al. A clinical prediction model to assess risk for pancreatic cancer among patients with new-onset diabetes. *Gastroenterology*. 2017;152(4):840-50.
20. Sharma A, Kandlakunta H, Nagpal SJS, et al. Model to determine risk of pancreatic cancer in patients with new-onset diabetes. *Gastroenterology*. 2018;155(3):730-9.
21. Lu C, Xu CF, Wan XY, Zhu HT, Yu CH, Li YM. Screening for pancreatic cancer in familial high-risk individuals: a systematic review. *World J Gastroenterol*. 2015;21(28):8678-86.
22. Chang MC, Wong JM, Chang YT. Screening and early detection of pamcreatic cancer in high risk population. *World J Gastroenterol*. 2014;20(9):2358-64.
23. Force USPST, Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, et al. Screening for pancreatic cancer: US Preventive Services Task Force Reaffirmation Recommendation Statement. *JAMA*. 2019;322(5):438-44.
24. Randi G, Franceschi S, La Vecchia C. Gallbladder cancer worldwide: geographical distribution and risk factors. *Int J Cancer*. 2006;118(7):1591-602.
25. Bosetti C, Bertuccio P, Negri E, La Vecchia C, Zeegers MP, Boffetta P. Pancreatic cancer: overview of descriptive epidemiology. *Mol Carcinog*. 2012;51(1):3-13.
26. Wang F, Zhu J, Yao P, Li X, He M, Liu Y, et al. Cohort profile: the Dongfeng-Tongji cohort study of retired workers. *Int J Epidemiol*. 2013;42(3):731-40.
27. Li P, Stuart EA, Allison DB. Multiple imputation: a flexible tool for handling missing data. *JAMA*. 2015;314(18):1966-7.
28. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):1-67.
29. Hayati Rezvan P, Lee KJ, Simpson JA. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Med Res Methodol*. 2015;15:30.
30. Gail MH, Lubin JH, Rubinstein LV. Likelihood calculations for matched case-control studies and survival studies with tied death times. *Biometrika*. 1981;68(3):703-7.
31. Agresti A. *Categorical Data Analysis (second edition)*. New York: Wiley, 2002.
32. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S plus to analyze and compare ROC curves. *BMC Bioinform*. 2011;12:77.
33. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67(1):1-48.
34. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28(5):1-26.
35. Borena W, Strohmaier S, Lukanova A, et al. Metabolic risk factors and primary liver cancer in a prospective study of 578,700 adults. *Int J Cancer*. 2012;131(1):193-200.
36. Häggström C, Jonsson H, Bjørge T, et al. Linear age-course effects on the associations between body mass index, triglycerides, and female breast and male liver cancer risk: an internal replication study of 800,000 individuals. *Int J Cancer*. 2020;146(1):58-67.
37. Alves-Bezerra M, Cohen DE. Triglyceride metabolism in the liver. *Compr Physiol*. 2017;8:1-8.

38. Hofmann AF. *Bile Composition*. Johnson LR, editor. New York: Elsevier; 2004 2004/01/01/. 176–84 p.
39. Zollner G, Trauner M. Mechanisms of cholestasis. *Clin Liver Dis*. 2008;12(1):1-26.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lyu J, Guan X, Zhou Y, Guo H, Cheng S, Wang C. Risk prediction of hepatobiliary and pancreatic cancers in elderly Chinese: The Dongfeng-Tongji cohort. *J Evid Based Med*. 2023;16:39–49.
https://doi.org/10.1111/jebm.12516