

Development of a predictive model for common bile duct stones in patients with clinical suspicion of choledocholithiasis: a single-center cohort study

Suppadech Tunruttanakul (✉ suppadech_t@spr.go.th)

Sawanpracharak Hospital <https://orcid.org/0000-0002-5418-682X>

Borirak Chareonsil

Sawanpracharak Hospital

Kotchakorn Verasmith

Sawanpracharak Hospital

Jayanton Patumanond

Chiang Mai University Faculty of Medicine

Chatchai Mingmalairak

Thammasat University Hospital

Research Article

Keywords: Choledocholithiasis, Clinical decision rules, Risk assessment

Posted Date: May 2nd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1520382/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Current choledocholithiasis guidelines heavily focus on patients with low or no risk; they may be inappropriate for populations with high rates of choledocholithiasis. We aimed to develop a predictive scoring model for choledocholithiasis in patients with relevant clinical manifestations.

Methods

Design: A multivariable predictive model development study based on a retrospective cohort of patients with clinical suspicion of choledocholithiasis.

Setting: A single 700-bed tertiary public hospital.

Participants: Patients who completed three reference tests (endoscopic retrograde cholangiography, magnetic resonance cholangiopancreatography, and intraoperative cholangiography) from January 2019 to June 2021.

Statistical analysis: The model was developed using logistic regression analysis. Predictor selection was conducted using a backward stepwise approach. Three risk groups were considered. Model performance was evaluated by area under the receiver-operating characteristic curve, calibration, classification measures, and decision curve analyses.

Results

Six hundred twenty-one patients were included; the choledocholithiasis prevalence was 59.9%. The predictors were age > 55 years, pancreatitis, cholangitis, cirrhosis, alkaline phosphatase level 125–250 or > 250 U/L, total bilirubin level > 4 mg/dL, common bile duct size > 6 mm, and common bile duct stone detection. Pancreatitis and cirrhosis each had a negative score. The sum of scores was -4.5 to 28.5. Patients were categorized into three risk groups: low-intermediate (score ≤ 5), intermediate (score 5–15), and high (score ≥ 15). Positive likelihood ratios were 0.16 and 3.47 in the low-intermediate and high risk groups, respectively.

The model had an area under the receiver-operating characteristic curve of 0.80 (95% confidence interval: 0.76, 0.83) and was well-calibrated; it exhibited better statistical suitability to the high-prevalence population, compared to current guidelines.

Conclusions

Our scoring model had good predictive ability for choledocholithiasis in patients with relevant clinical manifestations. Consideration of other factors is necessary for clinical application, particularly regarding the availability of expert physicians and specialized equipment.

Background

Cholelithiasis or common bile duct (CBD) stone is characterized by the presence of stones in the bile duct. The most common form is secondary CBD stone: stones originate in the gallbladder, then migrate to the bile duct (1). Management usually includes cholecystectomy (gallbladder removal) (2); this procedure is currently performed using a laparoscopic approach. CBD stones are suspected in symptomatic gallstone patients on the basis of relevant clinical manifestations, abnormal liver function test (LFT) results, or abnormal relevant imaging parameters (3). CBD stones can cause severe lethal complications (4); the current recommendation is that all detected stones should be treated (5). However, it is challenging to select the optimal investigation approach from the available options.

For example, endoscopic retrograde cholangiography (ERC) has therapeutic potential but can cause morbidity or (rarely) mortality (6). In contrast, intraoperative cholangiography (IOC) enables single-stage management (i.e., exploration combined with cholecystectomy) (7). Nevertheless, experienced surgeons and more specialized equipment are required for the treatment of CBD stones, particularly in the laparoscopic era (8). In this context, guidelines, recommendations, and scoring systems have been constructed (5, 9–13); however, such resources generally were not designed exclusively for patients with suspected CBD stones (5, 9, 10), and they have questionable relevance in high-prevalence populations (14). Notably, published scoring systems are not widely used (11–13). Therefore, this study was performed to develop a predictive model for CBD stones in patients with relevant clinical manifestations. We also aimed to build a practical model that complied with the TRIPOD guideline (15) and could be easily used in clinical practice.

Methods

Design and setting

This multivariable predictive model development study used data from a retrospective observational cohort of patients with suspected CBD stones. All patients were treated in Sawanpracharak Hospital (Thailand), a regional 700-bed tertiary public hospital. The patients in this study comprised both local and referral cases. All data were acquired from the hospital information system.

Participants

This study included patients who completed three main reference tests (ERC, IOC or operative bile duct exploration, and magnetic resonance cholangiopancreatography [MRCP]) from January 2019 to June 2021. All tests are considered standard for CBD stone diagnosis (4). The inclusion criteria for suspected CBD stones were:

- - symptomatic gallstone or cholecystitis with abnormal LFT results, primary imaging findings indicative of dilated bile duct, or presence of CBD stone
- - gallstone with jaundice

- - gallstone pancreatitis
- - cholangitis

Standard diagnostic guidelines were used to confirm the diagnosis of gallstone pancreatitis, cholecystitis, and cholangitis (16–18).

The exclusion criteria were:

- - previous biliary tract intervention (surgical or endoscopic)
- - suspected malignancy: painless obstructive jaundice (bilirubin > 5.85 mg/dL) with anorexia and weight loss, along with imaging findings indicative of bile duct dilatation without stones (19, 20). Patients were excluded if initial analysis suggested malignancy but later studies revealed CBD stones alone.

Predictors and outcome

Potential CBD stone predictors and interacting variables were identified in accordance with previous literature (3, 9, 21): patient age, sex, clinical manifestations, status-post (s/p) cholecystectomy, cirrhosis status, results of LFTs (levels of serum glutamic oxaloacetic transaminase [SGOT], serum glutamic pyruvic transaminase [SGPT], alkaline phosphatase [ALP], and total bilirubin [TB]), and relevant imaging findings. Cirrhosis was defined according to known clinical history or imaging-confirmed morphological liver cirrhosis. Relevant imaging findings were CBD size (in mm) and presence of CBD stones. Exploratory imaging comprised abdominal ultrasonography or computed tomography (CT) scans. Because we aimed to create a practical model, we categorized some predictors in accordance with the approaches in widely used CBD stone guidelines and meta-analyses (3, 21). Categorized variables were age, ALP level, TB level, and CBD size. The binary predictors were: age ≤ 55 vs. > 55 years and CBD size ≤ 6 vs. > 6 mm. The ternary predictors were: ALP level < 125 , $125\text{--}250$ (twofold greater than the normal limit), and > 250 U/L; TB level < 1.8 , $1.8\text{--}4$, and > 4 mg/dL. The CBD size was acquired from medical records (if available) or quantified by a participating radiologist using the hospital's picture archiving and communication system. The measurement location was immediately distal to the porta hepatis or mid-CBD. Bile duct dilatation status was not used to avoid ambiguous phrasing (e.g., minimal or borderline dilatation) and uncertain cut-off diameter.

Flow and timing for the determinant variable were as follows. Age was measured at the reference test date. In the hospital, a repeat LFT protocol is used prior to reference tests. However, physicians occasionally choose not to implement this protocol. Data for more than 7 days of LFTs were excluded. No repeat imaging protocol was established, although some physicians chose to perform repeat imaging. The most recent results were used for analysis.

The outcome was the presence of CBD stone according to the results of reference tests. The tests were chosen according to the attending physician's preference. CBD stones were considered "Present" (detected) if visualized in the endoscopic or operative field in the initial or subsequent investigational

session. If CBD stones were not visible (e.g., fluoroscopy or radiography analyses showed filling defects and patients were lost to follow-up [FU]), images were reviewed by either two endoscopists or one endoscopist and one radiologist. CBD stones were considered “Absent” (not detected) if the reference tests did not detect CBD stones during at least 5–6 months of FU to evaluate symptoms and LFT results; imaging findings were evaluated if available. Patients with fewer than 5–6 months of FU and patients who were lost to FU were contacted by phone to check for symptom persistence or therapeutic management in other hospitals. Negative responses to both questions were necessary for a CBD stone to be considered “Absent.” CBD stones were also considered “Absent” if patients died or could not be contacted. If patients underwent a repeat examination using one of the reference tests, the 5–6-month FU assessment was not required. Inconclusive outcomes were excluded.

Sample size and missing data

The sample size of 536 patients was established on the basis of LFT results and imaging parameters, using 90% statistical power and a two-sided alpha level of 0.05. The data used for calculation were collected from the historical records of 50 patients; the CBD stone prevalence was 65%.

Because of the study protocol, missing data solely involved imaging parameters (CBD size > 6 mm and presence of CBD stone); both parameters were binary. Concerning practical implications, missing data were managed by the mean-imputation method; each missing value was changed to 0.5.

Statistical analysis and model development

In univariable descriptive analysis, Fisher’s exact test was used for categorical data; the *t*-test or the Mann–Whitney U test were used for continuous data. Multivariable logistic regression analysis was the primary model development analytic method. Predictors were selected based on a backward stepwise approach. Predictors from the univariable analysis with *p*-values ≤ 0.20 were entered in logistic regression; they were eliminated when the multivariable *p*-value exceeded 0.05 or the odds ratio approached 1.0. Clinical relevance was also considered during the predictor selection process.

Score derivation and validation

Logit coefficient values of parameters remaining after selection were used to construct the score-based prediction model. The scores were derived from dividing the logit coefficients of all predictors by the lowest value. Therefore, the lowest logit coefficient was given a score of one. Divided scores were rounded to the nearest 0.5 or whole number. The sum of the total score for each patient was used to assess the model’s ability to predict CBD stone status. The model performance was evaluated by discriminative ability in terms of the area under the receiver-operating characteristic curve (AUC) (concordance index) and classification measures (e.g., sensitivity and specificity). Calibration (i.e., the relationship between predicted and observed risk) was performed by Hosmer–Lemeshow goodness-of-fit statistics and construction of a calibration plot. The ability to predict clinical outcomes was assessed using decision curve analysis (22). The model net benefit was calculated, plotted, and presented using

the net benefit curves of two strategies: treat all (treatment for all patients) and treat none (treatment for no patients). Bootstrap re-sampling procedures were used for internal validation of the scoring model.

For clinical applications, cut-off considerations were intended to guide clinicians in the selection of investigations and treatments. Currently, there is no optimal CBD stone threshold probability to suggest treatment modalities (5). We created a cut-off point by conducting a short survey and analyzing the classification properties for 10% increments of the model-predicted CBD stone probability.

The TRIPOD statement (15) suggests comparisons to existing models. However, to our knowledge, acceptable CBD stone scoring models are unavailable. Thus, we compared the proposed model with two widely used guidelines: the American Society of Gastrointestinal Endoscopy (ASGE) 2019 (revised version) guidelines (10) and the European Society of Gastrointestinal Endoscopy (ESGE) guidelines (5). The guidelines-predicted CBD stone probabilities were calculated using logistic regression analysis to compare AUC and decision curves.

Finally, sensitivity analysis was conducted to investigate outcome variability according to alteration of determinants.

Statistical significance was set at $p < 0.05$. All statistical analyses were conducted using STATA software, version 17 (StataCorp, College Station, TX, USA).

Results

Participants

In total, 1,185 patients were included in the initial review; 564 were excluded because they met the exclusion criteria, were missing large amounts of data, had duplicate records, had an inconclusive outcome, and/or had no pre-test LFTs. In total, 621 patients were included in model construction and analysis. The participant flow is illustrated in Fig. 1. The CBD stone prevalence was 59.9% (372 patients).

The distributions of variables between CBD stone groups are shown in Table 1. Most patients were elderly women who presented with cholangitis. The most common reference test was ERC (82.9%, 515 patients); IOC and MRCP were performed in 8.1% (50 patients) and 9.0% (56 patients) of the patients, respectively. The median interval between basic imaging and reference tests was 8 days (interquartile range: 2–25 days). Ultrasonography was the main primary imaging modality (approximately 75.2% of patients), while CT scan was performed in 24.8% of patients.

Table 1
Distribution of variables between groups according to CBD stone status

Predictors	CBD stone		p-value
	Present (n = 372)	Absent (n = 249)	
Mean age (\pm SD)	65.3 (17.3)	59.3 (16.1)	< 0.01
Age > 55 years, n (%)	274 (73.7)	158 (63.5)	< 0.01
Female, n (%)	221 (59.4)	159 (63.9)	0.28
Clinical manifestations, n (%)			
Abdominal pain	87 (23.4)	61 (24.5)	0.77
Pancreatitis	23 (6.2)	52 (20.9)	< 0.01
Jaundice	59 (15.9)	50 (20.1)	0.20
Cholecystitis	21 (5.7)	14 (5.6)	1.00
Cholangitis	182 (48.9)	72 (28.9)	< 0.01
Median days from clinical to reference test (IQR)	24 (9, 38.5)	26 (12, 41)	0.16
Clinical \leq 14 days ^a , n (%)	124 (33.3)	72 (28.9)	0.25
s/p Cholecystectomy, n (%)	41 (11.0)	20 (8.1)	0.27
Cirrhosis, n (%)	11 (3.0)	17 (6.8)	0.03
Median LFT results (IQR)			
SGOT (U/L)	50 (12, 606)	30 (12, 418)	< 0.01
SGPT (U/L)	48 (5, 794)	28 (5, 691)	< 0.01
ALP (U/L)	184.5 (51, 1117)	107 (41, 795)	< 0.01
TB (mg/dL)	1.23 (0.22, 22.97)	0.75 (0.22, 10.02)	< 0.01
Categorized LFT results, n (%)			
ALP 125–250 U/L	95 (25.5)	73 (29.3)	0.05
ALP > 250 U/L	148 (39.8)	30 (12.1)	< 0.01

^a Clinical \leq 14 days = Interval from clinical presentation to reference tests within 14 days

ALP alkaline phosphatase, CBD common bile duct, CT computed tomography, IQR interquartile range, LFT liver function test, SGOT serum glutamic oxaloacetic transaminase, SGPT serum glutamic pyruvic transaminase, s/p status-post, SD standard deviation, TB total bilirubin

Predictors	CBD stone		<i>p</i> -value
	Present (n = 372)	Absent (n = 249)	
TB 1.8–4 mg/dL	66 (17.7)	35 (14.1)	0.03
TB > 4 mg/dL	82 (22.0)	20 (8.0)	< 0.01
Imaging characteristics, n (%)			
CT scan	89 (24.1)	64 (25.9)	0.64
Presence of CBD stone	227 (61.0)	70 (28.1)	< 0.01
CBD size (mean [± SD])	12.2 (4.9)	8.8 (4.0)	< 0.01
CBD dilatation (> 6 mm)	336 (90.3)	160 (64.3)	< 0.01
^a Clinical ≤ 14 days = Interval from clinical presentation to reference tests within 14 days			
ALP alkaline phosphatase, CBD common bile duct, CT computed tomography, IQR interquartile range, LFT liver function test, SGOT serum glutamic oxaloacetic transaminase, SGPT serum glutamic pyruvic transaminase, s/p status-post, SD standard deviation, TB total bilirubin			

In the CBD stone “Present” group, three (0.8%) patients had benign bile duct stricture and eight (2.2%) patients had cancer. We included these patients in the CBD stone “Present” group during analysis because both conditions mostly required ERC; this situation can occur in clinical practice. In the CBD stone “Absent” group, 71 (28.5%) patients had inadequate FU; of these patients, 53 (21.3%) were contacted via telephone, six (2.4%) died, and 12 (4.8%) were lost to FU.

Missing values were identified regarding the imaging parameters of 14 (2.3%) patients; these missing values were caused by limited ultrasonographic examination related to the patient’s physical characteristics or presence of intestinal gas. Because both variables (CBD size > 6 mm and presence of CBD stone) were binary, we replaced any missing values with 0.5.

Model development and specification

Univariable analysis revealed potential predictors (Table 1). Among the significant differences, pancreatitis and cirrhosis were less frequent in the CBD stone “Present” group. The selection process removed the following variables from the scoring model: jaundice, SGPT, and SGOT. Cholangitis was identified as a non-significant predictor in multivariable analysis. However, it is a strong predictor in published guidelines (5, 10), and its *p*-value was near 0.05 (i.e., 0.14); thus, we retained cholangitis in the model. TB level 1.8–4 mg/dL was removed, although it is the second level of the significant ternary predictor TB. Its coefficient was near 0 (0.03), while its *p*-value was 0.92. Because the use of a coefficient near 0 as a denominator would cause extremely high score values, this predictor was excluded from the model.

A simplified (parsimonious) model is presented in Table 2. The predictors used in model construction were age > 55 years, pancreatitis, cholangitis, cirrhosis, ALP level 125–250 and > 250 U/L, TB level > 4 mg/dL, CBD size > 6 mm, and presence of CBD stone. ALP level 125–250 U/L had the lowest coefficient and served as the denominator. The item score ranged from – 5.5 for cirrhosis to 6.5 for ALP > 250 U/L. Pancreatitis and cirrhosis each had a negative score. The sum of scores was – 4.5 to 28.5. The mean score was significantly higher in the CBD stone “Present” group than in the CBD stone “Absent” group (mean ± standard deviation: 15.6 ± 6.2 vs. 8.3 ± 6.2, $p < 0.01$).

Table 2
Simplified (parsimonious) modeling with predictor odds ratios, β coefficients, and adjusted scores

Predictors	Odds ratios	95% CI	<i>p</i> -value	β	Item score
Intercept				-1.76	
Age > 55 years	1.84	1.21, 2.80	< 0.01	0.61	3.5
Pancreatitis	0.52	0.28, 0.96	0.04	-0.65	-3.5
Cholangitis	1.39	0.93, 2.08	0.11	0.33	2
Cirrhosis	0.36	0.15, 0.84	0.02	-1.03	-5.5
ALP (U/L)					
125–250	1.20	0.78, 1.86	0.41	0.18	1
> 250	3.35	2.02, 5.55	< 0.01	1.21	6.5
TB > 4 mg/dL	2.75	1.50, 5.05	< 0.01	1.01	5.5
CBD size > 6 mm	2.75	1.64, 4.60	< 0.01	1.01	5.5
CBD stone detected	2.61	1.76, 3.87	< 0.01	0.96	5.5
ALP alkaline phosphatase, β beta-coefficient, CBD common bile duct, CI confidence interval, TB total bilirubin					

Regarding risk-group or cut-off classification, a short survey was administered to gastroenterologists and other surgeons (n = 30) to identify the expected threshold probabilities for ERC and IOC. For the question regarding the expected CBD stone probability threshold for consideration of ERC, the responses were generally equally distributed: approximately 30% to > 50% (consistent with the ASGE suggestion [10]), 70–80%, and 90–100%. For IOC, the expected threshold probability was also generally equally distributed: < 10%, 20–30%, and > 50%. We presumed that these ranges of expected threshold probabilities were secondary to physician experience and the availability of equipment in a particular facility. Generally, physicians required high CBD stone probability for consideration of ERC. More available investigational options may be related to the higher expected probability. In contrast, although

the expected CBD stone probability could be high for IOC, there remained a large number of physicians who were unwilling to detect CBD stones using this method (presumably because of limited resources). While considering the survey results, we conducted another method that involved the separation of data into 10% increments of model-predicted CBD stone probability (shown in Supplemental Table 1) and calculating their diagnostic properties. The potential higher probability cut-offs were 70%, 80%, and 90%; these cut-off values were decided for ERC. All candidates had high specificities, ranging from 80.7–97.2%. However, the sensitivities were poor (sensitivities for 80% and 90% cut-off: 41.7% and 24.5%, respectively) and the numbers of patients who would benefit from the 80% and 90% probability cut-offs were low (178 and 98 patients above cut-off level, respectively). We used a 70% probability cut-off because it had optimal diagnostic properties and a reasonable number of patients above the cut-off level (287 patients). The potential lower probability cut-offs were 10%, 20%, and 30%; these cut-off values were decided for IOC. All candidates had greater than 90% sensitivity, despite poor specificity (5.6–27.3%). Their likelihood ratios were also generally similar. Because all three cut-offs exhibited comparable diagnostic properties, we used a 30% cut-off because it had the highest number of patients who would benefit from the cut-off level (numbers of patients below cut-off level for 10%, 20%, and 30% probability cut-off values: 15, 48, and 84, respectively). However, because the lower cut-off had up to 30% CBD stone probability, which is considerable, we designated this group as the low-intermediate group. The three risk groups were low-intermediate, intermediate, and high; their respective threshold probabilities were $\leq 30\%$, 30–70%, and $\geq 70\%$. The respective cut-off scores were ≤ 5 , 5–15, and ≥ 15 (for easier application in clinical practice, the ≥ 15 value is approximated from the ≥ 14.5 score for the $\geq 70\%$ cut-off level). The risk-group properties are shown in Table 3. Overall, 84 (13.5%), 277 (44.6%), and 260 (41.9%) patients were categorized into low-intermediate, intermediate, and high risk groups, respectively. The low-intermediate risk classification had high sensitivity (95.7%; 95% confidence interval (CI): 93.1%, 97.5%) but poor specificity (27.3%; 95% CI: 21.9%, 33.3%), while the high risk classification had low sensitivity (58.6%; 95% CI: 53.4%, 63.7%) but high specificity (83.1%; 95% CI: 77.9%, 87.6%). The intermediate-risk classification included equal numbers of CBD stone “Present” and “Absent” patients. However, this classification tended to predict a CBD stone “Absent” status (positive likelihood ratio [LHR+]: 0.66; 95% CI: 0.56, 0.79; $p < 0.01$).

Table 3
Scoring model characteristics and diagnostic properties among the three risk groups

Risk groups	Score	Prevalence (%)	CBD stone, n (%)		LHR+ (95% CI)	p-value
			Present (n = 372)	Absent (n = 249)		
Low-intermediate	≤ 5	13.5	16 (19.0)	68 (81.0)	0.16 (0.09, 0.27)	< 0.01
Intermediate	5–15	44.6	138 (49.8)	139 (50.2)	0.66 (0.56, 0.79)	< 0.01
High	≥ 15	41.9	218 (83.9)	42 (16.1)	3.47 (2.60, 4.64)	< 0.01
Classification properties (95% confidence interval)						
		Sensitivity	Specificity		PPV	NPV
Low-intermediate		95.7 (93.1, 97.5)	27.3 (21.9, 33.3)		66.3 (62.1, 70.3)	81.0 (70.9, 88.7)
Intermediate		37.1 (32.2, 42.2)	44.2 (37.9, 50.6)		49.8 (43.8, 55.9)	32.0 (27.1, 37.2)
High		58.6 (53.4, 63.7)	83.1 (77.9, 87.6)		83.8 (78.8, 88.1)	57.3 (52.1, 62.5)
CBD common bile duct, LHR + positive likelihood ratio, NPV negative predictive value, PPV positive predictive value						

Model performance

As shown in Fig. 2, the overall model discriminative property in terms of AUC was 0.80 (95% CI: 0.76, 0.83). Both calibration methods, the calibration plot (Fig. 3) and the Hosmer–Lemeshow goodness-of-fit statistics, showed a good or close correlation between the scoring model-predicted risk vs. observed risk of CBD stones. The well-calibrated plot, interpreted by the locally weighted scatterplot smoothing line slope, was consistently within 95% CI of the reference line. Hosmer–Lemeshow goodness-of-fit statistics showed a non-significant difference ($p = 1.00$), confirming the correlation

The risk curve (Fig. 4) depicts the three risk-group classifications as vertical dashed lines. The predicted risk of CBD stone increased (y-axis) in a manner that corresponded to the increased in our proposed score (x-axis). The circle size indicates the proportion of patients in each circular area.

The internally validated AUC of the scoring model slightly decreased to 0.76 (95% CI: 0.72, 0.81).

Clinical usefulness was determined by concurrent decision curve analysis with comparison to current CBD stone guidelines from the ASGE and the ESGE. Figure 5 shows a comparative AUC and decision

curve between the proposed scoring model and guidelines from the ASGE and ESGE. Decision curve analysis showed that the scoring model had a clinically beneficial outcome, compared to the treat all curve; this was indicated by the model net benefit curve above the treat all curve. The scoring model's net benefit was also superior to the net benefit of each set of guidelines. The model's receiver operating characteristic curve was closer to the graph's left upper corner, reflecting greater discriminative performance. Moreover, the scoring model's AUC was 0.80 (95% CI: 0.76, 0.83); this was significantly superior to the ASGE guidelines (AUC: 0.67; 95% CI: 0.63, 0.71; $p < 0.01$) and the ESGE guidelines (AUC: 0.67; 95% CI: 0.63, 0.71; $p < 0.01$).

Sensitivity analysis was conducted to test the model performance robustness after modification of variables that could affect the outcome. By removing all missing values (complete case analysis, $n = 607$), the AUC was 0.79 (95% CI: 0.75, 0.83). Upon removal of patients with benign bile duct stricture or malignancy from the CBD stone "Present" group ($n = 610$), the AUC was 0.80 (95% CI: 0.76–0.83). Finally, because patients who had undergone cholecystectomy and patients who exhibited cirrhosis can alter the determinant validity (23, 24), the AUCs after exclusion of these patients ($n = 533$) were 0.81 for the scoring model (95% CI: 0.77, 0.84), 0.68 for the ASGE guidelines (95% CI: 0.64, 0.72), and 0.68 for the ESGE guidelines (95% CI: 0.64, 0.72). In summary, the scoring model's AUC was generally consistent regardless of the missing value management approach and the removal of data for patients with benign bile duct strictures or malignancy. The exclusion of s/p cholecystectomy and cirrhotic patients minimally increased the AUCs of the scoring model and the guidelines.

Discussion

A model's overall performance can be interpreted from its LHR + and AUC values (25–27). For patients in the high risk group, the scoring model's LHR + was 3.47 (95% CI: 2.60, 4.64). For an LHR + of 2 to 5, use of the model could presumably influence the pre-test to post-test probability (27). With a pre-test probability of 59.9% (CBD stone prevalence in this study), the CBD stone probability (i.e., post-test probability or positive predictive value) shifted to 83.9%, approximately 20% higher than the pre-test value. For the low-intermediate risk classification, the LHR + was 0.16 (95% CI: 0.09, 0.27). For an LHR + between 0.1 and 0.2, the model had a moderate likelihood of influencing pre-test to post-test probability (27). The probability of stone absence increased from 40.1–81.0% (i.e., negative predictive value); the probability of CBD stone presence decreased from 59.9–19.0%. The AUC value reflects a model's overall performance. The scoring model had an AUC of approximately 0.80 (95% CI: 0.76, 0.83); its discrimination properties were acceptable to excellent (AUC 0.70–0.80) (26). The internally validated AUC slightly decreased but remained near 0.80 (0.76; 95% CI: 0.72, 0.81). The proposed model exhibited significantly better performance than did the ASGE and ESGE guidelines for CBD stone prediction in the high-prevalence population, according to the comparative validation (AUC and decision curve analysis) (Fig. 5).

Concerning model predictors, we found that pancreatitis was a negative predictor, while cholangitis did not reach statistical significance. Regarding the negative for score pancreatitis, our results are consistent with published findings that most CBD stones in pancreatitis patients often spontaneously pass into the

gastrointestinal tract (28); a less-invasive investigational approach is appropriate in such patients (17). Furthermore, cholangitis, a strong clinical predictor of CBD stones (5, 10), was a non-significant variable in our multivariable analysis. This outcome is also consistent with previous literature (29, 30). The use of cholangitis as a sole predictor could be an important reason for the limited predictive ability of current guidelines. Notably, ALP was a potent predictor. ALP level > 250 mg/dL had the highest odds ratio (3.35; 95% CI: 2.02, 5.55). The significance of the ALP and CBD stone relationship has been extensively analyzed (3, 31, 32). However, ALP has minimal importance in current guidelines. Our findings suggest that more attention to ALP may be useful in future guidelines or the construction of predictive models.

Our scoring system is based on assessment of patient-specific predictors. The sum of assigned predictor scores (Table 2) serves as the individual patient's model-based score. The individual patient's score is used to support the assessment of CBD stone probability, together with the risk group classification. According to risk curve analysis, a higher score was associated with a higher probability of CBD stone presence (Fig. 4). Our scoring model can also be used in s/p cholecystectomy and cirrhotic patients, although these factors can affect CBD size and LFT results (23, 24).

When implementing the model, additional factors should be considered with respect to the availability of expert physicians and specialized equipment. Although the model could reasonably reduce the probability of CBD stones for the low-intermediate risk group, the probability remained moderate (i.e., 20–30%). IOC (or laparoscopic ultrasound (33)) may be the most reasonable approach because cholecystectomy can be performed in the same setting (34). However, for physicians or hospitals without the capability to treat detected stones, there may be a need for patient transfer or the use of less-invasive investigations (e.g., MRCP or endoscopic ultrasonography) (35). Laparoscopic bile duct exploration (trans-cystic/trans-ductal) (36) or same-setting ERC (i.e., ERC combined with cholecystectomy) (7) are potential methods for removal of IOC-detected CBD stones. However, because patients with CBD dilatation only comprised 12.9% (n = 11) of our cohort, trans-ductal CBD exploration could not be applied because it requires a dilated duct (36). In the absence of alternative interventions for IOC-detected CBD stones, possible treatment options are trans-cystic biliary stent insertion followed by transfer for ERC (in an ERC-capable hospital) (37), or the acquisition of a clear cystic duct (e.g., via ligation or clipping of the cystic duct stump to prevent leakage related to high pressure from the retained CBD stone) followed by rapid transfer. Postoperative abdominal pain or cholangitis can occur in patients with persistent stones (38). Persistent CBD stones are unlikely to increase the probability of cystic duct stump leakage, although they can aggravate its severity (39).

Our proposed model sufficiently increased the CBD stone probability that is appropriate for consideration of ERC in the high risk group (i.e., from 59.9–83.9%). However, our short survey indicated that some physicians expect near 100% CBD stone probability; endoscopic ultrasonography and ERC in the same setting may be optimal (40). This approach can almost avoid the need for diagnostic (unnecessary) ERC. However, because most CBD stone patients are older adults, the prolonged procedural time, increased sedation (41), and cost can limit the application of combined endoscopic ultrasonography and ERC. The scoring model may improve patient selection for this combined approach.

The intermediate-risk group might constitute an indeterminate group. The CBD stone chance was moderate (49.8% in our cohort); a less invasive investigation (e.g., MRCP or endoscopic ultrasonography) may thus be more suitable. Nevertheless, IOC is appropriate for all risk groups if experienced surgeons and specialized equipment are available (42). In the Supplemental Figure, we show a proposed CBD stone investigation and treatment flow approach regarding specific risk groups; we provide example checklists for clinical application in Supplemental Table 2.

There were considerable limitations in our study. First, we reviewed data from reference tests. Some patients with suspected CBD stones were not included in our data; other patients had few unusual findings in LFTs or imaging result abnormalities, and attending physicians chose observation as management for such patients. Thus, there were no reference test records for these patients. However, we considered the outcome validity to be an essential focus of the study; we did not modify the study protocol. With a similar potential selection bias issue, our reference tests were not included all CBD stone confirmatory tests. Endoscopic ultrasonography was not available in the study hospital during the study period. Second, a retrospective design is not the optimal data collection approach for a model development study because it involves various potential biases (15). Third, for the proposed application, LFTs should be examined within 7 days before using the score-based model to assess the CBD stone risk or choose a reference test that is compatible with our study protocol. Finally, the CBD prediction model was developed using data from patients with relevant clinical manifestations and a high-prevalence population. Thus, the findings cannot be applied to a low-prevalence population until they have been confirmed in additional studies.

Conclusions

Our proposed scoring model demonstrated reasonable ability to predict CBD stones; it is suitable for use in patients with relevant clinical manifestations or in a high-prevalence population. However, because there is variability among institutes concerning the investigation and treatment of CBD stones, the proposed model requires the consideration of whether specialized physicians or equipment are available. For application of the model to a low-prevalence population, additional studies are needed.

Abbreviations

ALP Alkaline phosphatase

ASGE American Society of Gastrointestinal Endoscopy

AUC Area under the receiver-operating characteristic curve

CBD Common bile duct

CI Confidence interval

CT Computed tomography

ERC Endoscopic retrograde cholangiography

ESGE European Society of Gastrointestinal Endoscopy

FU Follow-up

IOC Intraoperative cholangiography

LHR+ Positive likelihood ratio

MRCP Magnetic resonance cholangiopancreatography

LFT Liver function test

SGOT Serum glutamic oxaloacetic transaminase

SGPT Serum glutamic pyruvic transaminase

s/p Status-post

TB Total bilirubin

Declarations

Ethical approval and consent to participate

The study protocol was approved by the Human Research Ethics Committee of Thammasat University, Faculty of Medicine [MTU-EC-OO-0-169/64], and the Sawanpacharak Hospital Ethical Committee for Research in Human Subjects.

The requirement for patient consent was waived because of the retrospective study design and the use of de-identified data.

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information file.

Competing interests

All authors have no conflicts of interest or financial ties to disclose.

Funding

None(Funding for publication cost might be obtained after the manuscript is accepted).

Authors' contributions

All authors conceived and designed the study. ST and JP was responsible for statistical analysis. ST, BC, and KV participated in writing. ST, JP, CM participated in critical revision. All authors read and approved the final version of the manuscript.

Acknowledgments

We thank Ryan Chastain-Gross, Ph.D., from Edanz (<https://www.edanz.com/ac>) for editing a draft of this manuscript.

We thank Dr. Thawee Ratanachu-ek (<https://orcid.org/0000-0002-8579-1547>) for his valuable clinical suggestions.

References

1. Molvar C, Glaenzer B. Choledocholithiasis. Evaluation, Treatment, and Outcomes. *Semin Intervent Radiol.* 2016;33(4):268–76.
2. Cui ML, Cho JH, Kim TN. Long-term follow-up study of gallbladder in situ after endoscopic common duct stone removal in Korean patients. *Surg Endosc.* 2013;27(5):1711–6.
3. Gurusamy KS, Giljaca V, Takwoingi Y, Higgie D, Poropat G, Stimac D, et al. Ultrasound versus liver function tests for diagnosis of common bile duct stones. *Cochrane Database Syst Rev.* 2015(2):CD011548.
4. Freitas ML, Bell RL, Duffy AJ. Choledocholithiasis: evolving standards for diagnosis and management. *World J Gastroenterol.* 2006;12(20):3162–7.
5. Manes G, Paspatis G, Aabakken L, Anderloni A, Arvanitakis M, Ah-Soune P, et al. Endoscopic management of common bile duct stones: European Society of Gastrointestinal Endoscopy (ESGE) guideline. *Endoscopy.* 2019;51(5):472–91.
6. Freeman ML, Nelson DB, Sherman S, Haber GB, Herman ME, Dorsher PJ, et al. Complications of endoscopic biliary sphincterotomy. *N Engl J Med.* 1996;335(13):909–18.
7. Ghazal AH, Sorour MA, El-Riwini M, El-Bahrawy H. Single-step treatment of gall bladder and bile duct stones: a combined endoscopic-laparoscopic technique. *Int J Surg.* 2009;7(4):338–46.
8. Salama AF, Abd Ellatif ME, Abd Elaziz H, Magdy A, Rizk H, Basheer M, et al. Preliminary experience with laparoscopic common bile duct exploration. *BMC Surg.* 2017;17(1):32.
9. Liu TH, Consorti ET, Kawashima A, Tamm EP, Kwong KL, Gill BS, et al. Patient evaluation and management with selective use of magnetic resonance cholangiography and endoscopic retrograde cholangiopancreatography before laparoscopic cholecystectomy. *Ann Surg.* 2001;234(1):33–40.

10. Committee ASoP, Buxbaum JL, Abbas Fehmi SM, Sultan S, Fishman DS, Qumseya BJ, et al. ASGE guideline on the role of endoscopy in the evaluation and management of choledocholithiasis. *Gastrointest Endosc.* 2019;89(6):1075–105. e15.
11. Menezes N, Marson LP, debeaux AC, Muir IM, Auld CD. Prospective analysis of a scoring system to predict choledocholithiasis. *Br J Surg.* 2000;87(9):1176–81.
12. Nathan T, Kjeldsen J, Schaffalitzky de Muckadell OB. Prediction of therapy in primary endoscopic retrograde cholangiopancreatography. *Endoscopy.* 2004;36(6):527–34.
13. Trondsen E, Edwin B, Reiertsen O, Fagertun H, Rosseland AR. Selection criteria for endoscopic retrograde cholangiopancreatography (ERCP) in patients with gallstone disease. *World J Surg.* 1995;19(6):852–6. discussion 7.
14. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol.* 2009;62(1):5–12.
15. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1–73.
16. Kiriyaama S, Kozaka K, Takada T, Strasberg SM, Pitt HA, Gabata T, et al. Tokyo Guidelines 2018: diagnostic criteria and severity grading of acute cholangitis (with videos). *J Hepatobiliary Pancreat Sci.* 2018;25(1):17–30.
17. Tenner S, Baillie J, DeWitt J, Vege SS, American College of G. American College of Gastroenterology guideline: management of acute pancreatitis. *Am J Gastroenterol.* 2013;108(9):1400-15; 16.
18. Yokoe M, Hata J, Takada T, Strasberg SM, Asbun HJ, Wakabayashi G, et al. Tokyo Guidelines 2018: diagnostic criteria and severity grading of acute cholecystitis (with videos). *J Hepatobiliary Pancreat Sci.* 2018;25(1):41–54.
19. Garcea G, Ngu W, Neal CP, Dennison AR, Berry DP. Bilirubin levels predict malignancy in patients with obstructive jaundice. *HPB (Oxford).* 2011;13(6):426–30.
20. Pu LZ, Singh R, Loong CK, de Moura EG. Malignant Biliary Obstruction: Evidence for Best Practice. *Gastroenterol Res Pract.* 2016;2016:3296801.
21. Committee ASoP, Maple JT, Ikenberry SO, Anderson MA, Appalaneni V, Decker GA, et al. The role of endoscopy in the management of choledocholithiasis. *Gastrointest Endosc.* 2011;74(4):731–44.
22. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26(6):565–74.
23. Park SM, Kim WS, Bae IH, Kim JH, Ryu DH, Jang LC, et al. Common bile duct dilatation after cholecystectomy: a one-year prospective study. *J Korean Surg Soc.* 2012;83(2):97–101.
24. Ahmed Z, Ahmed U, Walayat S, Ren J, Martin DK, Moole H, et al. Liver function tests in identifying patients with liver disease. *Clin Exp Gastroenterol.* 2018;11:301–7.
25. Shapiro DE. The interpretation of diagnostic tests. *Stat Methods Med Res.* 1999;8(2):113–34.

26. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010;5(9):1315–6.
27. Chu K. An introduction to sensitivity, specificity, predictive values and likelihood ratios. *Emerg Med.* 1999;11(3):175–81.
28. Acosta JM, Ledesma CL. Gallstone migration as a cause of acute pancreatitis. *N Engl J Med.* 1974;290(9):484–7.
29. He H, Tan C, Wu J, Dai N, Hu W, Zhang Y, et al. Accuracy of ASGE high-risk criteria in evaluation of patients with suspected common bile duct stones. *Gastrointest Endosc.* 2017;86(3):525–32.
30. Kuzu UB, Odemis B, Disibeyaz S, Parlak E, Oztas E, Saygili F, et al. Management of suspected common bile duct stone: diagnostic yield of current guidelines. *HPB (Oxford).* 2017;19(2):126–32.
31. Isherwood J, Garcea G, Williams R, Metcalfe M, Dennison AR. Serology and ultrasound for diagnosis of choledocholithiasis. *Ann R Coll Surg Engl.* 2014;96(3):224–8.
32. Sheen AJ, Asthana S, Al-Mukhtar A, Attia M, Toogood GJ. Preoperative determinants of common bile duct stones during laparoscopic cholecystectomy. *Int J Clin Pract.* 2008;62(11):1715–9.
33. Aziz O, Ashrafian H, Jones C, Harling L, Kumar S, Garas G, et al. Laparoscopic ultrasonography versus intra-operative cholangiogram for the detection of common bile duct stones during laparoscopic cholecystectomy: a meta-analysis of diagnostic accuracy. *Int J Surg.* 2014;12(7):712–9.
34. Ali FS, DaVee T, Bernstam EV, Kao LS, Wandling M, Hussain MR, et al. Cost-effectiveness analysis of optimal diagnostic strategy for patients with symptomatic cholelithiasis with intermediate probability for choledocholithiasis. *Gastrointest Endosc.* 2022;95(2):327–38.
35. Meeralam Y, Al-Shammari K, Yaghoobi M. Diagnostic accuracy of EUS compared with MRCP in detecting choledocholithiasis: a meta-analysis of diagnostic test accuracy in head-to-head studies. *Gastrointest Endosc.* 2017;86(6):986–93.
36. Gupta N. Role of laparoscopic common bile duct exploration in the management of choledocholithiasis. *World J Gastrointest Surg.* 2016;8(5):376–81.
37. Gomez D, Cox MR. Laparoscopic Transcystic Stenting and Postoperative ERCP for the Management of Common Bile Duct Stones at Laparoscopic Cholecystectomy. *Ann Surg.* 2018;267(5):e86-e8.
38. Lee DH, Ahn YJ, Lee HW, Chung JK, Jung IM. Prevalence and characteristics of clinically significant retained common bile duct stones after laparoscopic cholecystectomy for symptomatic cholelithiasis. *Ann Surg Treat Res.* 2016;91(5):239–46.
39. Shaikh IA, Thomas H, Joga K, Amin AI, Daniel T. Post-cholecystectomy cystic duct stump leak: a preventable morbidity. *J Dig Dis.* 2009;10(3):207–12.
40. Moutinho-Ribeiro P, Peixoto A, Macedo G. Endoscopic Retrograde Cholangiopancreatography and Endoscopic Ultrasound: To Be One Traveler in Converging Roads. *GE - Portuguese Journal of Gastroenterology.* 2018;25(3):138–45.

41. Gornals JB, Esteban JM, Guarner-Argente C, Marra-Lopez C, Repiso A, Sendino O, et al. Endoscopic ultrasound and endoscopic retrograde cholangiopancreatography: Can they be successfully combined? *Gastroenterol Hepatol.* 2016;39(9):627–42.
42. Zhu J, Li G, Du P, Zhou X, Xiao W, Li Y. Laparoscopic common bile duct exploration versus intraoperative endoscopic retrograde cholangiopancreatography in patients with gallbladder and common bile duct stones: a meta-analysis. *Surg Endosc.* 2021;35(3):997–1005.

Figures

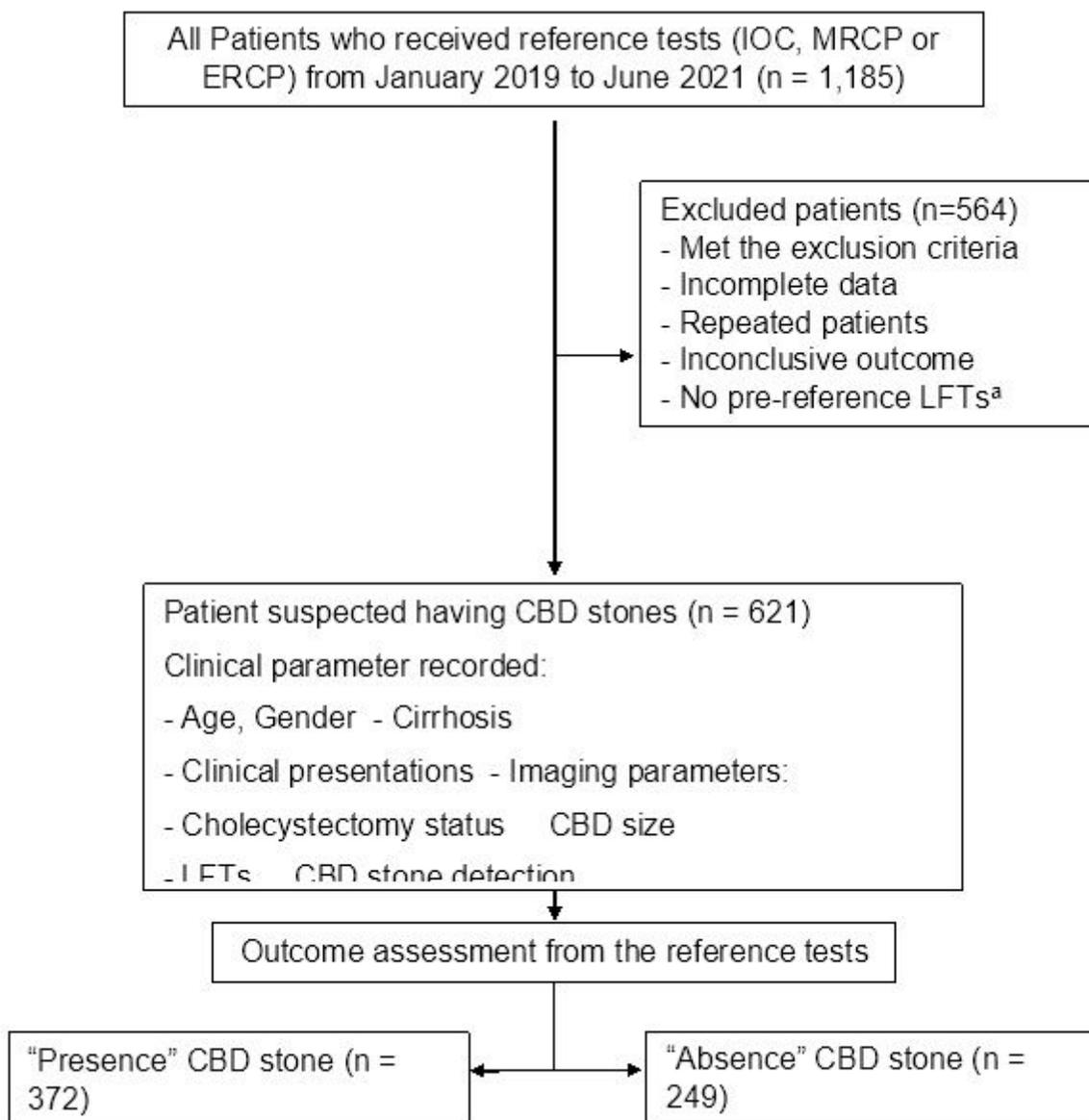


Figure 1

Study participant flow diagram

^a Pre-reference LFTs = LFTs within 7 days before reference tests

CBD common bile duct, ERC endoscopic retrograde cholangiography, IOC intraoperative cholangiography, LFTs liver function tests, MRCP magnetic resonance cholangiopancreatography

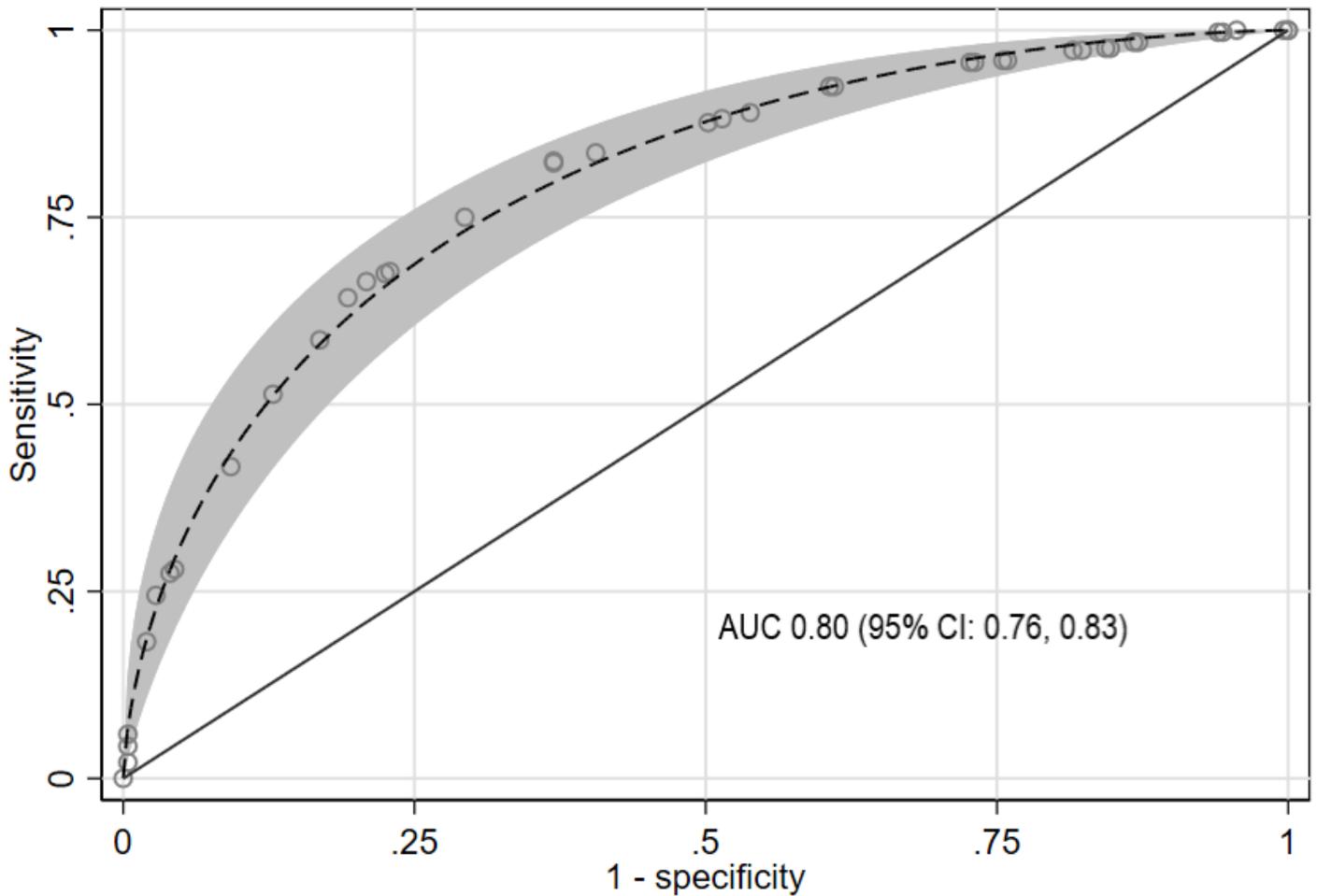


Figure 2

Parametric ROC with 95% confidence band for CBD stone prediction using the scoring model

AUC area under the receiver operating characteristic curve, CBD common bile duct, CI confidence interval, ROC receiver operating characteristic curve

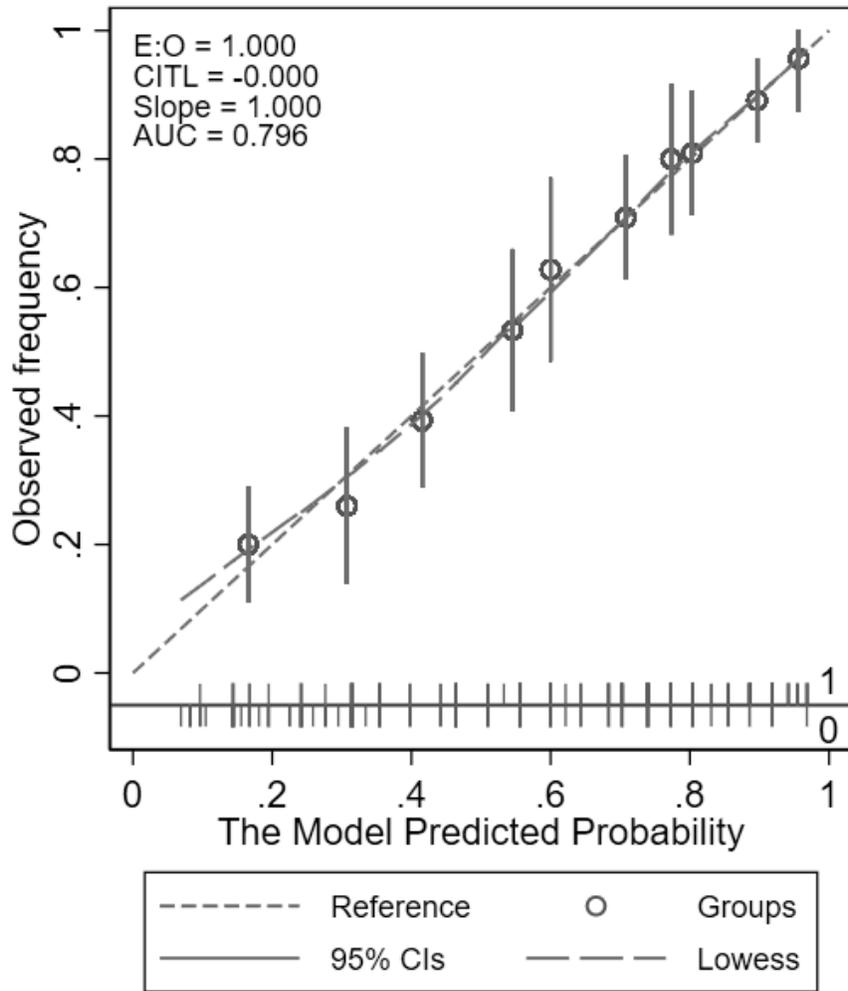


Figure 3

Calibration plot comparing the score-predicted and observed risks of common bile duct stone

AUC area under the receiver operating characteristic curve, CIs confidence intervals, CITL calibration-in-the-large, LOWESS locally weighted scatterplot smoothing

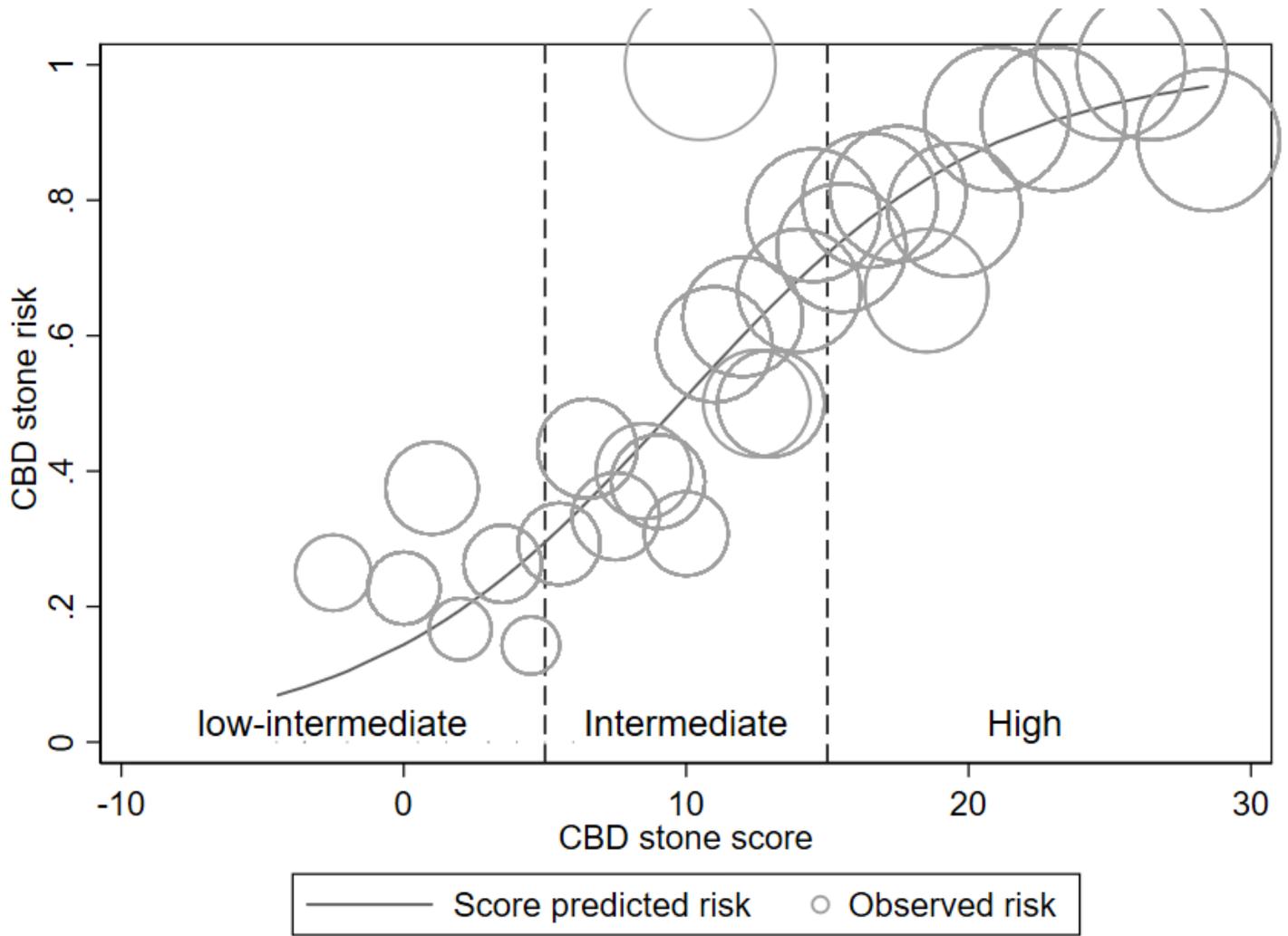


Figure 4

Risk curve

Risk curve illustrating the score-predicted CBD stone risk (solid line) and the observed stone risk (hollow circles) according to risk group (vertical dash line). The relative number of patients corresponds to the circle's size.

CBD common bile duct

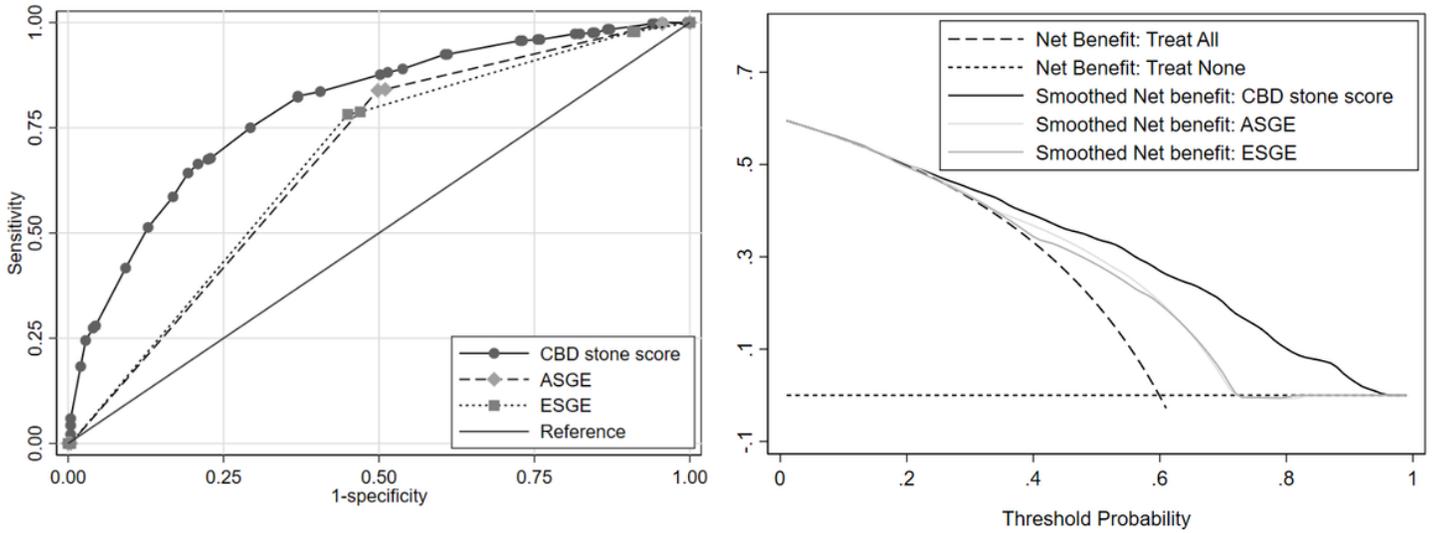


Figure 5

Comparing validation of CBD stone score performance to CBD stone guidelines

Discriminative ability with ROC is shown in left panel and clinical utility with decision curve analysis is shown in right panel.

ASGE American Society of Gastrointestinal Endoscopy, CBD common bile duct, ESGE European Society of Gastrointestinal Endoscopy, ROC receiver operating characteristic curve

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Onlinegroupimage2.png](#)
- [SuppTables.docx](#)