



Original Article

Machine learning models compared to existing criteria for noninvasive prediction of endoscopic retrograde cholangiopancreatography-confirmed choledocholithiasis[☆]

Camellia Dalai^a, John M Azizian^a, Harry Trieu^b, Anand Rajan^a, Formosa C Chen^c, Tien Dong^d, Simon W Beaven^{d,e}, James H. Tabibian^{d,e,*}

^a UCLA-Olive View Internal Medicine Residency Program, Department of Medicine, Olive View-UCLA Medical Center, Sylmar, CA, USA

^b Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

^c Department of Surgery, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

^d Vatche and Tamar Manoukian Division of Digestive Diseases, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

^e Division of Gastroenterology, Department of Medicine, Olive View-UCLA Medical Center, Sylmar, CA, USA

ARTICLE INFO

Article history:

Received 31 March 2021

Received in revised form

5 August 2021

Accepted 20 October 2021

Keywords:

Machine learning models (MLMs)

Endoscopic retrograde

cholangiopancreatography (ERCP)

Noninvasive imaging

Bile duct disorders

Common bile duct stones

Gallstones

ABSTRACT

Background and aims: Noninvasive predictors of choledocholithiasis have generally exhibited marginal performance characteristics. We aimed to identify noninvasive independent predictors of endoscopic retrograde cholangiopancreatography (ERCP)-confirmed choledocholithiasis and accordingly developed predictive machine learning models (MLMs).

Methods: Clinical data of consecutive patients undergoing first-ever ERCP for suspected choledocholithiasis from 2015 to 2019 were abstracted from a prospectively-maintained database. Multiple logistic regression was used to identify predictors of ERCP-confirmed choledocholithiasis. MLMs were then trained to predict ERCP-confirmed choledocholithiasis using pre-ERCP ultrasound (US) imaging only as well as using all available noninvasive imaging (US, computed tomography, and/or magnetic resonance cholangiopancreatography). The diagnostic performance of American Society for Gastrointestinal Endoscopy (ASGE) “high-likelihood” criteria was compared to MLMs.

Results: We identified 270 patients (mean age 46 years, 62.2% female, 73.7% Hispanic/Latino, 59% with noninvasive imaging positive for choledocholithiasis) with native papilla who underwent ERCP for suspected choledocholithiasis, of whom 230 (85.2%) were found to have ERCP-confirmed choledocholithiasis. Logistic regression identified choledocholithiasis on noninvasive imaging (odds ratio (OR) = 3.045, $P = 0.004$) and common bile duct (CBD) diameter on noninvasive imaging (OR = 1.157, $P = 0.011$) as predictors of ERCP-confirmed choledocholithiasis. Among the various MLMs trained, the random forest-based MLM performed best; sensitivity was 61.4% and 77.3% and specificity was 100% and 75.0%, using US-only and using all available imaging, respectively. ASGE high-likelihood criteria demonstrated sensitivity of 90.9% and specificity of 25.0%; using cut-points achieving this specificity, MLMs achieved sensitivity up to 97.7%.

Conclusions: MLMs using age, sex, race/ethnicity, presence of diabetes, fever, body mass index (BMI), total bilirubin, maximum CBD diameter, and choledocholithiasis on pre-ERCP noninvasive imaging predict ERCP-confirmed choledocholithiasis with good sensitivity and specificity and outperform the ASGE criteria for patients with suspected choledocholithiasis.

© 2021 The Third Affiliated Hospital of Sun Yat-sen University. Publishing services by Elsevier B. V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Gallstone disease is the leading inpatient gastrointestinal disorder in the United States, with an estimated annual cost of \$10 billion.^{1,2} Despite its ubiquity, it continues to pose etiopathogenic,

[☆] Edited by Yuxia Jiang, Peiling Zhu and Genshu Wang.

* Corresponding author. Division of Gastroenterology, Department of Medicine, Olive View-UCLA Medical Center, Sylmar, CA, USA.

E-mail address: jtabibian@dhs.lacounty.gov (J.H. Tabibian).

diagnostic, and therapeutic uncertainties.^{3–7} Approximately, 15% of patients with gallstone disease will also experience choledocholithiasis, which poses additional management challenges.⁸ Given choledocholithiasis is typically secondary to cholelithiasis, predictors of the former have historically been extrapolated from those of the latter, including the “4 Fs”, namely: obesity (“fat”), female sex, and middle reproductive age (“fertile and forty”).^{5,8–10} The accuracy of the 4 Fs, however, has not been well-examined nor recently investigated with respect to choledocholithiasis. Moreover, as only a small proportion of patients with cholelithiasis go on to develop choledocholithiasis, the need for more nuanced and selective noninvasive predictors is evident.^{11,12} In a seminal guideline, the American Society for Gastrointestinal Endoscopy (ASGE) proposed choledocholithiasis risk criteria wherein common bile duct (CBD) stone on transabdominal ultrasound (US), clinical ascending cholangitis, and total bilirubin >4 mg/dL were deemed as “very strong” predictors and CBD dilation >6 mm on US with gallbladder *in situ* and total bilirubin 1.8–4.0 mg/dL as “strong” predictors of choledocholithiasis.^{13,14} While providing a basic clinical framework, these criteria have been found to have suboptimal diagnostic performance characteristics compared to the gold standard of choledocholithiasis confirmed by endoscopic retrograde cholangiopancreatography (ERCP).^{15–18}

Numerous studies, prior to and after the aforementioned ASGE guideline, have aimed to identify independent predictors of choledocholithiasis, but many of these were prior to the era of magnetic resonance cholangiopancreatography (MRCP), not based on contemporary cohorts (i.e. published >10–20 years ago), lacking a diverse patient population, and/or contradictory to each other.^{16,19–24} Likewise, attempts to formulate an algorithm to predict choledocholithiasis have not yielded a widely adopted or validated risk assessment instrument.^{15,25–27} Thus, aside from the subset of patients with an obvious obstructing stone seen on noninvasive imaging or signs of acute cholangitis without alternative explanation (e.g., acute cholecystitis), it can often be unclear which patients need (therapeutic) ERCP.

Our institution is one of three hospitals within the Los Angeles County Department of Health Services (LADHS), the second largest municipal healthcare system in the United States.²⁸ The prevalence of ethnorracial minorities in LADHS, particularly majority Hispanic/Latino patients, provides a unique opportunity to study choledocholithiasis, especially given the association between certain ethnorracial backgrounds and gallstone disease.⁶ Therefore, in the present study, we examined the clinical epidemiology of suspected and ERCP-confirmed choledocholithiasis in our patient population, and in particular: (i) assessed the features and characteristics of a contemporary cohort of patients with suspected choledocholithiasis, (ii) identified independent predictors of ERCP-confirmed choledocholithiasis using multiple logistic regression, (iii) utilized machine learning models (MLMs) to develop a tool to predict ERCP-confirmed choledocholithiasis, and (iv) compared the performance of MLMs to the ASGE choledocholithiasis risk criteria. Additionally, we used our findings to develop a clinician-oriented, free, MLM-based web application to facilitate risk-stratification of patients with suspected choledocholithiasis.

2. Methods

2.1. Study setting and population

This study was conducted at Olive View-UCLA Medical Center (OVMC), a 377-bed LADHS tertiary care teaching hospital and was approved by its institutional review board. Using a prospectively maintained endoscopy database, we retrospectively reviewed all ERCPs performed from 1 November 2015 to 31 December 2019 in

patients aged 18 years and older. Basic patient demographics (age, sex, and race/ethnicity) and the indication for each ERCP were then abstracted from the electronic medical record using a standardized data collection form. For ERCPs performed with the indication of suspected (or “rule out”) choledocholithiasis, additional data (biochemical, radiologic, and endoscopic) were collected.

ERCPs performed for indications other than suspected choledocholithiasis, including biliary stricture, bile leak, malignancy, and “others” (e.g., stent exchange, stent removal, and pancreatography) were excluded. In addition, only the index (i.e. first-ever) ERCP for suspected choledocholithiasis was included (i.e. subsequent ERCPs for stent removal or other indications were excluded, as were patients with prior biliary sphincterotomy for any reason). Patients who did not undergo pre-ERCP US were excluded for consistency with the ASGE choledocholithiasis risk criteria, which utilize predictors that are to be evaluated specifically on US as the noninvasive imaging modality.

2.2. Study outcome measure and variables

The primary study outcome was ERCP-confirmed choledocholithiasis. Clinical variables analyzed included demographic, biochemical, radiologic, and cholangiographic data including: age, race/ethnicity, sex, body mass index (BMI, kg/m²), diabetes history, history of cholecystectomy pre-ERCP, peak serum total bilirubin pre-ERCP, peak temperature pre-ERCP, maximum CBD diameter on pre-ERCP noninvasive imaging (US, computed tomography (CT), or MRCP), and presence of choledocholithiasis on pre-ERCP noninvasive imaging. ERCP-confirmed choledocholithiasis was defined as the presence of bile duct stone or obstructing debris/sludge visualized fluoroscopically during ERCP or as seen directly by white light endoscopy (e.g. in the duodenal lumen following ductal sweeping). Acute cholangitis was defined as the presence of objective fever without alternative explanation pre-ERCP and in the setting of suspected choledocholithiasis.

2.3. Statistical analyses

Two-sample *t*-tests and chi-squared (χ^2) tests were used to compare demographic, laboratory, radiologic, and other pre-ERCP clinical parameters between patients with and without ERCP-confirmed choledocholithiasis. Patients were divided into two groups based on the presence of choledocholithiasis on pre-ERCP noninvasive imaging, and the same parameters were then compared between patients with and without ERCP-confirmed choledocholithiasis within these two groups.

We fit two multiple logistic regression models to examine the utility of the aforementioned parameters in predicting patients with ERCP-confirmed choledocholithiasis. In the first model, maximum CBD diameter and choledocholithiasis were assessed on US only. In the second model, all available noninvasive imaging modalities (US, CT, and MRCP) were assessed for these variables; the greatest pre-ERCP CBD diameter (if not concordant between modalities) was used. The decision to evaluate the imaging-dependent predictors two ways (i.e. one model using US only and the second model using all noninvasive imaging modalities) was made due to the fact that a large proportion of patients often only undergo US prior to ERCP, and the ASGE criteria references only US as the noninvasive imaging modality. In order to maintain a ratio of roughly 10 negative outcomes (absence of ERCP-confirmed choledocholithiasis) to 1 predictor and prevent overfitting, the number of predictors in multiple logistic regression modeling was limited to four. Age, pre-ERCP total bilirubin, maximum CBD diameter measured on noninvasive imaging, and evidence of

choledocholithiasis on noninvasive imaging were selected as predictors.

Next, we trained MLMs to predict the presence of ERCP-confirmed choledocholithiasis. The dataset was divided into training and testing sets with 80% of the observations assigned to the former and the remaining 20% to the latter. The testing set included a group of patients that were held out for the purpose of evaluating the performance of our MLMs. Our study trained 4 different supervised learning models on the training set: a generalized linear model (GLM), support vector machine (SVM) with linear kernels, SVM with radial basis function (RBF) kernels, and random forest. Six patients (2.2%) were excluded from the MLMs because they were missing BMI measurements, and several of the MLMs used BMI as a feature (i.e. variable). In the GLM and both SVM models, age, pre-ERCP total bilirubin, maximum CBD diameter on noninvasive imaging, and choledocholithiasis on noninvasive imaging were used as features (i.e. variables). The random forest models were allowed to select from features including female sex, diabetes, white race, age, fever, BMI, maximum CBD diameter on noninvasive imaging, presence of choledocholithiasis on noninvasive imaging, and total bilirubin prior to ERCP upon which to split each tree. As with the aforementioned logistic regression models, each of the four MLMs was trained with maximum CBD diameter and evidence of choledocholithiasis evaluated by US only and then separately using all available imaging modalities, yielding 8 total models. We used 10-fold cross-validation with 3 repeats for resampling when tuning the SVM with RBF kernel hyperparameters. Each MLM was then validated on the testing set, and receiver operator characteristic (ROC) curves and area under the ROC curve (AUROC) were generated. The optimal probability cut-point above which a patient would be considered to have ERCP-confirmed choledocholithiasis was calculated using Youden's index and the point closest to (0, 1) method. Youden's index maximizes the sum of sensitivity and specificity while the point closest to (0, 1) method minimizes the Euclidean distance between the ROC curve and the (0, 1) point.²⁹ Accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were then calculated for each MLM at its optimal cut-point.

Lastly, the ability of the ASGE guideline's choledocholithiasis risk criteria to predict ERCP-confirmed choledocholithiasis was examined utilizing the testing set. The ASGE guideline indicates the presence of at least one "very strong" predictor (CBD stone on US, clinical ascending cholangitis, or total bilirubin >4 mg/dL) or the presence of both "strong" predictors: (dilated CBD on US > 6 mm and total bilirubin level between 1.8 and 4 mg/dL) as predicting a high likelihood of choledocholithiasis.³⁰ These predictions were compared to actual ERCP findings, and accuracy, sensitivity, specificity, PPV, and NPV were calculated. The performance of the ASGE criteria was then compared to that of our MLMs.

Descriptive statistics were performed using Stata/IC 16.1 (StataCorp, College Station, TX, USA). Logistic regression and machine learning experiments were performed using R 4.0.2 and the caret, random forest, and pROC libraries. A *P*-value <0.05 was considered statistically significant.

3. Results

3.1. Overview of ERCPs performed during the study period

Of the 641 ERCPs performed, 289 (45.1%) were index ERCPs performed for an indication of suspected choledocholithiasis (Supplementary Table 1). Of these 289 patients, presence of choledocholithiasis and maximum CBD diameter were assessed on pre-ERCP US in 270, and these patients were thus included for further

study (Fig. 1). The mean age of these 270 patients was 46 years, 62.2% were female, and 73.7% were Hispanic/Latino.

3.2. Characteristics and univariate analyses of patients undergoing ERCP for suspected choledocholithiasis

Of the 270 ERCPs performed for suspected choledocholithiasis, choledocholithiasis was confirmed in 230 (85.2%), as shown in Table 1. Among patients with ERCP-confirmed choledocholithiasis, 64.8% were female and 13.0% had diabetes compared to 47.5% and 32.5%, respectively, in choledocholithiasis-negative ERCP patients (*P*-values of 0.037 and 0.002). Median of maximum CBD diameter on noninvasive imaging was 9.0 mm in the ERCP-confirmed choledocholithiasis group and 8.0 mm in the choledocholithiasis-negative ERCP group (*P* = 0.011).

In the group without evidence of choledocholithiasis on noninvasive imaging, diabetes was more common in choledocholithiasis-negative ERCP patients compared to ERCP-confirmed choledocholithiasis patients (40.7% vs. 10.3%, *P* < 0.001) (Table 2), but otherwise the two subgroups were very similar. In the group with evidence of choledocholithiasis on noninvasive imaging, there was no statistically significant difference between patients with ERCP-confirmed choledocholithiasis and patients with choledocholithiasis-negative ERCP (Table 2).

3.3. Multiple logistic regression identifies positive and negative independent predictors of ERCP-confirmed choledocholithiasis

In the multiple logistic regression model fit using predictors assessed on US only, every 1 mm increase in maximum CBD diameter was associated with increased odds of ERCP-confirmed choledocholithiasis (odds ratio (OR) = 1.157, *P* = 0.011) (Fig. 2a). In the model fit using predictors assessed on all available imaging modalities, the presence of choledocholithiasis was associated with increased odds of ERCP-confirmed choledocholithiasis (OR = 3.045, *P* = 0.004) (Fig. 2b).

3.4. MLMs predict ERCP-confirmed choledocholithiasis with good accuracy, sensitivity, and specificity

The training set consisted of 212 patients, of which 180 had ERCP-confirmed choledocholithiasis, while the testing set consisted of 52 patients, of which 44 had ERCP-confirmed choledocholithiasis. When imaging-dependent features were evaluated on US only, the random forest model demonstrated the greatest AUROC (0.791) of the four supervised learning models (Fig. 3a). When imaging-dependent features were evaluated on all available imaging modalities, the random forest model again performed best with an AUROC of 0.801 (Fig. 3b). The four most important features in both random forest models, as determined by greatest mean decrease in Gini, were total bilirubin pre-ERCP, age, BMI, and maximum CBD diameter on noninvasive imaging (Supplementary Table 2).

The optimal cut-point as determined by the Youden index and point closest to (0, 1) for the random forest model trained on US measurements only were 0.852 and 0.793, respectively (Table 3). At a cut-point of 0.852, 27 of 44 choledocholithiasis-positive and 8 of 8 choledocholithiasis-negative cases were correctly identified. This yielded an accuracy of 67.3%, sensitivity of 61.4%, specificity of 100%, PPV of 100%, and NPV of 32.0%. At a cut-point of 0.793, 32 of 44 ERCP-confirmed choledocholithiasis cases and 6 of 8 choledocholithiasis-negative ERCP cases were correctly identified. This yielded an accuracy of 73.1%, sensitivity of 72.7%, specificity of 75.0%, PPV of 94.1%, and NPV of 33.3% (Table 3).

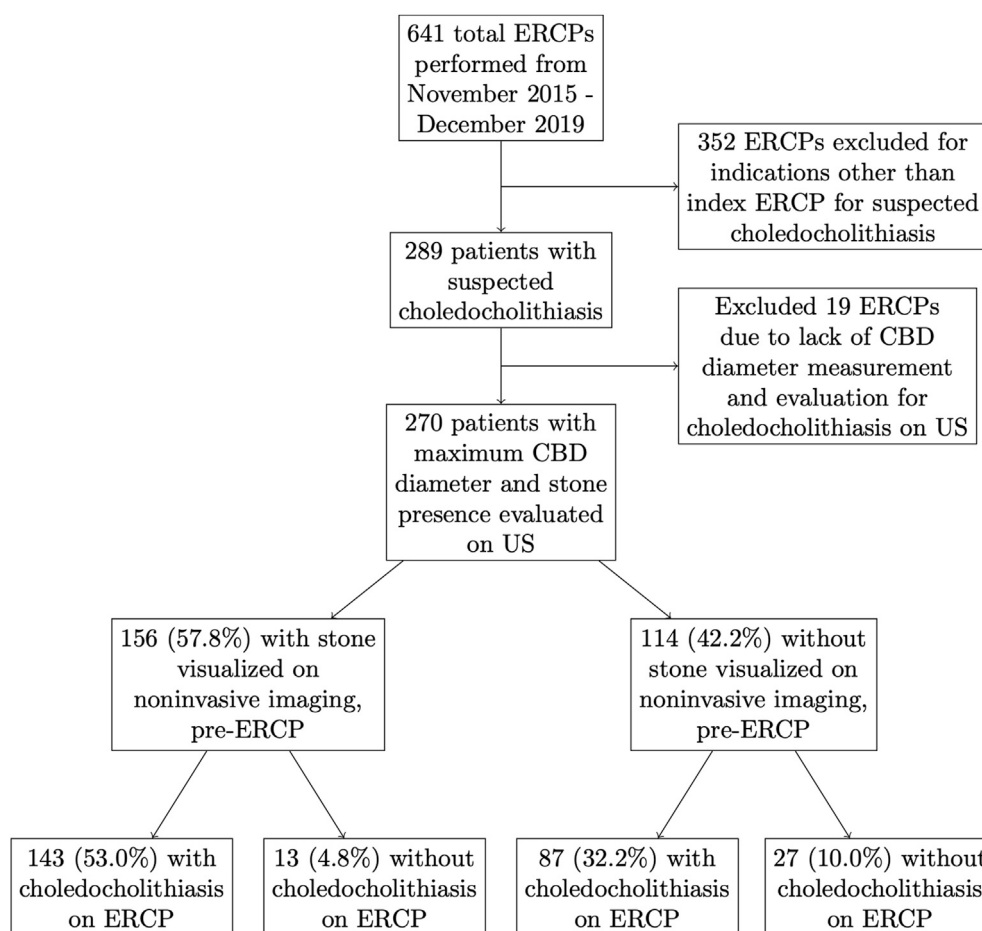


Fig. 1. ERCP indication and outcome flow diagram. Study flow diagram demonstrating proportion of ERCPs performed for suspected choledocholithiasis, presence or absence of choledocholithiasis on noninvasive imaging, and subsequent ERCP result (confirmed choledocholithiasis or absence thereof). Abbreviations: CBD, common bile duct; ERCP, endoscopic retrograde cholangiopancreatography; US, ultrasound.

The optimal cut-point as determined by both the Youden index and point closest to (0, 1) method for the random forest model trained on all available imaging modalities was 0.825. At this cut-point, ERCP-confirmed choledocholithiasis was correctly identified in the testing set 34 of 44 times while absence of ERCP-confirmed choledocholithiasis was correctly identified 6 of 8 times. This yielded an accuracy of 76.9%, sensitivity of 77.3%, specificity of 75.0%, PPV of 94.4%, and NPV of 37.5%.

3.5. MLMs outperform ASGE high-likelihood criteria at predicting ERCP-confirmed choledocholithiasis in the testing set

The ASGE high-likelihood criteria correctly predicted ERCP-confirmed choledocholithiasis in 40 of 44 cases, while absence of choledocholithiasis was correctly predicted 2 of 8 times. This yielded an accuracy of 80.8%, sensitivity of 90.9%, specificity of 25.0%, PPV of 87.0%, and NPV of 33.3%. The point on the ROC curve for the random forest model trained on US only corresponding to a specificity of 25.0% was a sensitivity of 90.9%, PPV of 87.0%, and NPV of 33.3%. By comparison, when trained on all available imaging modalities, the point on the ROC curve corresponding to a specificity of 25.0% had a sensitivity of 97.7%, PPV of 87.8%, and NPV of 66.7%. The GLM and SVM-based models achieved a sensitivity higher than 90.9% at the cut-point which achieves 25.0% specificity when trained on US-based measurements only, as shown in Fig. 3a and b. Although an ROC curve could not be generated for the ASGE criteria

given there was no boundary to vary, the point corresponding to its sensitivity and specificity was plotted. Our random forest-based model trained on all available noninvasive imaging modalities is available for use online at <https://harrytrieu.shinyapps.io/choledocholithiasisrisk/>.

4. Discussion

ERCP, the gold standard for diagnosing and treating choledocholithiasis, is invasive and costly, thus making accurate *a priori* patient selection using noninvasive predictors crucial. We identified noninvasive predictors of ERCP-confirmed choledocholithiasis in our predominantly Hispanic/Latino patient population and trained multiple MLMs to predict the presence or absence of choledocholithiasis on ERCP using noninvasive clinical and demographic parameters. We validated the performance of these MLMs and compared them to the current ASGE high-likelihood criteria for choledocholithiasis.

Of the various MLMs we trained to predict the presence of ERCP-confirmed choledocholithiasis, we found the random forest model performed best. The random forest model trained on predictors using all available imaging modalities demonstrated a sensitivity of 77.3%, specificity of 75%, PPV of 94.4%, and NPV of 37.5% at the optimal cut-point (Youden index). While the ASGE high-likelihood criteria yielded higher sensitivity (90.9%) and poorer specificity (25%) compared to the random forest model at the optimal cut-

Table 1

Demographic, clinical, biochemical, and radiological features of all patients who underwent ERCP for suspected choledocholithiasis grouped by ERCP findings.

Characteristics	ERCP-confirmed choledocholithiasis (n = 230)	Choledocholithiasis-negative ERCP (n = 40)	P-value
Age at ERCP (years), median (IQR)	46 (32–57)	47 (32–64)	0.640
Age >40 years, n (%)	139 (60.4)	23 (57.5)	0.727
Female, n (%)	149 (64.8)	19 (47.5)	0.037
Race/ethnicity, n (%)			0.295
White	14 (6.1)	6 (15.0)	–
Hispanic/Latino	173 (75.2)	26 (65.0)	–
Black	3 (1.3)	0 (0)	–
Asian	7 (3.0)	1 (2.5)	–
Other/unknown	33 (14.4)	7 (17.5)	–
BMI (kg/m ²), median (IQR) ^a	29.0 (26.0–34.0)	29.1 (26.5–35.0)	0.772
BMI >30 kg/m ² , n (%) ^a	91 (40.6)	16 (40.0)	0.941
Diabetes, n (%)	30 (13.0)	13 (32.5)	0.002
Fever, n (%)	40 (17.4)	11 (27.5)	0.132
Classic cholelithiasis risk factors (BMI >30 kg/m ² , female, age >40 years), n (%)			
0 risk factor	16 (7.1)	5 (12.5)	0.249
1 risk factor	80 (35.7)	17 (42.5)	0.412
2 risk factors	92 (41.1)	13 (32.5)	0.308
3 risk factors	36 (16.1)	5 (12.5)	0.566
History of cholecystectomy, n (%)	54 (23.5)	9 (22.5)	0.893
Peak bilirubin pre-ERCP (mg/dL), median (IQR)	2.1 (0.5–3.7)	2.1 (0.4–4.1)	0.823
Noninvasive imaging performed, n (%)			
Ultrasound	230 (100)	40 (100)	–
CT	99 (43.0)	17 (42.5)	0.949
MRCP	64 (27.8)	13 (32.5)	0.546
EUS performed, n (%)	12 (5.3)	3 (8.1)	0.487
Maximum CBD diameter (mm) on noninvasive imaging, median (IQR)	9.0 (6.0–11.0)	8.0 (4.0–9.5)	0.011
Noninvasive imaging positive for choledocholithiasis, n (%)	147 (63.9)	13 (32.5)	<0.001
Ultrasound positive for choledocholithiasis, n (%)	82 (35.7)	10 (25.0)	–
CT positive for choledocholithiasis, n (%)	36 (15.7)	1 (2.5)	–
MRCP positive for choledocholithiasis, n (%)	57 (24.8)	5 (12.5)	–

Data are shown as n (%) or median (IQR).

Abbreviations: BMI, body mass index; CBD, common bile duct; CT, computed tomography; ERCP, endoscopic retrograde cholangiopancreatography; EUS, endoscopic ultrasound; MRCP, magnetic resonance cholangiopancreatography; IQR, interquartile range.

^a Six patients (2.2%) were missing BMI measurements.

point, when the random forest models were evaluated using cut-points which achieved a specificity equal to the ASGE criteria, the model trained on all imaging modalities demonstrated greater sensitivity (97.7%).

An advantage to using the MLMs to determine whether or not to proceed with ERCP is that they predict ERCP-confirmed choledocholithiasis using a single set of noninvasive clinical parameters, whereas the ASGE criteria stratify patients into high-, intermediate-, and low-risk for choledocholithiasis based on different sets of parameters, some of which have relatively arbitrary cutoffs (e.g., bilirubin of 4 mg/dL). Another advantage is the fact that MLMs can achieve different combinations of sensitivity and specificity by varying the cut-point, whereas the ASGE criteria produce a fixed sensitivity and specificity.

Identifying patients who have choledocholithiasis and those who do not in a cohort where there is already high clinical suspicion is an inherently challenging task. In our testing set, 84.6% of patients had ERCP-confirmed choledocholithiasis, suggesting that clinical suspicion of experienced endoscopists is already quite adept at correctly identifying these patients. As such, there may be greater utility in correctly predicting which patients will not have ERCP-confirmed choledocholithiasis. In this vein, our MLMs may be helpful in identifying patients who have had spontaneous passage of biliary stones after presentation (e.g., a gallstone pancreatitis patient in whom the stone passes spontaneously before ERCP is performed) and thus avoiding an unnecessary invasive procedure in patients with a lower-likelihood of ERCP-confirmed choledocholithiasis and intervenable findings. The MLM trained on US-based measurements may be ideal for use when a patient has only undergone transabdominal US and high specificity is desired, while

the MLM trained on all available imaging useful when the patient has undergone other abdominal imaging modalities.

It is likely that the reason no one set of risk stratification criteria has thus far exhibited uniformly strong diagnostic performance is the heterogeneity of patients with suspected choledocholithiasis and the presence of distinct subgroups; for instance, there are patients with suspected choledocholithiasis with vs. without acute (gallstone) pancreatitis, with vs. without choledocholithiasis (or bile duct dilation) on noninvasive imaging, and with vs. without gallbladder intact. It is conceivable, though, that with more training data, an MLM can be developed that performs well in all subgroups and circumstances. A multi-center study is underway in this regard.

Our study has several strengths. First, it provides clinical data regarding choledocholithiasis from a minority-predominant patient population that has not previously been reported. This is a vulnerable patient group wherein healthcare disparities are prevalent and gallstone disease is common. Indeed, studies have shown that patients of lower socioeconomic status, who are uninsured, or Medicaid-insured are less likely to undergo a cholecystectomy in a timely manner and have worse outcomes after cholecystectomy.^{2,11} Second, in contrast to similar studies which aimed to identify predictors of ERCP-confirmed choledocholithiasis,³¹ our analysis did not exclude patients with history of cholecystectomy, in whom it has been reported that 10% or more will subsequently be diagnosed with choledocholithiasis.³² Moreover, we found that this subset was even larger than expected (nearly 25% of patients), further reinforcing the need to include such patients, since MLMs excluding such individuals would overlook this non-insignificant population subset. Third, we utilized numerous advanced analytic techniques to predict ERCP-confirmed choledocholithiasis; we

Table 2

Comparison of demographic, clinical, biochemical, and radiological features of patients grouped by evidence of choledocholithiasis on noninvasive imaging and presence or absence of ERCP-confirmed choledocholithiasis.

Characteristics	Patients with pre-ERCP noninvasive imaging positive for choledocholithiasis (n = 156)			Patients with pre-ERCP noninvasive imaging negative for choledocholithiasis (n = 114)		
	ERCP-confirmed choledocholithiasis (n = 143)	Choledocholithiasis-negative ERCP (n = 13)	P-value	ERCP-confirmed choledocholithiasis (n = 87)	Choledocholithiasis-negative ERCP (n = 27)	P-value
Age at ERCP (years), median (IQR)	46 (33–57)	36 (29–60)	0.517	46 (31–56)	57 (34–64)	0.288
Age >40 years, n (%)	87 (60.8)	5 (38.5)	0.147	52 (59.8)	18 (66.7)	0.706
Female, n (%)	101 (70.6)	6 (46.2)	0.098	48 (55.2)	13 (48.2)	0.379
Race/ethnicity, n (%)			0.963			0.194
White	10 (7.0)	1 (7.7)	–	4 (4.6)	5 (18.5)	–
Hispanic/Latino	112 (78.3)	10 (76.9)	–	61 (70.1)	16 (59.3)	–
Black	1 (0.7)	0	–	2 (2.3)	0	–
Asian	5 (3.5)	0	–	2 (2.3)	1 (3.7)	–
Other/unknown	19 (13.3)	2 (15.4)	–	14 (16.1)	5 (18.5)	–
BMI (kg/m ²), median (IQR) ^b	29.0 (25.0–34.0)	28.1 (27.0–35.0)	0.976	29.0 (26.0–34.0)	29.3 (26.0–37.0)	0.995
BMI >30 kg/m ² , n (%) ^b	56 (39.2)	4 (30.8)	0.564	35 (40.2)	12 (44.4)	0.950
Diabetes, n (%)	21 (14.7)	2 (15.4)	0.914	9 (10.3)	11 (40.7)	<0.001
Fever, n (%)	26 (18.2)	3 (23.1)	0.629	14 (16.1)	8 (29.6)	0.150
Absence of classic cholelithiasis risk factors (BMI >30 kg/m ² , female, age >40 years), n (%)	9 (6.3)	2 (15.4)	0.217	7 (8.0)	3 (11.1)	0.716
History of cholecystectomy, n (%)	36 (25.2)	3 (23.1)	0.909	18 (20.7)	6 (22.2)	0.953
Peak bilirubin pre-ERCP (mg/dL), median (IQR)	3.1 (1.8–5.2)	5.2 (3.5–6.0)	0.150	4.5 (3.3–7.5)	3.1 (1.9–5.8)	0.064
Noninvasive imaging, n (%)						
Ultrasound	143 (100.0)	13 (100.0)	–	87 (100.0)	27 (100.0)	–
CT	62 (43.4)	4 (30.8)	0.423	37 (42.5)	13 (48.2)	0.746
MRCP	58 (40.6)	5 (38.5)	0.944	6 (6.9)	8 (29.6)	0.002
EUS performed, n (%)	5 (3.5)	0	–	7 (8.0)	3 (12.0)	0.602
Maximum CBD diameter (mm) on noninvasive imaging, median (IQR) ^a	10.0 (8.0–12.0)	9.0 (6.0–10.0)	0.058	7.0 (5.0–10.0)	6.0 (4.0–9.0)	0.616

Data are shown as n (%) or median (IQR).

Abbreviations: BMI, body mass index; CBD, common bile duct; CT, computed tomography; ERCP, endoscopic retrograde cholangiopancreatography; EUS, endoscopic ultrasound; MRCP, magnetic resonance cholangiopancreatography; IQR, interquartile range.

^a For patients who underwent more than one noninvasive imaging modality, the greatest maximum noninvasively-measured CBD diameter was used.

^b Six patients (2.2%) were missing BMI measurements.

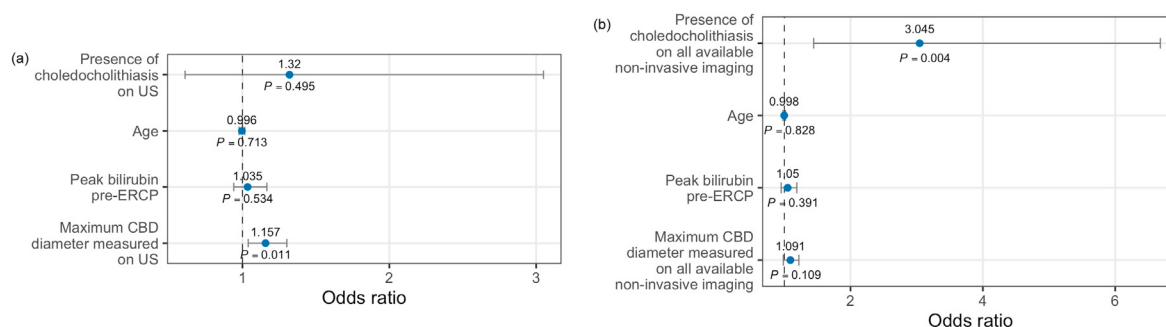


Fig. 2. Multiple logistic regression model with demographic, biochemical, and radiological predictors of ERCP-confirmed choledocholithiasis assessed using (a) US only and (b) all available noninvasive imaging modalities. Abbreviations: CBD, common bile duct; ERCP, endoscopic retrograde cholangiopancreatography; US, ultrasound.

anticipate future growth in this regard and further validation in our population and others. Finally, the formulation of an easy-to-use, online, clinician-oriented application, as developed herein, can be a useful tool to help ascertain the degree of likelihood that a given patient will have ERCP-confirmed choledocholithiasis.

Our study also has some limitations. It was a single-center retrospective study, and although the sample size was comparable to that of other published studies on ERCP-proven choledocholithiasis, the small number of negative outcomes limited the number of predictors that could be included in the multiple logistic

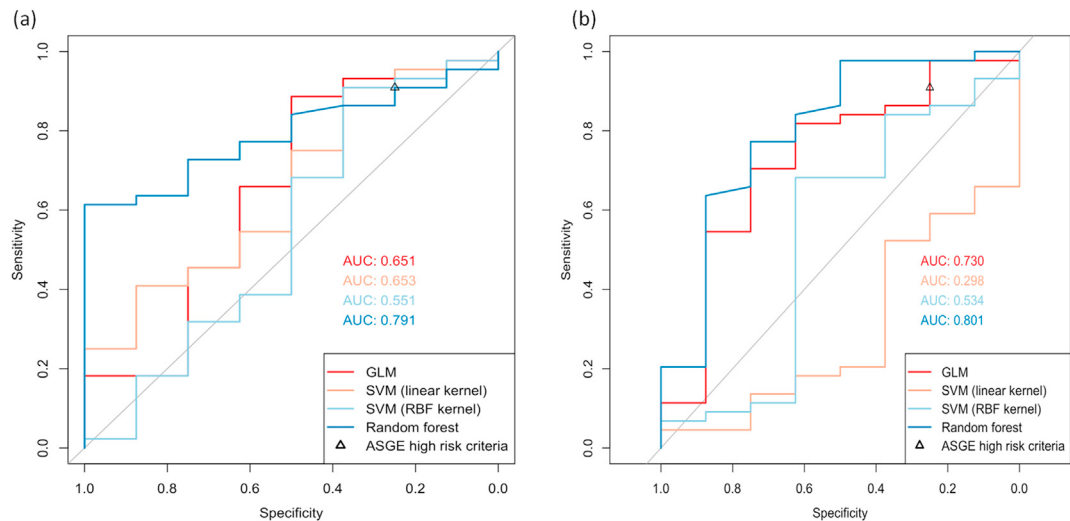


Fig. 3. Receiver operator characteristic curves for MLMs trained to predict the presence of ERCP-confirmed choledocholithiasis. Model fit using predictors assessed on (a) US only and (b) all available noninvasive imaging modalities. AUC could not be calculated for the ASGE high-likelihood criteria because applying the criteria to our dataset does not generate class membership properties. The performance of the ASGE criteria is expressed as a single point here. Abbreviations: ASGE, American Society for Gastroenterology; AUC, area under the receiver operator characteristic curve; ERCP, endoscopic retrograde cholangiopancreatography; GLM, generalized linear model; MLM, machine learning model; RBF, radial basis function; SVM, support vector machine; US, ultrasound.

Table 3
Performance of MLMs at optimal cut-point determined using Youden index and ASGE high-likelihood criteria.

Learning models	Optimal cut-point	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
All available noninvasive imaging modalities:						
GLM	0.886	71.2	70.5	75.0	93.9	31.6
SVM with linear kernel	0.849	19.2	4.6	100.0	100.0	16.0
SVM with radial basis function kernel	0.841	67.3	68.2	62.5	90.9	26.3
Random forest	0.825	76.9	77.3	75.0	94.4	37.5
ASGE high-likelihood criteria	—	80.8	90.9	25.0	87.0	33.3
US only:						
GLM	0.785	82.7	88.6	50.0	90.7	44.4
SVM with linear kernel	0.849	82.7	90.9	37.5	88.9	42.9
SVM with radial basis function kernel	0.844	82.7	90.9	37.5	88.9	42.9
Random forest	0.852	67.3	61.4	100.0	100.0	32.0
Random forest (w/optimal cut point as determined by point closest to (0,1))	0.793	73.1	72.7	75.0	94.1	33.3

Abbreviations: ASGE, American Society for Gastroenterology; GLM, generalized linear model; MLMs, machine learning models; NPV, negative predictive value; PPV, positive predictive value; SVM, support vector machine; US, ultrasound.

regression models. In addition, while endoscopic ultrasound (EUS) is available at our facility, it was infrequently used in our cohort as it was generally not necessary for clinical decision-making (i.e. it would not have changed the management plan of performing ERCP given the high *a priori* suspicion of choledocholithiasis); moreover, while less invasive than ERCP, it is still an invasive technique, whereas our study focus was on noninvasive predictors. Although our MLMs achieved good performance, we expect their ability to identify patients who will or will not have ERCP-confirmed choledocholithiasis to improve as more training data is acquired and features and model parameters are refined. Finally, the study population was largely underserved, uninsured/under-insured, and majority of Hispanic/Latino, which may limit the applicability of this study and the utility of our MLMs in other populations.

5. Conclusions

In conclusion, the random forest MLM trained on age, sex, race/ethnicity, diabetes, fever, BMI, total bilirubin, and maximum CBD diameter and choledocholithiasis assessed on all available noninvasive imaging achieved good sensitivity and specificity (77.3% and 75.0%, respectively). The random forest model trained on the same

features and maximum CBD diameter and presence of choledocholithiasis assessed on US only achieved 61.4% sensitivity and 100.0% specificity. When the random forest models were validated using a cut point which achieved a specificity equal to that of the ASGE high-likelihood criteria, they achieved equal or superior sensitivity (97.7% vs. 90.9%).

Considering clinician suspicion for choledocholithiasis is already quite sensitive, our random forest models, with their high specificity, could be useful for identifying patients who do not need ERCP as a next step. We have made our random forest-based MLM trained on all available imaging modalities freely available as an online application (<https://harrytrieu.shinyapps.io/choledocholithiasisrisk/>) to help guide the decision of whether or not to proceed with ERCP.

Authors' contributions

C. Dalai and J. Azizian contributed equally to this work. C. Dalai and J. Azizian abstracted data, interpreted data analysis findings, and drafted the manuscript. H. Trieu and T. Dong provided statistical analysis, formulated machine learning modeling algorithms, and contributed to drafting the methods, results, and discussion

sections of the manuscript. A. Rajan abstracted data. S. Beaven, F. Chen and J. H. Tabibian provided critical revisions. J. H. Tabibian provided study supervision.

STROBE statement

Manuscript was prepared and revised according to the STROBE statement checklist of items.

Disclosure

An abstract based on this study has been selected as an ASGE Poster of Distinction for presentation during the Digestive Disease Week 2021 ERCP Poster Session on May 23rd.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgements

We thank Stefanie Vassar and Sarmen Hakopian of the Olive View-UCLA Medical Center and Dr. William Raynor for providing statistical planning and guidance for this project. J. H. Tabibian was supported in part by the United States National Center for Advancing Translational Sciences grant UL1 TR000135.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.livres.2021.10.001>.

References

- Peery AF, Dellon ES, Lund J, et al. Burden of gastrointestinal disease in the United States: 2012 update. *Gastroenterology*. 2012;143:1179–1187(e3). <https://doi.org/10.1053/j.gastro.2012.08.002>.
- Russo MW, Wei JT, Thiny MT, et al. Digestive and liver diseases statistics, 2004. *Gastroenterology*. 2004;126:1448–1453. <https://doi.org/10.1053/j.gastro.2004.01.025>.
- Acalovschi M. Cholesterol gallstones: from epidemiology to prevention. *Postgrad Med J*. 2001;77:221–229. <https://doi.org/10.1136/pmj.77.906.221>.
- Bass G, Gilani SN, Walsh TN. Validating the 5Fs mnemonic for cholelithiasis: time to include family history. *Postgrad Med J*. 2013;89:638–641. <https://doi.org/10.1136/postgradmedj-2012-131341>.
- Chapman BA, Wilson IR, Frampton CM, et al. Prevalence of gallbladder disease in diabetes mellitus. *Dig Dis Sci*. 1996;41:2222–2228. <https://doi.org/10.1007/BF02071404>.
- Everhart JE, Khare M, Hill M, Maurer KR. Prevalence and ethnic differences in gallbladder disease in the United States. *Gastroenterology*. 1999;117:632–639. [https://doi.org/10.1016/s0016-5085\(99\)70456-7](https://doi.org/10.1016/s0016-5085(99)70456-7).
- Sturdik I, Krajcovicova A, Leskova Z, et al. P750 Prevalence and risk factors of cholelithiasis in patients with Crohn's disease. *Journal of Crohn's and Colitis*. 2017;11(suppl_1):S465–S466. <https://doi.org/10.1093/ecco-jcc/jjx002.873>.
- Attasaranya S, Fogel EL, Lehman GA. Cholelithiasis, ascending cholangitis, and gallstone pancreatitis. *Med Clin North Am*. 2008;92:925–960. <https://doi.org/10.1016/j.mcna.2008.03.001>.
- Onken JE, Brazier SR, Eisen GM, et al. Predicting the presence of cholelithiasis in patients with symptomatic cholelithiasis. *Am J Gastroenterol*. 1996;91:762–767.
- Ko CW, Lee SP. Epidemiology and natural history of common bile duct stones and prediction of disease. *Gastrointest Endosc*. 2002;56(6 Suppl):S165–S169. <https://doi.org/10.1067/mge.2002.129005>.
- Freitas ML, Bell RL, Duffy AJ. Cholelithiasis: evolving standards for diagnosis and management. *World J Gastroenterol*. 2006;12:3162–3167. <https://doi.org/10.3748/wjg.v12.i20.3162>.
- Huang RJ, Barakat MT, Girotra M, Banerjee S. Practice patterns for cholecystectomy after endoscopic retrograde cholangiopancreatography for patients with cholelithiasis. *Gastroenterology*. 2017;153:762–771(e2). <https://doi.org/10.1053/j.gastro.2017.05.048>.
- ASGE Standards of Practice Committee, Maple JT, Ben-Menachem T, et al. The role of endoscopy in the evaluation of suspected cholelithiasis. *Gastrointest Endosc*. 2010;71:1–9. <https://doi.org/10.1016/j.gie.2009.09.041>.
- ASGE Standards of Practice Committee, Buxbaum JL, Abbas Fehmi SM, et al. ASGE guideline on the role of endoscopy in the evaluation and management of cholelithiasis. *Gastrointest Endosc*. 2019;89:1075–1105(e15). <https://doi.org/10.1016/j.gie.2018.10.001>.
- Yu CY, Roth N, Jani N, et al. Dynamic liver test patterns do not predict bile duct stones. *Surg Endosc*. 2019;33:3300–3313. <https://doi.org/10.1007/s00464-018-06620-x>.
- Adams MA, Hosmer AE, Wamsteker EJ, et al. Predicting the likelihood of a persistent bile duct stone in patients with suspected cholelithiasis: accuracy of existing guidelines and the impact of laboratory trends. *Gastrointest Endosc*. 2015;82:88–93. <https://doi.org/10.1016/j.gie.2014.12.023>.
- Suarez AL, LaBarre NT, Cotton PB, Payne KM, Coté GA, Elmunzer BJ. An assessment of existing risk stratification guidelines for the evaluation of patients with suspected cholelithiasis. *Surg Endosc*. 2016;30:4613–4618. <https://doi.org/10.1007/s00464-016-4799-8>.
- He H, Tan C, Wu J, et al. Accuracy of ASGE high-risk criteria in evaluation of patients with suspected common bile duct stones. *Gastrointest Endosc*. 2017;86:525–532. <https://doi.org/10.1016/j.gie.2017.01.039>.
- Abboud PA, Malet PF, Berlin JA, et al. Predictors of common bile duct stones prior to cholecystectomy: a meta-analysis. *Gastrointest Endosc*. 1996;44:450–455. [https://doi.org/10.1016/s0016-5107\(96\)70098-6](https://doi.org/10.1016/s0016-5107(96)70098-6).
- Kama NA, Atli M, Doganay M, Kologlu M, Reis E, Dolapci M. Practical recommendations for the prediction and management of common bile duct stones in patients with gallstones. *Surg Endosc*. 2001;15:942–945. <https://doi.org/10.1007/s00464-001-0005-7>.
- Prat F, Meduri B, Ducot B, Chiche R, Salimbeni-Bartolini R, Pelletier G. Prediction of common bile duct stones by noninvasive tests. *Ann Surg*. 1999;229:362–368. <https://doi.org/10.1097/0000658-199903000-00009>.
- Tse F, Barkun JS, Barkun AN. The elective evaluation of patients with suspected cholelithiasis undergoing laparoscopic cholecystectomy. *Gastrointest Endosc*. 2004;60:437–448. [https://doi.org/10.1016/s0016-5107\(04\)01457-9](https://doi.org/10.1016/s0016-5107(04)01457-9).
- Yang MH, Chen TH, Wang SE, et al. Biochemical predictors for absence of common bile duct stones in patients undergoing laparoscopic cholecystectomy. *Surg Endosc*. 2008;22:1620–1624. <https://doi.org/10.1007/s00464-007-9665-2>.
- Parra Pérez V, Vargas Cárdenas G, Astete Benavides M, et al. Cholelithiasis predictors in high-risk population subjected to endoscopic retrograde pancreatocolangiography at "Hospital Nacional Arzobispo Loayza. *Rev Gastroenterol Peru*. 2007;27:161–171.
- Al-Jiffry BO, Khayat S, Abdeen E, Hussain T, Yassin M. A scoring system for the prediction of cholelithiasis: a prospective cohort study. *Ann Saudi Med*. 2016;36:57–63. <https://doi.org/10.5144/0256-4947.2016.57>.
- Khoury T, Kadhah A, Mahamid M, Mari A, Sbeit W. Bedside score predicting retained common bile duct stone in acute biliary pancreatitis. *World J Clin Cases*. 2020;8:1414–1423. <https://doi.org/10.12998/wjcc.v8.i8.1414>.
- Miura F, Okamoto K, Takada T, et al. Tokyo Guidelines 2018: initial management of acute biliary infection and flowchart for acute cholangitis. *J Hepatobiliary Pancreat Sci*. 2018;25:31–40. <https://doi.org/10.1002/jhbp.509>.
- About Us - More DHS. <https://dhs.lacounty.gov/more-dhs/about-us/>. Accessed August 24, 2020.
- Unal I. Defining an optimal cut-point value in ROC analysis: an alternative approach. *Comput Math Methods Med*. 2017;3762651. <https://doi.org/10.1155/2017/3762651>.
- ASGE Standards of Practice Committee, Buxbaum JL, Abbas Fehmi SM, et al. ASGE guideline on the role of endoscopy in the evaluation and management of cholelithiasis. *Gastrointest Endosc*. 2019;89:1075–1105(e15). <https://doi.org/10.1016/j.gie.2018.10.001>.
- Jovanovic P, Salkic NN, Zerem E. Artificial neural network predicts the need for therapeutic ERCP in patients with suspected cholelithiasis. *Gastrointest Endosc*. 2014;80:260–268. <https://doi.org/10.1016/j.gie.2014.01.023>.
- Uchiyama K, Onishi H, Tani M, et al. Long-term prognosis after treatment of patients with cholelithiasis. *Ann Surg*. 2003;238:97–102. <https://doi.org/10.1097/01.sla.0000077923.38307.84>.