Revista de Gastroenterología de México xxx (xxxx) xxx–xxx

REVISTA DE GASTROENTEROLOGÍA DE MÉXICO
www.elsevier.es/rgmx

ORIGINAL ARTICLE

An App model that utilizes a logistic regression algorithm for predicting choledocholithiasis: A prospective clinical trial

F. García-Villarreal[a], L.M. Torres-Treviño[b], C. Herrera-Figueroa[a], J.O. Jáquez-Quintana[a], A.A. Garza-Galindo[a], C.A. Cortez-Hernández[a], D. García-Compeán[a], R.A. Jiménez-Castillo[a], H.J. Maldonado-Garza[a], J.A. González-González[a,*]

[a] *Departamento de Medicina Interna, Servicio de Gastroenterología y Endoscopia Digestiva, Hospital Universitario ''Dr. José E. González'', Universidad Autónoma de Nuevo León, Monterrey, Nuevo León, Mexico*
[b] *Facultad de Ingeniería Mecánica y Eléctrica, Universidad Autónoma de Nuevo León, Monterrey, Nuevo León, Mexico*

Received 18 March 2024; accepted 23 May 2024

KEYWORDS
Choledocholithiasis;
Logistic regression;
Intermediate risk

Abstract
Introduction and aim: The diagnostic yield of the current criteria for assigning the risk of choledocholithiasis (CL) is inaccurate. The aim of our work was to develop a logistic regression model for predicting CL diagnosis in patients catalogued as either intermediate or high risk for CL, according to the criteria of the American Society for Gastrointestinal Endoscopy (ASGE).
Material and methods: We conducted an analytic, observational, cross-sectional study for evaluating the diagnostic yield of a logistic regression model in adults with intermediate or high risk for CL. A receiver operating characteristic (ROC) curve analysis was done to determine the best cutoff point for predicting the diagnosis of CL. Endoscopic retrograde cholangiopancreatography (ERCP) was utilized as the gold standard for diagnosing CL.
Results: A total of 148 patients suspected of presenting with CL were studied. In our cohort, 71 had immediate risk and 77 had high risk. CL diagnosis was confirmed in 102 patients (69%). Our model showed an area under the curve (AUC) of 0.68. In patients with an intermediate risk for CL, the AUC value was 0.72 and the positive predictive value (PPV) was 70%. In patients with a high risk for CL, the AUC value was 0.78 and the PPV was 89%.
Conclusion: Our model appears to better predict the diagnosis of CL than the ASGE criteria for patients with an intermediate or high risk for the disease. Our model can guide clinical decisions in patients with suspected CL.

* Corresponding author at: Hospital Universitario ''Dr. José Eleuterio González'', Av. Madero y Gonzalitos S/N Mitras Centro, Monterrey. Tel.: +8441019421.
E-mail address: jalbertogastro@gmail.com (J.A. González-González).

PALABRAS CLAVE
Coledocolitiasis;
Regresión logística;
Riesgo intermedio

Un modelo tipo App que utiliza un algoritmo de regresión logística para predecir coledocolitiasis. Un ensayo clínico prospectivo

Resumen
Introducción y objetivo: El rendimiento diagnóstico de los criterios actuales para asignar el riesgo de coledocolitiasis (CL) es impreciso. El objetivo de nuestro trabajo fue desarrollar un modelo de regresión logística para predecir el diagnóstico de CL en pacientes catalogados como riesgo intermedio y alto de CL, según los criterios de la Sociedad Americana de Endoscopia Gastrointestinal (ASGE).
Material y métodos: Realizamos un estudio transversal, observacional y analítico para evaluar el rendimiento diagnóstico de un modelo de regresión logística en adultos con riesgo intermedio y alto de CL. Se realizó un análisis de curva característica operativa del receptor (COR) para determinar el mejor punto de corte para predecir el diagnóstico de CL. Se utilizó la colangiopancreatografía retrógrada endoscópica (CPRE) como estándar de oro para el diagnóstico de CL.
Resultados: Se estudiaron 148 pacientes con sospecha de CL. En nuestra cohorte, 71 presentaron riesgo intermedio y 77 riesgo alto. El diagnóstico de CL se confirmó en 102 pacientes (69%). En la cohorte, nuestro modelo mostró un área bajo la curva (ABC) de 0.68. En pacientes con riesgo intermedio de CL, el valor de ABC fue de 0.72 y el valor predictivo positivo (VPP) fue del 70%. En pacientes con riesgo alto de CL, el valor de ABC fue de 0.78 y el VPP de 89%.
Conclusión: Nuestro modelo parece predecir mejor el diagnóstico de CL que los criterios de la ASGE para pacientes de riesgo intermedio y alto. Nuestro modelo puede orientar las decisiones clínicas en pacientes con sospecha de CL.

Introduction and aim

Choledocholithiasis (CL) is a frequent cause of extrahepatic bile duct obstruction that can be diagnosed in up to 15% of patients with cholecystolithiasis.[1,2] Given that there can be complications due to the presence of CL (acute cholangitis, acute pancreatitis), its early diagnosis and treatment is of vital importance. Diagnosis is currently based on clinical, radiologic, and laboratory parameters. Considering body mass index, age, sex, the presence of bile duct dilation, and liver function tests, the American Society for Gastrointestinal Endoscopy (ASGE) and the European Society for Gastrointestinal Endoscopy (ESGE) developed predictive scales for CL.[2,3]

The ASGE criteria classify patients into low risk, intermediate risk, and high risk for CL, with a probability of 10%, 10–50%, and >50%, respectively. However, accuracy varies from 40% to 85% for high risk and 30%–40% for intermediate risk.[2] Due to the wide variability in the diagnostic yield of the abovementioned criteria in clinical practice, other methods utilized for predicting CL diagnosis, such as logistic regression and symbolic regression, have been implemented.[4]

The gold standard for diagnosing CL is endoscopic retrograde cholangiopancreatography (ERCP), with 94% sensitivity and 100% specificity, but this method is not exempt from serious complications and should be performed mainly in the context of therapeutic indications.[5]

Therefore, we evaluated the diagnostic accuracy of our application through a logistic regression model utilized in patients with an intermediate or high risk for CL.

Material and methods

An analytic, observational, cross-sectional study was conducted at a single center within the time frame of February 1, 2022, and February 1, 2023. As background in 2021, a logistic regression model was obtained through artificial intelligence for predicting the diagnosis of CL. Initially, the model was validated in a retrospective cohort of patients diagnosed with intermediate or high risk for CL that underwent ERCP during 2020 at the gastroenterology service of our hospital. From that precedent, the model was prospectively applied to hospitalized patients with intermediate or high risk for CL, to evaluate its diagnostic yield. Only patients with the presence of a stone during ERCP were considered to have a definitive diagnosis of CL. The model was developed by one of the authors (LM T-T), who is a Doctor in Artificial Intelligence. An established cutoff point of $\geq$0.6 for discerning whether the model predicted a positive result (the presence of a stone in ERCP) or a negative cutoff point <0.6 (no stone in ERCP) was determined. The App model was applied at the bedside of 148 patients admitted to our hospital. Inclusion criteria were patients $\geq$18 years of age, with clinical suspicion, and with intermediate or high risk of CL through laboratory or imaging tests. Exclusion criteria were patients at low risk for CL by the ASGE classification (considering that those patients did not require invasive studies before cholecystectomy), patients with previous cholecystectomy, previous ERCP or biliary surgery, pregnant women, patients with cirrhosis of the liver, clinical suspicion of cholangitis, ASA III, and patients that did

Age	Sex (1 for M; 2 for F)	Time (0,24,48) hours	AST	ALT	Total bilirubin	Direct bilirubin	ALP	CBD diameter (mm)	CBD stone (probability)	CBD stone (optimized)	ERCP results (gold standard)
55	1	0	306	112	21.1	12.8	282	2	1.05%	NO	NO
39	2	0	107	137	12.8	7.7	281	14	52.83%	NO	NO
24	2	0	28	162	1.2	1	284	9	95.86%	YES	NO
34	2	0	1099	648	4.5	3.4	204	10	86.12%	YES	YES
74	2	0	36	39	2.9	1.7	154	1	38.67%	NO	YES
30	2	0	91	97	26.8	17.3	280	18	18.82%	NO	YES
27	1	0	120	221	10.9	6.8	183	17	63.02%	NO	YES
84	2	0	291	143	2	1.2	449	19	47.12%	NO	YES
85	2	0	137	146	7.9	5.2	756	19	43.63%	NO	YES
28	2	0	356	430	5.3	3.4	185	11	95.46%	YES	YES
55	2	0	70	93	1.2	0.8	443	7	75.82%	NO	YES
36	2	0	19	13	5.5	3.1	207	7	62.07%	NO	NO
37	2	0	72	120	21.8	16.4	277	1	45.88%	NO	YES
25	2	0	319	689	4.9	3.2	242	7	99.48%	YES	YES
37	2	0	733	510	3.6	2.7	249	11	90.63%	YES	NO
85	2	0	153	176	2	1.2	256	1	47.78%	NO	YES
25	2	0	242	226	7.1	4.8	555	12	87.84%	YES	YES
51	1	0	392	642	13.3	8	268	11	73.54%	NO	YES
74	2	0	77	36	1.7	1	505	18	57.51%	NO	YES
36	1	0	216	496	5.8	4	250	8	93.17%	YES	NO
42	1	0	144	293	8.6	5.3	168	13	65.11%	NO	YES
32	2	0	104	235	1	0.4	173	3	93.27%	YES	YES
73	1	0	83	158	7	4.5	172	1	19.49%	NO	NO
67	1	0	170	156	13.2	10.7	698	14	31.90%	NO	YES

Figure 1 Logistic regression model used to evaluate the clinical, laboratory, and imaging characteristics of 148 patients with intermediate or high risk of choledocholithiasis and to predict the diagnosis of choledocholithiasis.
ALP: alkaline phosphatase; ALT: alanine aminotransferase; AST: aspartate aminotransferase; CBD: common bile duct; F: female; M: male.

not complete their follow-up at our hospital. Laboratory tests were carried out at hospital admission and at 24 and 48 h after hospitalization. The laboratory tests performed upon hospital admission were utilized to classify patients according to the ASGE criteria.

Logistic regression model

Our logistic regression model included age, sex, time of ERCP to hospital admission, AST, ALT, alkaline phosphatase, total bilirubin, and the diameter of the common bile duct measured by abdominal ultrasound, for their analysis (Fig. 1). Through the logistic regression model, the contribution of each variable in predicting CL was evaluated and utilized for the control of other confounding factors.

Three experienced endoscopists carried out all the ERCPs in our study. The results of the model were not taken into consideration for making medical decisions.

Statistical analysis

Frequencies (%), medians (q25-q75), or means ± standard deviation were reported in the descriptive analysis. A receiver operating characteristic (ROC) curve analysis was done to establish the best cutoff point for predicting the diagnosis of CL. Model sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) were reported. The Python program was utilized for the statistical analysis.

Ethical considerations

The protocol was reviewed and approved by the ethics committee of our institution (GA22-00005). Written statements of informed consent were obtained before participation in the study.

Results

A total of 148 patients with either intermediate or high risk for CL were recruited. Median age was 43 years (range:16—85) and 110 (74%) of the patients were women. As mean values, AST was 214.37 ± 184.25 U/l, ALT 288.13 ± 228.17 U/l, alkaline phosphatase 290.49 ± 174.57 U/l, total bilirubin 6.87 ± 5.13 mg/dl, and the diameter of the common bile duct 9.2 ± 5.4 mm. In our cohort, 71 (48%) patients were classified as having an intermediate risk for CL and 77 (52%) as having a high risk for CL. Table 1 shows all laboratory test results. ERCP was performed on 125 (85%) patients ≥48 h from hospital admission. CL diagnosis through the identification of a stone by ERCP was made in 102 (69%) patients. ECRP demonstrated the presence of a stone in the common bile duct in 26 (36.6%) patients at intermediate risk for CL and in 41 (53.2%) patients at high risk for CL.

Table 1 Clinical and laboratory characteristics of 148 patients with clinical suspicion of choledocholithiasis upon their admission to the *Hospital Universitario ''Dr. José Eleuterio González''*.

Number of patients	148
Age (years)	43 (16−85)
Sex (n, %)	148
Male	38 (26)
Female	110 (74)
Time of ERCP performance with respect to hospital admission	
<48 h	23 (16)
>48 h	125 (85)
AST (IU/l)	214.37 ± 184.25
ALT (IU/l)	288.13 ± 228.17
Total bilirubin (mg/dl)	6.87 ± 5.13
Direct bilirubin (mg/dl)	4.43 ± 3.51
Alkaline phosphatase (IU/l)	290.49 ± 174.57
Common bile duct (mm)	9.28 ± 5.4
Probability according to ASGE criteria	
High risk	77 (52)
Intermediate risk	71 (48)
Choledocholithiasis present in the prediction model (n, %)	148
Yes	81 (55)
No	67 (45)
ERCP-confirmed choledocholithiasis (n, %)	148
Yes	102 (69)
No	46 (31)

ALT: alanine aminotransferase; ASGE: American Society for Gastrointestinal Endoscopy; ST: aspartate aminotransferase; ERCP: endoscopic retrograde cholangiopancreatography.

Prediction of choledocholithiasis made by our logistic regression model

The analysis of our model in the cohort revealed an area under the ROC curve (AUC) value of 0.68, indicating moderate predictive capaciVty. The AUC value for patients with intermediate risk was 0.72, with 65% sensitivity, 65% specificity, 70% PPV, 59% NPV, and 71% accuracy. In high-risk patients, the AUC value was 0.78, with 66% sensitivity, 67% specificity, 89% PPV, 32% NPV, and 89% accuracy (Fig. 2).

Discussion

The development of different noninvasive, economic tools (neural networks, machine learning) for predicting the presence of CL is vitally important, given that current methods (magnetic resonance imaging, ERCP) for its diagnosis are costly, not exempt from risks for the patient, and not always widely available. Unfortunately, the development of CL prediction tools continues to produce heterogeneous results.

Currently, the ASGE criteria define patients as having low, intermediate, and high risks for CL, but when applied to different populations, these definitions vary greatly. Matt Ridley expressed this concept in his book, *The Agile Gene*.[6] He wrote that persons are similar because they are different and different because they are similar. In such a context, the applicability of any score or criterion becomes quite difficult.

In 2017, Narváez et al. applied the ASGE criteria to patients at the *Hospital Universitario ''Dr. José Eleuterio González''*. Those authors reported a diagnostic accuracy of CL in the high-risk patients of 59%, with 85% sensitivity and 24% specificity. In the intermediate-risk patients, accuracy was 41%, with 14% sensitivity and 75% specificity, indicating unnecessary ERCP in almost half the patients.[7] With those data in mind, we developed a logistic regression model to be prospectively applied as an App at the bedside of the patient suspected of presenting with CL. In our model, the PPV was 89% and accuracy was 89% for high-risk patients. Our results were similar to those reported by Dalai et al., who analyzed 270 patients at high risk for CL, utilizing artificial intelligence. Those authors described 91% sensitivity, 25% specificity, 87% PPV, 33% NPV, and 81% accuracy.[8] Additionally, in 2014, a published prospective study by Jovanovic et al. utilized an artificial neuronal network to determine

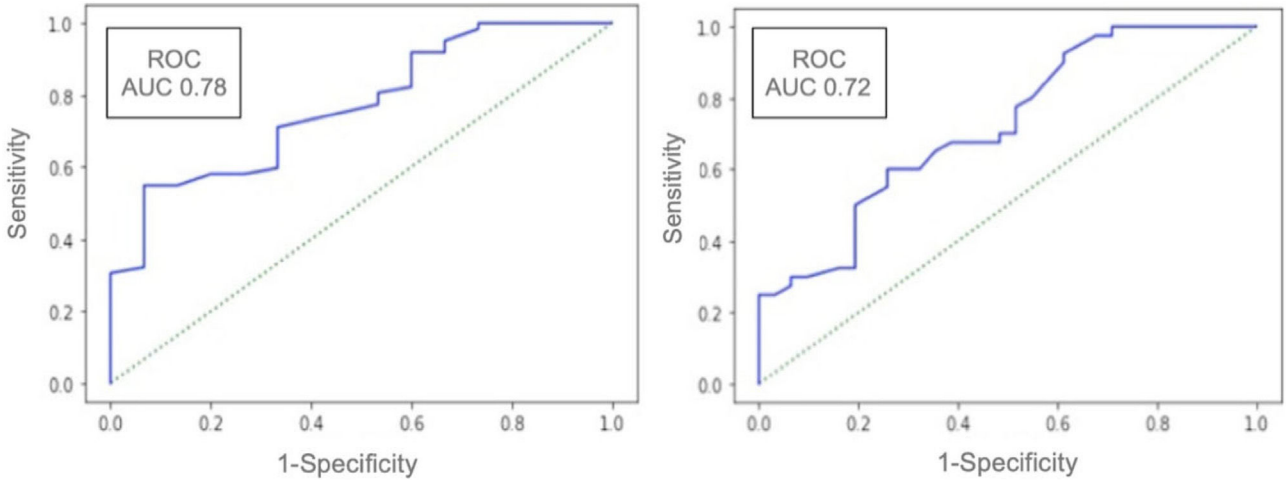

Figure 2 ROC curve analysis to predict the definitive diagnosis of choledocholithiasis, in high and intermediate-risk patients, respectively.

Table 2 Comparison of the diagnostic yield of our logistic regression model and the different statistical models presented in the literature, in patients at high risk for choledocholithiasis.

	Sensitivity	Specificity	PPV	NPV
Our predictive model (logistic regression)	66%	67%	89%	32%
Narváez et al.[7] (ASGE criteria)	86%	24%	60%	56%
Dalai et al.[8] (mechanized learning)	91%	25%	87%	33%
He et al.[10] (ASGE criteria)	70%	74%	64%	79%
Jagtap et al.[11] (ASGE criteria)	75%	97%	90%	91%
Herrera et al.[12] (symbolic regression)	61%	85%	87%	57%
Ovalle et al.[13] (ASGE criteria)	69%	52%	79%	38%

NPV: negative predictive value; PPV: positive predictive value.

Table 3 Comparison of diagnostic yield of our logistic regression model and the different statistical models presented in the literature in patients at intermediate risk for choledocholithiasis.

	Sensitivity	Specificity	PPV	NPV
Our predictive model (logistic regression)	65%	65%	70%	59%
Narváez et al.[7] (ASGE criteria)	14%	76%	44%	40%
Jagtap et al.[11] (ASGE criteria)	24%	20%	10%	42%
Herrera et al.[12] (symbolic regression)	73%	77%	55%	88%

NPV: negative predictive value; PPV: positive predictive value.

Table 4 Classification criteria for the risk of choledocholithiasis, according to the American Society for Gastrointestinal Endoscopy (ASGE) and its suggested treatment.[1]

Risk classification	Clinical criteria	Treatment
High risk	Choledocholithiasis present in a noninvasive imaging study or cholangitis or total bilirubin >4 and dilated common bile duct	ERCP
Intermediate risk	Altered liver function tests or age >55 years or dilated common bile duct (> 6 mm with gallbladder *in situ*)	Magnetic resonance cholangiopancreatography or endoscopic ultrasound
Low risk	None of the above (symptomatic cholecystolithiasis with none of the abovementioned factors)	Cholecystectomy

Source: this table is based on the guideline by Buxbaum et al.[2]
ERCP: endoscopic retrograde cholangiopancreatography.

the risk for ERCP indication in patients with suspected CL. The AUC value was 0.88 (95% CI 83–93%), significantly higher than that of our model, but showed a similar PPV of 92% in patients at high risk of CL. That model correctly classified 92% of patients who needed an ERCP.[9]

A study by Steinway et al., utilizing a machine learning-based method for predicting CL in 1,378 patients, compared the diagnostic accuracy of the gradient boosting machine-learning method versus the 2019 ASGE and ESGE criteria, finding accuracy of 71%, 62%, and 62%, respectively, results not significantly different from ours.[4]

In a study on 1,171 patients, He et al. reported that the specificity of the ASGE criteria for CL in high-risk patients was 74% (95% CI 72–77%) and the PPV was 64% (95% CI 61–675). Even though the high-risk criteria demonstrated a probability above 50% of presenting with CL, more than one-third of the patients underwent diagnostic ERCP.[10] In a retrospective article by Jagtap et al. that compared the ASGE and ESGE criteria for CL in high-risk patients, a higher PPV was obtained with the ESGE criteria.[11]

Table 2 shows the results of a group of studies on patients at high risk of CL, describing sensitivity from 61% to 91%. In addition, specificity varied from 24% to 97%. All studies were conducted on populations with a different prevalence of choledocholithiasis.[7,8,10–13]

Our predictive model applied to the patients with intermediate risk for CL had 65% sensitivity, 65% specificity, 70% PPV, 59% NPV, and 71% accuracy for CL diagnosis. Our results are similar to those reported in the literature (Table 3).[7,11,12]

Most of the prediction models reviewed in the literature have good diagnostic yield in patients at high risk for CL. These data can help confirm the indication for therapeutic ERCP or the necessity for bile duct examination (Table 4).

Nevertheless, the medical decision in patients in the intermediate-risk group continues to be a theme of interest because the PPV of our model in that group of patients was

70%, which was a marginal value, similar to that of several other studies. Therefore, an additional imaging study, such as magnetic resonance imaging or endoscopic ultrasound, is still necessary for confirming the diagnosis of CL.[14]

It should be highlighted that our logistic regression model showed improvement in diagnostic accuracy, with respect to the ASGE criteria. However, evaluating its performance in a different population is necessary because of the high prevalence variability of bile duct stones.

The limitations of our study include the fact that it was conducted at a single center, ERCP was performed $\geq$ 48 h from hospital admission, we did not have laboratory test data from the same day as the ERCP, the majority of our patients were young women with a mean age under 40 years, and our sample size was based on consecutive sampling of patients with intermediate or high risk of CL during one year. Most likely, the diagnostic yield of our model would be more statistically robust if the 3 CL risk groups had been evaluated. Nevertheless, we only selected patients at intermediate or high risk for CL, because in low-risk patients ERCP is not needed and cholecystectomy is indicated.

The strong point of our article is that the model was developed utilizing adequate methodology and it was prospectively applied for diagnosing CL. All ERCPs were performed by expert endoscopists, and the model was applied at the patient's bedside.

Conclusion

Our logistic expression model showed an improvement in diagnostic accuracy, with respect to the ASGE criteria. We believe our findings can be useful for guiding the physician in his/her clinical decision-making in patients with suspected choledocholithiasis.

Financial disclosure

No financial support was received in relation to this study/article.

Declaration of competing interest

The authors declare that there is no conflict of interest.

References

1. Zouki J, Sidhom D, Bindon R, et al. Choledocholithiasis: a review of management and outcomes in a regional setting. Cureus. 2023;15:e50223, http://dx.doi.org/10.7759/cureus.50223.
2. Buxbaum JL, Fehmi SMA, Sultan S, et al. ASGE guideline on the role of endoscopy in the evaluation and management of choledocholithiasis. Gastrointest Endosc. 2019;89:1075–105.e15, http://dx.doi.org/10.1016/j.gie.2018.10.001.
3. Manes G, Paspatis G, Aabakken L, et al. Endoscopic management of common bile duct stones: European Society of Gastrointestinal Endoscopy (ESGE) guideline. Endoscopy. 2019;51:472–91, http://dx.doi.org/10.1055/a-0862-0346.
4. Steinway SN, Tang B, Telezing J, et al. A machine learning-based choledocholithiasis prediction tool to improve ERCP decision making: a proof-of-concept study. Endoscopy. 2023;56:165–71, http://dx.doi.org/10.1055/a-2174-0534.
5. Gurusamy KS, Giljaca V, Takwoingi Y, et al. Endoscopic retrograde cholangiopancreatography versus intraoperative cholangiography for diagnosis of common bile duct stones. Cochrane Database Syst Rev. 2015:CD010339, http://dx.doi.org/10.1002/14651858.CD010339.pub2.
6. Ridley M. The agile gene: How nature turns on nurture. New York: Perennial; 2004. p. 352. Reprint edición.
7. Narváez-Rivera RM, González-González JA, Monreal-Robles R, et al. Accuracy of ASGE criteria for the prediction of choledocholithiasis. Rev Esp Enferm Dig. 2016;108:309–14, http://dx.doi.org/10.17235/reed.2016.4212/2016.
8. Dalai C, Azizian JM, Trieu H, et al. Machine learning models compared to existing criteria for noninvasive prediction of endoscopic retrograde cholangiopancreatography-confirmed choledocholithiasis. Liver Res. 2021;5:224–31, http://dx.doi.org/10.1016/j.livres.2021.10.001.
9. Jovanovic P, Salkic NN, Zerem E. Artificial neural network predicts the need for therapeutic ERCP in patients with suspected choledocholithiasis. Gastrointest Endosc. 2014;80:260–8, http://dx.doi.org/10.1016/j.gie.2014.01.023.
10. 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;86:525–32, http://dx.doi.org/10.1016/j.gie.2017.01.039.
11. Jagtap N, Yashavanth H, Tandan M, et al. Clinical utility of ESGE and ASGE guidelines for prediction of suspected choledocholithiasis in patients undergoing cholecystectomy. Endoscopy. 2020;52:569–73, http://dx.doi.org/10.1055/a-1117-3451.
12. Herrera-Figueroa CA. Modelo de regresión simbólica para le predicción de necesidad de CPRE en pacientes con sospecha de colédocolitiasis: validación prospectiva. UANL; 2021 http://eprints.uanl.mx/id/eprint/20595
13. Ovalle-Chao C, Guajardo-Nieto D, Elizondo-Pereo R. Performance of the predictive criteria of the American Society for Gastrointestinal Endoscopy in the diagnosis of choledocholithiasis at a secondary care public hospital in the State of Nuevo León, Mexico. Rev Gastroenterol Mex (Engl Ed). 2022;88:322–32, http://dx.doi.org/10.1016/j.rgmxen.2022.06.005.
14. Afzalpurkar S, Giri S, Kasturi S, et al. Magnetic resonance cholangiopancreatography versus endoscopic ultrasound for diagnosis of choledocholithiasis: an updated systematic review and meta-analysis. Surg Endosc. 2023;37:2566–73, http://dx.doi.org/10.1007/s00464-022-09744-3.