



HOSPITAL GENERAL DE ZONA 3 DE AGUASCALIENTES

UNIVERSIDAD AUTONOMA DE AGUASCALIENTES

**MODELO PREDICTOR DE COLEDOCOLITIASIS EN
PACIENTES DEL ÓRGANO DE OPERACIÓN
ADMINISTRATIVA DESCONCENTRADA EN
AGUASCALIENTES**

**Tesis presentada por Ariadna Beltran Rico
para obtener el grado de especialista en cirugía general**

**Dra. Ivonne González Espinosa
TUTORA**

**Dr. Rogelio Salinas Gutiérrez
COTUTOR**

**Dr. José Luis Bizueto Monroy
COTUTOR**

Aguascalientes, Ags; a 29 de noviembre del 2025



Gobierno de
México



INSTITUTO MEXICANO DEL SEGURO SOCIAL
SEGURIDAD Y SOLIDARIDAD SOCIAL



Dirección de Prestaciones Médicas

Unidad de Educación e Investigación
Coordinación de Investigación en Salud
División de Innovación y Regulación de la
Investigación en Salud
Comité de Ética en Investigación

Ciudad de México, a 30 de junio de 2025.

Dra. Ivonne González Espinosa

Investigador Responsable
Coordinación Clínica de Cirugía
Hospital General de Zona 3
Presente

En relación al protocolo titulado: "MODELO PREDICTOR DE COLEDOLITIASIS EN PACIENTES DEL ÓRGANO DE OPERACIÓN ADMINISTRATIVA DESCONCENTRADA EN AGUASCALIENTES", con número de registro **2024-785-056**, el Comité de Ética en Investigación **CONBIOÉTICA-09-CEI-009-20160601**, **revisó y aprobó** la solicitud de reaprobación anual del proyecto por el periodo del 25 de junio de 2025 al 24 de junio de 2026. A su vez, se aprueba el cronograma actualizado.

Sin otro particular, reciba un cordial saludo.

Atentamente

Dra. Adriana Leticia Valdez González

Presidenta
Comité de Ética en Investigación
Coordinación de Investigación en Salud
Centro Médico Nacional Siglo XXI

Dr. José Efrén Israel Grijalva Otero

Presidente del Comité de Investigación
Coordinación de Investigación en Salud
Centro Médico Nacional Siglo XXI
COFEPRIS CI17 CI 09 015 006

RBB/mcjbm/
F-CNIC-2024-075



2025
Año de
La Mujer
Indígena

C.M.N Siglo XXI, Ave. Cuauhtémoc No. 330, Piso 4 Edificio Bloque B, Anexo a la Unidad de Congresos, Col. Doctores, Alcaldía Cuauhtémoc, C. P. 06720, Ciudad de México, Tel. (55) 5627 6900, Ext. 21230, www.imss.gob.mx



GOBIERNO DE
MÉXICO



DIRECCIÓN DE PRESTACIONES MÉDICAS
Unidad de Educación e Investigación
Coordinación de Investigación en Salud

Dictamen de Aprobación

Jueves, 10 de octubre de 2024

Ref. 09-B5-61-2800/202400/

Doctor (a) IVONNE GONZÁLEZ ESPINOSA
COOR CLINICA DE CIRUGIA, HOSP GRAL ZONA 3
Aguascalientes

Presente:

Informo a usted que el protocolo titulado: **MODELO PREDICTOR DE COLEDOCOLITIASIS EN PACIENTES DEL ÓRGANO DE OPERACIÓN ADMINISTRATIVA DESCONCENTRADA EN AGUASCALIENTES**, fue sometido a la consideración de este Comité Nacional de Investigación Científica.

Los procedimientos propuestos en el protocolo cumplen con los requerimientos de las normas vigentes, con base en las opiniones de los vocales, del Comité de Ética en Investigación, Comité de Investigación, y Comité de Bioseguridad del Comité Nacional de Investigación Científica del IMSS, por lo que se ha emitido el dictamen de **APROBADO**, con número de registro: R-2024-785-056.

De acuerdo a la normatividad vigente, deberá enviar el "Informe de Seguimiento Técnico" a este Comité en el mes de junio de cada año, en donde deberá describir el desarrollo del proyecto a su cargo.

Este dictamen sólo tiene vigencia de un año. Por lo que en caso de ser necesario requerirá solicitar una re aprobación al Comité de Ética en Investigación del Comité Nacional de Investigación Científica, al término de la vigencia del mismo.

Atentamente,

Dr. José Efrén Israel Grijalva Otero
Presidente del
Comité Nacional de Investigación Científica

Anexo comentarios:
Se anexa dictamen
SNN/ . F-CNIC-2024-075



CARTA DE APROBACIÓN DE TRABAJO DE TESIS

AGUASCALIENTES, AGS, A DE 20 DE SEPTIEMBRE DE 2025

**COMITÉ DE INVESTIGACIÓN Y ÉTICA EN INVESTIGACIÓN EN SALUD 101
HOSPITAL GENERAL DE ZONA No.1, AGUASCALIENTES**

**DRA. JANNETT PADILLA LÓPEZ
COORDINADORA AUXILIAR MÉDICO DE INVESTIGACIÓN EN SALUD
P R E S E N T E**

Por medio de la presente le informo que el (la) Residente de la Especialidad de Cirugía General del Hospital General de Zona No. 3 del Instituto Mexicano del Seguro Social de la Delegación Aguascalientes.

DRA. ARIADNA BELTRAN RICO

Ha concluido satisfactoriamente con el trabajo de titulación denominado:

**“MODELO PREDICTOR DE COLEDOLITIASIS EN PACIENTES DEL
ORGANO DE OPERACIÓN ADMINISTRATIVA DESCONCENTRADA EN
AGUASCALIENTES”**

Número de Registro: R-2024-785-056 del Comité Local de Ética en Investigación No. 1018 y el comité de Investigación en Salud No. 101.

Elaborado de acuerdo con la opción de titulación: **TESIS**.

La Dra. Ariadna Beltran Rico, asistió a las asesorías correspondientes y realizó las actividades apegadas al plan de trabajo, por lo que no tengo inconvenientes para que se proceda a la impresión definitiva ante el comité que usted preside, para que sean realizados los trámites correspondientes a su especialidad. Sin otro particular, agradezco la atención que sirva a la presente, quedando a sus órdenes para cualquiera aclaración.

ATENTAMENTE:

**Dra. Elizabeth Guadalupe
Hernández Infante**

**Dr. José Luis Bizueto
Monroy**

Dra. Ivonne González Espinosa

**COORDINADOR CLÍNICO DE
EDUCACIÓN E INVESTIGACIÓN
EN SALUD**

PROFESOR TITULAR

ASESOR O DIRECTOR DE TESIS



CARTA DE CONCLUSIÓN DE TRABAJO DE TESIS

AGUASCALIENTES, AGS, A DE 20 DE SEPTIEMBRE DE 2025

DR. CARLOS ARMANDO SÁNCHEZ NAVARRO
COMITÉ LOCAL DE INVESTIGACIÓN EN SALUD 101
DELEGACIÓN AGUASCALIENTES

P R E S E N T E

Por medio de la presente le informo que el (la) Residente de la Especialidad de cirugía general del Hospital General de Zona No. 3 del Instituto Mexicano del Seguro Social de la Delegación Aguascalientes.

DRA ARIADNA BELTRAN RICO

Ha concluido satisfactoriamente con el trabajo de titulación denominado:

**“MODELO PREDICTOR DE COLEDOCOLITIASIS EN PACIENTES DEL
ORGANO DE OPERACIÓN ADMINISTRATIVA DESCONCENTRADA EN
AGUASCALIENTES”**

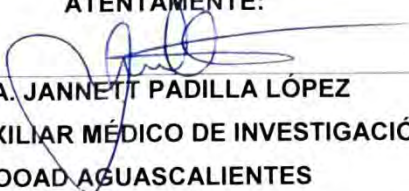
con Número de Registro R-2024-785-056 del Comité Local de Ética en Investigación No. 1018 y el comité de Investigación en Salud No. 101.

Elaborado de acuerdo con la opción de titulación: **TESIS**.

La Dra. Ariadna Beltran Rico asistió a las asesorías correspondientes con su director de tesis y realizó las actividades para la realización del protocolo de investigación, con apego al plan de trabajo, dando cumplimiento a la normatividad de investigación vigente en el Instituto Mexicano del Seguro Social.

Sin otro particular, agradezco a usted su atención, enviándole un cordial saludo.

ATENTAMENTE:


DRA. JANNETT PADILLA LÓPEZ

COORDINADORA AUXILIAR MÉDICO DE INVESTIGACIÓN EN SALUD
COAD AGUASCALIENTES



CARTA DE CONCLUSIÓN DE TRABAJO DE TESIS

AGUASCALIENTES, AGS, A DE 20 DE SEPTIEMBRE DE 2025

DR. SERGIO RAMIREZ GONZALEZ

DECANO DEL CENTRO DE CIENCIAS DE LA SALUD

P R E S E N T E

Por medio de la presente le informo que el (la) Residente de la Especialidad de cirugía general del Hospital General de Zona No. 3 del Instituto Mexicano del Seguro Social de la Delegación Aguascalientes.

DRA ARIADNA BELTRAN RICO

Ha concluido satisfactoriamente con el trabajo de titulación denominado:

**“MODELO PREDICTOR DE COLEDOLITIASIS EN PACIENTES DEL
ORGANO DE OPERACIÓN ADMINISTRATIVA DESCONCENTRADA EN
AGUASCALIENTES”**

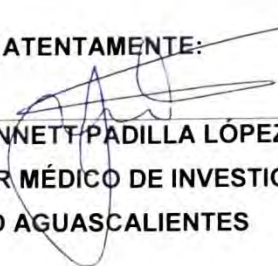
con Número de Registro R-2024-785-056 del Comité Local de Ética en Investigación No. 1018 y el comité de Investigación en Salud No. 101.

Elaborado de acuerdo con la opción de titulación: **TESIS**.

La Dra. Ariadna Beltran Rico asistió a las asesorías correspondientes con su director de tesis y realizó las actividades para la realización del protocolo de investigación, con apego al plan de trabajo, dando cumplimiento a la normatividad de investigación vigente en el Instituto Mexicano del Seguro Social.

Sin otro particular, agradezco a usted su atención, enviándole un cordial saludo.

ATENTAMENTE:


DRA. JANNETT PADILLA LÓPEZ

**COORDINADORA AUXILIAR MÉDICO DE INVESTIGACIÓN EN SALUD
OAO AGUASCALIENTES**



DICTAMEN DE LIBERACIÓN ACADÉMICA PARA INICIAR LOS TRÁMITES DEL
EXAMEN DE GRADO - ESPECIALIDADES MÉDICAS



Fecha de dictaminación dd/mm/aa: 04/02/2026

NOMBRE: BELTRAN RICO ARIADNA ID: 345398
ESPECIALIDAD: CIRUGIA GENERAL LGAC (del posgrado): PACIENTE QUIRURGICO EN EL PERIOPERATORIO
TIPO DE TRABAJO: (X) Tesis () Trabajo práctico
SEDE HOSPITALARIA: INSTITUTO MEXICANO DEL SEGURO SOCIAL
TITULO: MODELO PREDICTOR DE COLEDOCOLITIASIS EN PACIENTES DEL ORGANO DE OPERACION ADMINISTRATIVA DESCONCENTRADA EN AGUASCALIENTES

IMPACTO SOCIAL (señalar el impacto logrado): HERRMIENTA PRECISA, SIN COSTO QUE BRINDA APOYO AL DIAGNOSTICO DE COLEDOCOLITIASIS

INDICAR SI - NO - NA (No aplica) SEGÚN CORRESPONDA:

Elementos para la revisión académica del trabajo de tesis o trabajo práctico:

SI El trabajo es congruente con las LGAC de la especialidad médica
SI La problemática fue abordada desde un enfoque multidisciplinario
SI Existe coherencia, continuidad y orden lógico del tema central con cada apartado
SI Los resultados del trabajo dan respuesta a las preguntas de investigación o a la problemática que aborda
SI Los resultados presentados en el trabajo son de gran relevancia científica, tecnológica o profesional según el área
SI El trabajo demuestra más de una aportación original al conocimiento de su área
SI Las aportaciones responden a los problemas prioritarios del país
NO Generó transferencia del conocimiento o tecnológica
SI Cumple con la ética para la investigación (reporte de la herramienta antiplagio)

El egresado cumple con lo siguiente:

SI Cumple con lo señalado por el Reglamento General de Posgrado
SI Cumple con los requisitos señalados en el plan de estudios
SI Cuenta con los votos aprobatorios del comité tutorial
SI Cuenta con la aprobación del (la) Jefe de Enseñanza y/o Hospital
SI Coincide con el título y objetivo registrado
SI Tiene el CVU de la SECIHTI actualizado
NA Tiene el artículo aceptado o publicado y cumple con los requisitos institucionales

Con base a estos criterios, se autoriza se continúen con los trámites de titulación y programación del examen de grado

Sí X

No

FIRMAS

Revisó:

NOMBRE Y FIRMA DEL SECRETARIO DE INVESTIGACIÓN Y POSGRADO:

MCB E SILVIA PATRICIA GONZÁLEZ FLORES

Autorizó:

NOMBRE Y FIRMA DEL DECANO:

DR. EN FARM. SERGIO RAMÍREZ GONZÁLEZ

Nota: procede el trámite para el Depto. de Apoyo al Posgrado

En cumplimiento con el Art. 136 fracción II, inciso g) del Reglamento General de Posgrado, que a la letra señala: autorización de la persona titular del Decanato del Centro de Ciencias de la Salud.



Outlook

CIRU - Artículo registrado correctamente

Desde Online submission manuscript <no-reply@permanyemail.com>

Fecha Vie 26/09/2025 08:44 PM

Para ariadnabeltran95@hotmail.com <ariadnabeltran95@hotmail.com>



**CIRUGÍA
Y CIRUJANOS**

ORGANO DE DIFUSIÓN CIENTÍFICA
DE LA ACADEMIA MEXICANA DE CIRUGÍA
FUNDADA EN 1922



PERMANYER MÉXICO
www.permanyer.com

Estimado/a Dr/Dra Ariadna ,

Gracias por su interés en nuestra publicación. Le confirmamos que el artículo '[MODELO PREDICTOR DE COLEDOCOLITIASIS EN PACIENTES DEL ÓRGANO DE OPERACIÓN ADMINISTRATIVA DESCONCENTRADA EN AGUASCALIENTES](#)' (CIRU/0462/25) se ha registrado correctamente en nuestro sistema.

El Comité Editorial de Revista Cirugía y Cirujanos con todo gusto revisará su manuscrito y en breve nos comunicaremos con Usted.

Muchas gracias y saludos cordiales,

El Editor
Revista Cirugía y Cirujanos



Temístocles 315, Dept. 404. - Col Polanco,
Del. Miguel Hidalgo México D.F., 11560 |
ricardo.morales@permanyer.com

AGRADECIMIENTOS

Agradezco a los pacientes por su confianza, paciencia y su fuerza en la recuperación en sus padecimientos. Por permitirme acompañarlos.

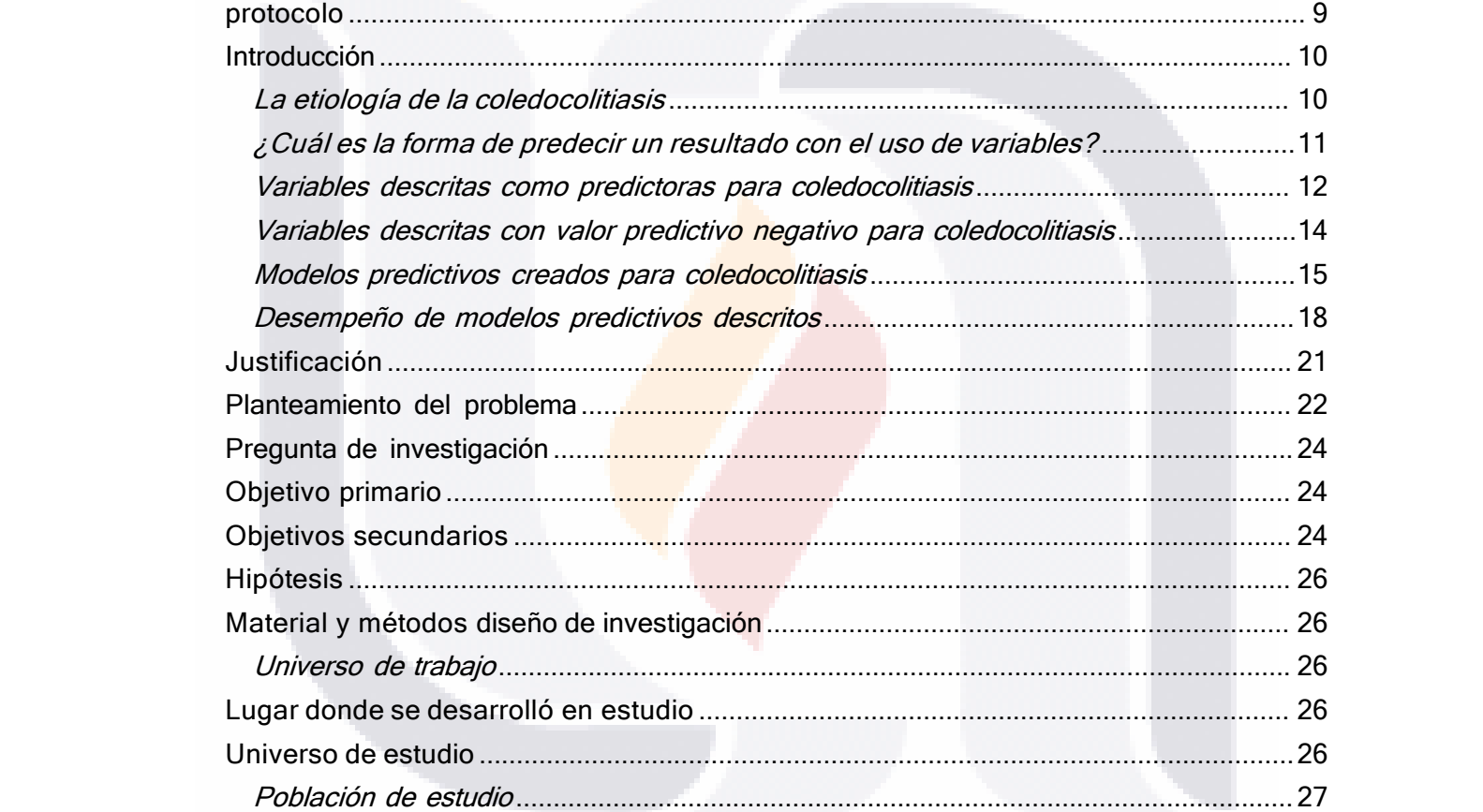
Agradezco a la Dra. Ivonne González por la oportunidad de aprender, su gran apoyo, consejos y las facilidades que brindó para desarrollar mi trabajo, es el ejemplo de cirujana que las mujeres aspiramos a ser. Gracias al Dr. José Luis Bizueto por su dedicación, fuerte compromiso con nuestra enseñanza, darnos el impulso y la inspiración para enfrentar los problemas.

Al Dr. Rogelio Gutiérrez, por el cual tengo una gran admiración. Le agradezco por su valioso tiempo, su gran paciencia y por enseñarme las matemáticas de una manera que no conocía.

A Javier González por su apoyo incondicional, su gran amor, contención y sus desvelos a mi lado. Sin ti no hubiera sido posible transitar este camino.

Gracias a Leticia Rico y Octavio Beltrán por dejarme volar, enseñarme a vivir y a soñar. Porque perdura en mi la huella del camino enseñado.

Agradezco a los administrativos, compañeros de trabajo, jefes de cirugía de los hospitales que me brindaron apoyo y recursos para desarrollar mi investigación.



Índice

Resumen 5

Abstract 7

Unidades médicas de investigación 8

Datos de contacto del director de las unidades médicas donde se desarrollará el protocolo 9

Introducción 10

La etiología de la coledocolitiasis 10

¿Cuál es la forma de predecir un resultado con el uso de variables? 11

Variables descritas como predictoras para coledocolitiasis 12

Variables descritas con valor predictivo negativo para coledocolitiasis 14

Modelos predictivos creados para coledocolitiasis 15

Desempeño de modelos predictivos descritos 18

Justificación 21

Planteamiento del problema 22

Pregunta de investigación 24

Objetivo primario 24

Objetivos secundarios 24

Hipótesis 26

Material y métodos diseño de investigación 26

Universo de trabajo 26

Lugar donde se desarrolló en estudio 26

Universo de estudio 26

Población de estudio 27

Unidad de observación 27

Unidad de análisis 27

Criterios de inclusión 27

Criterios de no inclusión 28

Criterios de exclusión 28

Tamaño de muestra 28

Operacionalización de las variables 29

Procedimientos para recolección de datos 32

Instrumentos a utilizar 34

<i>Métodos para el control y calidad de los datos</i>	35
<i>Análisis estadístico de los datos</i>	35
<i>Aspectos éticos</i>	37
<i>Aspectos de bioseguridad</i>	38
<i>Recursos, financiamiento y factibilidad</i>	38
Resultados	39
<i>Procesamiento de los resultados</i>	41
<i>Construcción del modelo: regresión logística múltiple</i>	44
<i>Modelo de regresión logística con la variable "IMC"</i>	46
<i>Modelo de regresión logística con la variable "Riesgo ASGE"</i>	47
<i>Modelo de regresión logística con la variable "Fosfatasa alcalina (FA)"</i>	49
<i>Modelo de regresión logística con la variable "gama glutamiltransferasa (GGT)"</i>	51
<i>Modelo de regresión logística con la variable "Lito"</i>	52
<i>Selección del modelo predictivo</i>	55
<i>Desempeño del modelo predictivo</i>	55
<i>Otros modelos alternativos</i>	58
<i>La importancia de los umbrales en las pruebas diagnósticas</i>	61
.....	64
<i>Desempeño de los criterios ASGE 2019</i>	64
Discusión	65
Conclusiones	68
Glosario	69
Referencias bibliográficas	71
Anexos	74
<i>Anexo A. Formato de solicitud de expedientes</i>	74
<i>Anexo B. Base de datos para captura de variables</i>	75
<i>Anexo C. Coeficiente de correlación de las variables continuas</i>	76
<i>Anexo D. Scripts del software R utilizados</i>	77

Índice de tablas

Tabla 1. Variables , significado y las unidades de medición.

32

Tabla 2. Variables incluidas en el estudio.....

43

Tabla 3. Odds ratio, los intervalos de confianza, valor P de cada variable y los coeficientes B.....

54

Tabla 4. Predicciones realizadas con el modelo.....

56

Tabla 5. Métricas de predicción de la prueba diagnóstica.....

57

Tabla 6. Predicciones realizadas con el modelo 18 variables.....

59

Tabla 7. Métricas de predicción de la prueba diagnóstica con el modelo 18 variables.....

59

Tabla 8. Predicciones realizadas con el modelo terciario de 5 variables.....

60

Tabla 9. Métricas de predicción de la prueba diagnóstica con el modelo 5 variables.....

61

Tabla 10. Matriz de confusión criterios ASGE.....

64

Tabla 11. Muestra el desempeño de los criterios ASGE 2019.....

65

Índice de gráficos

Ilustración 1. Calculo de muestra. 29

Ilustración 2. Distribución del tamaño de muestra entre los 3 hospitales. 29

Ilustración 3. Variables que mostraron poder predictivo y fueron usadas para la elaboración del modelo 44

Ilustración 4. Curva ROC del modelo 57

Ilustración 5. Curva ROC del modelo del modelo 18 variables 59

Ilustración 6. Curva ROC del modelo de 5 variables 61

Ilustración 7. Grafica de valor predictivo positivo y valor predictivo negativo 62

Ilustración 8. Grafica de sensibilidad, lespecificidad y acurracy 63

Ilustración 9. Curva ROC de cada uno de los modelos creados 64



Resumen

Título: Modelo predictor de coledocolitiasis en pacientes del órgano de operación administrativa desconcentrada en Aguascalientes.

Introducción: la coledocolitiasis es la presencia de cálculos en la vía biliar. Existen características de los pacientes como predictores. La Sociedad Americana de Endoscopia Gastrointestinal (por sus siglas en ingles ASGE) y la Sociedad Europea de Endoscopia Gastrointestinal (por sus siglas en ingles ESGE), indican que el alto riesgo para la coledocolitiasis deben proceder a colangiopancreatografía retrograda endoscópica (CPRE). Estudios han mostrado una especificidad del 79,5%. Modelos de análisis multivariado ha demostrado mayor sensibilidad, especificidad y valor predictivo positivo. **Objetivo:** Generar un modelo predictivo diagnóstico para coledocolitiasis en adultos de las unidades del Órgano de Operación Administrativa Desconcentrada en Aguascalientes y compararlo con los criterios internacionales actuales. **Metodología:** Se realizó un estudio analítico, de casos y controles, con pacientes de riesgo intermedio y alto de coledocolitiasis. Se identificó edad, sexo, IMC, diabetes mellitus, hipertensión arterial, hábito tabáquico y alcohol, diagnóstico de envió a CPRE o exploración de vías biliares, grado de coledocolitiasis ASGE, niveles de fosfatasa alcalina, dilatación del colédoco, visualización de lito, niveles de bilirrubina total, de AST, ALT, GGT, colangitis y/o pancreatitis biliar. Se utilizó el software R para construcción del modelo predictivo con una regresión logística múltiple, se midió su odds ratio ajustado (OR adj) y de los criterios ASGE 2019, se obtuvo sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo y precisión del área bajo la curva. Se compararon las métricas contra los de los criterios ASGE 2019.

Resultados: el 56.18% se corroborar la presencia de coledocolitiasis. El sexo femenino representó el 68.80%. La GGT mostró una fuerte correlación con la presencia de litiasis intraductal ($p < 0.011$). El área bajo la curva es mayor en el modelo predictivo creado comparado con la obtenida con los criterios ASGE 2019

(0.63 vs 0.59).

Conclusión: el modelo predictivo proporciona una herramienta precisa, sin costo, aporta un valor predictivo positivo, una razón de verosimilitud positiva y un área bajo la curva considerablemente. Esto es beneficioso en el área médica ya que aumentará la probabilidad de certeza en los diagnósticos.

PALABRAS CLAVE

Modelo predictivo, Coledocolitiasis, conducto biliar común.

Abstract

Title: Predictive model for choledocholithiasis in patients of the deconcentrated administrative operation body (órgano de operación administrativa desconcentrada) in aguascalientes.

Introduction: choledocholithiasis is the presence of stones in the bile duct. Patient characteristics exist as predictors. The American Society for Gastrointestinal Endoscopy (ASGE) and the European Society of Gastrointestinal Endoscopy (ESGE) indicate that patients at high risk for choledocholithiasis should proceed to Endoscopic Retrograde Cholangiopancreatography (ERCP). Studies have shown a specificity of 79.5%. Multivariate analysis models have demonstrated higher sensitivity, specificity, and positive predictive value. **Objective:** to generate a diagnostic predictive model for choledocholithiasis in adults from the units of the Deconcentrated Administrative Operation Body in Aguascalientes and compare it with the current international criteria. **Methodology:** an analytical, case-control study was conducted with patients at intermediate and high risk of choledocholithiasis. Age, sex, BMI, diabetes mellitus, arterial hypertension, smoking habit, and alcohol consumption were identified, along with the diagnosis for referral to ERCP or bile duct exploration, the ASGE choledocholithiasis grade, levels of alkaline phosphatase, common bile duct dilation, stone visualization, levels of total bilirubin, AST, ALT, GGT, cholangitis and/or biliary pancreatitis. The R software was used for the construction of the predictive model with a multiple logistic regression, and its adjusted odds ratio (Odds ratio adj) was measured. Sensitivity, specificity, positive predictive value, negative predictive value, and Area Under the Curve (AUC) precision were obtained for both the model and the ASGE 2019 criteria. The metrics were compared against those of the ASGE 2019 criteria. **Results:** the presence of choledocholithiasis was corroborated in 56.18% of cases. The female sex represented 68.80%. GGT showed a strong correlation with the presence of intraductal stones ($p < 0.011$). The Area Under the Curve is greater in the created predictive model compared to that obtained with the ASGE 2019

criteria (0.63 vs 0.59). **Conclusion:** the predictive model provides an accurate, cost-free tool, offering a positive predictive value, a positive likelihood ratio, and a considerable Area Under the Curve. This is beneficial in the medical area as it will increase the probability of certainty in diagnoses.

KEYWORDS: predictive model, choledocholithiasis, common bile duct.

Unidades médicas de investigación

Hospital General de Zona No.1, Aguascalientes.

Dirección: boulevard José María Chávez No. 1202, Fraccionamiento Lindavista
C.P 20270, Aguascalientes, Municipio de Aguascalientes, Estado de Aguascalientes.

Hospital General de Zona No.2, Aguascalientes

Dirección: avenida Los Conos 102 Aguascalientes, Col. Ojocaliente, Aguascalientes,
C.P 20190 Aguascalientes, Municipio de Aguascalientes, Estado de Aguascalientes.

Hospital General de Zona No.3, Jesús María, Aguascalientes.

Dirección: prolongación Gral. Ignacio Zaragoza No. 905, Ejido Jesús María C.P 20908,
Jesús María, Municipio de Jesús María, Estado de Aguascalientes

Datos de contacto del director de las unidades médicas donde se desarrollará el protocolo

Nombre: Dra. Cynthia Lorena Pantoja Minguela

OOAD: Hospital General de Zona No.1, Aguascalientes.

Dirección de la unidad: boulevard José María Chávez No. 1202, Fraccionamiento Lindavista C.P 20270, Aguascalientes, Municipio de Aguascalientes, Estado de Aguascalientes.

Teléfono: 449 913 9050

Correo electrónico: cynthia.pantoja@imss.gob.mx

Nombre: Dr. Yamid Brajin Sánchez Rodríguez

OOAD: Hospital General de Zona No.2, Aguascalientes

Dirección de la unidad: avenida Los Conos 102 Aguascalientes, Col. Ojocaliente, Aguascalientes, C.P 20190 Aguascalientes, Municipio de Aguascalientes, Estado de Aguascalientes.

Teléfono: 449 970 3660

Correo electrónico: yamid.sanchez@imss.gob.mx

Nombre: Dra. Ana Cecilia Valdivia Martínez

OOAD: Hospital General de Zona No.3, Jesús María, Aguascalientes.

Dirección de la unidad: prolongación Gral. Ignacio Zaragoza No. 905, Ejido Jesús María C.P 20908, Jesús María, Municipio de Jesús María, Estado de Aguascalientes.

Teléfono: 449 153 5900

Correo electrónico: ana.valdivia@imss.gob.mx

Introducción

La coledocolitiasis se define como la presencia de cálculos en la vía biliar principal. (1) Podemos clasificarla por su localización en aquella que el lito se encuentra dentro la vía biliar dentro del hígado (intrahepática) y donde el lito se encuentra en la parte de la vía biliar fuera del hígado (extrahepática). También es clasificada por su origen en primaria cuando la litogénesis se origina dentro de la vía biliar debido a estasis de la bilis y secundaria, cuando la litogénesis se formó dentro de la vesícula biliar y estos migraron a la vía biliar común por su paso a través del cístico. Esta última es la más común en el 85% de los casos. (2,3)

La etiología de la coledocolitiasis

La causa de la coledocolitiasis es multifactorial. Existen características de los pacientes que se han analizado como predictores de coledocolitiasis, se describen de manera historia las “4 F” obesidad (Fat), sexo femenino (Female), edad fértil (Fertile) y cuarenta (Forty). (4) Los factores que intervienen para el proceso de litogénesis es la dieta, las hormonas y la predisposición genética. Diariamente el cuerpo secreta alrededor de 600 ml de bilis en 24 h. Los elementos que componen la bilis son: ácidos biliares, lecitina, colesterol, bilirrubina conjugada, proteínas y electrolitos. (5)

Los ácidos biliares (el ácido cólico y quenodexosicólico) se sintetizan en el hígado usando como materia prima el colesterol, se conjugan con glucina, taurina. Tiene la función de ser solubilizante y su función es ayudar a la absorción de las grasas en el intestino. Estas sales se van a absorber en el íleon terminal para ser absorbidos por los hepatocitos desde la circulación portal. (5)

El colesterol se va a diluir en la bilis solo si hay una armoniosa cantidad de lípidos, lecitina y ácidos biliares, de lo contrario sufre precipitación en forma de cristales. La colecistoquinina es la hormona que se libera en la mucosa duodenal ante el estímulo de grasas y proteínas, estimula la contracción de la vesícula y disminuir el tono del esfínter de Oddi para permitir el flujo de bilis hacia el duodeno. (5)

El proceso de litogénesis es una consecuencia de un aumento en la proporción de colesterol en la bilis respecto a las sales, esto aunado a una contractibilidad disminuida del epitelio biliar que condiciona la estasis de bilis. (6) Primero acontece una nucleación del colesterol con agregación paulatina de cristales precipitados que poco a poco van realizando el crecimiento del cálculo. El cálculo ocasiona una

disfunción del epitelio vesicular favoreciendo la estasis de bilis y perpetuando la formación. (5)

Existen 2 tipos de litos: de colesterol mixto, son los más prevalentes en un 80%, y los cálculos pigmentados en un 20%. (5) Se conoce que los cálculos de tamaño grande con aquellos mayores de 10 mm. Aquellos litos que presentan una mayor dureza y formas irregulares son características que los convierten en litos difíciles de eliminar. Asimismo, existe variantes anatómicas que pueden ocasionar la presencia de un conducto colédoco anormal presentando oblicuidad, estrecheces perivaterianas que pueden causar que los litos se impacten e incluso sean refractarios a la extracción por procedimientos terapéuticos. (6)

¿Cuál es la forma de predecir un resultado con el uso de variables?

En el área médica existen varios estados patológicos que muchas veces se ven influenciados por varias características previas presentes que en conjunto pueden desarrollar la historia de la enfermedad.

Por ese motivo es de interés conocer la acción que tienen uno o varios antecedentes para un hecho (enfermedad) se produzca. La enfermedad, extrapolándola se llamará variante dependiente (Y) y es dicotómica, eso quiere decir que la formula planteada nos arrojará la presencia o ausencia de la enfermedad.

La fórmula que representa la regresión logística es la siguiente:

$$\pi(x) = \frac{\exp(B_0 + B_1x)}{1 + \exp(B_0 + B_1x)}$$

Donde $\pi(x)$ es la probabilidad de obtener 1 “éxito”, en este caso representaría que la patología a indagar este presente. Y depende del valor X (la variable predictora).

Si se realiza el despeje de $B_0 + B_1x$ encontramos que:

$$\ln\left(\frac{\pi}{1 - \pi}\right) = B_0 + B_1x$$

Donde el coeficiente $\ln\left(\frac{\pi}{1-\pi}\right)$ representa el Odds ratio (OR) o razón de momios, lo que significa que es la probabilidad de que se presente un éxito entre que no se presente. Esta razón siempre debe ser mayor a 1 para significar una verdadera diferencia asociativa en que se presente el resultado.

Cada una de las variables independientes creará una curva que mostrará el grado de probabilidad que a cierta magnitud logran un resultado "exitoso". El grado en el que la curva en una gráfica se presenta se interpreta como coeficiente β (7)

El coeficiente β o pendiente de la recta representa el grado de inclinación con respecto a la línea que representa la media en la gráfica del modelo de regresión. Cuanto mayor sea el coeficiente beta, más contribución tendrá esta variable en el resultado. (8)

La regresión logística al incluir múltiples variables independientes es el mejor modelo para este tipo de análisis ya que es más informativo porque al mostrar como contribuye cada variable a la probabilidad de la variable dependiente realizando un ajustando a las demás. Así pues, estado patológicos en los que intervienen varios componentes o variables, es importante en análisis de todas al unísono. (8)

Variables descritas como predictoras para coledocolitiasis

Menezes, et al (2000). Evaluó una muestra prospectivamente de 211 pacientes de Gran Bretaña. Estimando una incidencia de coledocolitiasis del 15 %. (9) Sheshang et al (2016) observaron prospectivamente una muestra de 407 pacientes con coledocolitiasis sintomática de un hospital de Mumbai, India. Determinó una prevalencia de 18.9% coledocolitiasis en pacientes con enfermedad litiasica sintomática. (10)

Khoury et al (2020) en un estudio retrospectivo llevado a cabo en Nahariya, Israel del 1ero de enero al 31 diciembre 2018 donde participaron 112 pacientes con coledocolitiasis confirmada por ultrasonido endoscópico, se identificaron los siguientes predictores: edad promedio de 68 años, género femenino de 50.9% y 49.1 masculinos, hipertensión arterial en el 60.4%, diabetes mellitus en 47.2%, tabaquismo en 11.3% enfermedades cardiacas isquémicas 20.7%, hiperlipidemia 35.8% y falla cardiaca en 15.1%. La media encontrada en los parámetros laboratoriales fue: alanino aminotransferasa 136.3 U/L, aspartato aminotransferasa 86.1 U/L, fosfatasa alcalina 240.5 U/L, GGT 417 U/L, bilirrubina total 2.5, hemoglobina 12.3 g/dl y leucocitos de 8.2 células x 10³. En cuanto a la dilatación del colédoco la media fue de 7.7 mm. (11)

En estadística local tenemos la aportación de Ovalle-Chao et al (2022) en un hospital de 2do nivel en Nuevo León, México con 352 pacientes con diferentes riesgos de coledocolitiasis: 5 pacientes riesgo bajo, 125 intermedio y 222 riesgo alto. Se realizó la comprobación de coledocolitiasis mediante CPRE o explicación laparoscopia de vías biliares. Dentro de esta muestra encontraron que 88.4% eran mujeres, 11.6% hombres y la media de edad fue 35.56 años. Se confirmó la coledocolitiasis en 256 pacientes (el 72.73% de la muestra) de estos la media de edad fue de 36.2 años, los que tuvieron menos de 55 años representaron el 72.88% (223 pacientes), el sexo femenino 72.35% (225 mujeres). En las pruebas de laboratorio los niveles de bilirrubina mayores a 4 mg/dl se registraron en 76.17%, las enzimas hepáticas anormales en el 73.35% (234 pacientes) y en estudios de imagen se incluyeron los colédocos con tamaño mayor a 6 mm (75.25%, 225 pacientes) y la visualización del lito dentro de las vías biliares en 69.31% (192 pacientes). En cuanto a los predictores de coledocolitiasis reportados por la ASGE que con mayor frecuencia se reportaron fueron: la alteración en el perfil hepático (90.63%), dilatación del colédoco más de 6 mm por US (84.94%) y la combinación de la dilatación del colédoco con niveles de bilirrubina total mayor a 4 mg/dl (54.55%). (12)

En el estudio de Tunruttanakul et al (2022), se empleó una muestra de 621 pacientes con sospecha de coledocolitiasis en Tailandia. Para la confirmación diagnóstica se empleó la CPRE, exploración de vías biliares o colangiografía transoperatoria y aunado a estala colangioresonancia. Se encontró que la media de edad de los pacientes con coledocolitiasis en la muestra fue 65.3 (73.7%) y las mujeres representaron el 59.4% (221 pacientes). Como manifestaciones clínicas se reportó dolor abdominal en un 23.4%, pancreatitis biliar en 6.2% e ictericia en 15.9%. El 5.7% de los pacientes padecían colecistitis, la colangitis fue desarrollada en el 48.9%. En lo que respecta al perfil bioquímico la fosfatasa alcalina en el rango de 125 – 250 U/L se encontró en un 25.5% y 39.8% niveles mayores a 250 U/L. Los niveles de bilirrubina total entre 1.8-4 mg/dl en un 17.7% y valores mayores a 4 mg/dl en el 22.0%. En los estudios de imagen se encontró lito intraluminal en el 61.0% con un diámetro mayor a 6 mm en el 90.3% En contraste también se observaron alteraciones nivel bioquímico e imagenológico en el grupo de pacientes que no padecían coledocolitiasis (249 pacientes) su distribución se presentó de la siguiente manera: niveles de fosfatasa alcalina mayores a 250 mg/dl en el 12.1%, bilirrubinas totales de

1.8 a 4 mg/dl en 14.1% y valores mayores a 4 mg/dl igual a 8 %. Se identificó por los mismos estudios de imagen mencionados presencia de litos hasta en un 28.1% y dilatación del colédoco en 64.3% (13)

En una muestra de 270 pacientes con sospecha de coledocolitiasis en un hospital de los Ángeles realizada por Dalai et al (2021) corroboró el diagnóstico en un 84.6%. (230 pacientes) mediante CPRE. La frecuencia de las características observadas fue: 60.4% tenían más de 40 años (139 pacientes), el 64.8% eran mujeres (149 pacientes), el 75.2% eran de origen latino (173 pacientes) encontrando promedio de índice de masa corporal (IMC) de 29.0 kg/m², además el 40.6% presentaban el IMC mayor a 30 kg/m². La fiebre se presentó únicamente en el 17.4% (40 pacientes). De las 4 F descritas en la literatura (female, fat, forty and fértil) solo se la presentaron el 16.21% (36 pacientes). De manera arbitraria a las guías ASGE 2019 donde brinda importancia a los valores de bilirrubina mayor a 4 mg/dl en conjunto con la dilatación del colédoco como un predictor fuerte de coledocolitiasis, esta cohorte reveló que los pacientes a los que se les descartó coledocolitiasis por CPRE (40 pacientes) también presentaron elevación de la bilirrubina con un pico máximo promedio de 2.1 mg/dl y dilatación del colédoco promedio de 8 mm. (4)

En prevalencias a nivel local. Marín et al (2020) identificó en población del estado de Aguascalientes, los factores más prevalentes en los casos reportados de coledocolitiasis, estos fueron: el sexo femenino (72.73%), media de edad de 43 años, una media de niveles de bilirrubina total de 6.1 mg/dl, con bilirrubina directa media de 4.7 mg/dl, media del diámetro del colédoco de 10.6 mm, se lograron identificar litos en el colédoco con el ultrasonido en el 30.26% con un tamaño medio de 13.4 mm. (3)

Variables descritas con valor predictivo negativo para coledocolitiasis

Como predictores negativos, En los Ángeles, Estados Unidos. Dalai et al (2021) en pacientes que se descartó la coledocolitiasis, la diabetes mellitus fue prevalente en un 40.7% (contra 10.3% en los que se diagnosticó coledocolitiasis). (4) Por otro lado Tunruttanakul, et al (2022) se encontró que la pancreatitis fue un predictor negativo (odds ratio=0.52) y la colangitis no fue significativamente relevante (odds ratio= 1.39). Esto contradictorio a las guías ASGE que indican colangitis como un fuerte predictor de coledocolitiasis. (13)

Modelos predictivos creados para coledocolitiasis

N. Menezes et al (2000) estudiaron una muestra de 233 pacientes británicos entre octubre 1992 y diciembre 1994 en Dumfries & Galloway con sospecha de coledocolitiasis utilizó como muestra de control pacientes a los que se les practicó CPRE o colangiografía transoperatoria para confirmación diagnóstica. De estos pacientes 77 presentaron litos en la vía biliar y 156 fue ausente. Se documentaron las variables de edad, sexo, la presentación clínica, el resultado de niveles de bilirrubina, aspartato aminotransferasa, alanino aminotransferasa y fosfatasa alcalina, así como imagenológicos. Realizaron una asociación de las variables utilizando Chi cuadrada con test Yates para factores binarios y Mann-Whitney U para factores ordinarios y cuantitativos. Con estas variables se realizó un análisis multivariable de regresión logística. Se encontró para la edad mayor a 55 años ($p < 0.05$), el sexo femenino ($p < 0.05$), ausencia de colangitis ascendente ($p < 0.05$), ausencia de ictericia con ($p < 0.05$), niveles de transaminasas normales ($p < 0.01$), diámetro del conducto biliar sin dilatación ($p < 0.001$) y ausencia de litos en el conducto biliar ($p < 0.05$). Se estableció un sistema de puntuación con un punto de cohorte de 3 puntos o más. Pacientes con 3 puntos o menos se observó disminución en el riesgo de coledocolitiasis y el riesgo de practicar un CPRE innecesaria. En el estudio 136 pacientes obtuvieron una puntuación de 0 a 2, de estos el 4% (5 pacientes) tenían coledocolitiasis observada por colangiografía transoperatoria laparoscópica con litos menores a 5 mm. De lo contrario 55 pacientes una puntuación de 3 o más, el 42 % (23 pacientes) tenían cálculos en el conducto biliar común. (9)

Kamath et al (2016) realizó un estudio observacional prospectivo en Mumbai para determinar si las variables clínicas, de laboratorio e imagenológicas descritas en la literatura predecían el diagnóstico de coledocolitiasis con la elaboración de una regresión logística. Observó 407 pacientes con enfermedad litiasica sintomática, diagnosticando en ellos a 77 pacientes con coledocolitiasis. Se utilizó la CPRE como gold estándar para el diagnóstico. Se seleccionaron las variables que mostraron una odd ratio significativa, todas tuvieron un IC 95%. Las variables fueron las siguientes: transaminasa glutámica oxalacética (OR=2.138), transaminasa glutámica pirúvica (OR= 1.398), fosfatasa alcalina (OR = 10.956), amilasa (OR= 1.498) dilatación del conducto en ultrasonido (OR=1.946) y niveles de bilirrubina totales (OR= 12.321). De las anteriores mencionadas, la transaminasa glutámica pirúvica y la amilasa no mostraron significancia ($p < 0.05$), por lo que fueron descartadas. Para la realización

del modelo se utilizaron 4 factores: dilatación del conducto colédoco mayor a 6 mm, bilirrubina total mayor a 2 mg/dl, fosfatasa alcalina mayor a 190 mg/dl y transaminasa glutámica pirúvica mayor a 40 U/L. Cada uno de estos arrojó una puntuación de 1. (10)

Tunruttanakul, et al (2022) desarrollaron modelos predictivos multivariados con datos de una cohorte observacional retrospectiva de pacientes con sospecha de cálculos de conducto biliar común (CBD) en el Hospital Sawanpracharak, Tailandia. Se incluyeron pacientes referidos al hospital con colangiografía pancreática retrógrada endoscópica (82,9%, 515 pacientes), exploración de vías biliares (8,1%; 50 pacientes) y colangiopancreatografía resonancia magnética (CPRM 9,0%; 56 pacientes). Siendo estos estudios el gold estándar utilizado para el diagnóstico. Con dichos datos se calcularon los coeficientes Beta de los parámetros identificados como predictores para construir un modelo predictivo basado en puntuaciones. Se utilizó la suma de la puntuación total de cada paciente para evaluar la capacidad para predecir el estado de los cálculos en el conducto biliar común. Como primer resultado las variables denotaron un odds ratio significativo positivo o negativo las cuales fueron: edad mayor a 55 años (OR = 1.84, β = 0.61), pancreatitis (OR = 0.52, β = -0.65), colangitis (OR = 1.39, β = 0.33) cirrosis (OR = 0.36, β = -1.03), nivel de fosfatasa alcalina de 125-250 U/L (OR = 1.20, β = 0.18), nivel fosfatasa alcalina mayor a 250 U/L (OR = 3.35, β = 1.21), niveles de bilirrubina mayor a 4 mg/dl (OR = 2.75, β = 1.01) colédoco mayor a 6 mm (OR = 2.75, β = 1.01) y litos detectados (OR = 2.61, β = 0.96). Con los coeficientes B obtenidos, se realizó una sumatoria total. La puntuación media fue mayor en el grupo de cálculos en el CBD que en el grupo de los ausentes con una media $\pm$ desviación estándar de 15.6 ± 6.2 contra 8.3 ± 6.2 , una $p < 0.01$ y análisis de regresión logística multivariable fue el principal análisis de desarrollo del modelo método, brindando una puntuación relacionada con cada una de las categorías que clasifica ASGE: riesgo bajo una puntuación menor o igual a 5, riesgo intermedio 5.5- 14.5 y riesgo alto mayor o igual a 15. Los pacientes se catalogaron conforme a los criterios conocidos por las guías en bajo (13.5 % pacientes), intermedio (44.6% pacientes) y alto riesgo (41.9% pacientes) de coledocolitiasis. (13)

Dalai et al (2021) realizaron una cohorte retrospectiva con 270 pacientes con alta sospecha de coledocolitiasis de un hospital de los Ángeles, se revisaron las CPRES realizadas desde el uno de noviembre 2015 al 31 diciembre 2019, dividiendo a los

pacientes en los que se encontró hallazgos imagenológicos de coledocolitiasis (dilatación del colédoco mayor a 6 mm) únicamente por US abdominal y por varios estudios no invasivos además del US, como la tomografía (116 pacientes), colagioresonancia (77 pacientes) y ultrasonido endoscópico (15 pacientes). Se hizo uso de CPRE como gold estándar. Se midió el odds ratio de las variables identificando en su resultado primario, las de mayor significancia que fueron: nivel total de bilirrubina, edad, índice de masa corporal y grosor del colédoco. Con ayuda de aprendizaje automatizado realizó 7 modelos predictivos, observando mayor precisión diagnóstica (con una AUC 0.791) con el modelo de bosque aleatorizado. Debido a que los endoscopistas en este estudio demostraron identificar correctamente al 84.6% de los pacientes que padecieron coledocolitiasis, la utilidad de este modelo radica en encontrar a los que no tienen coledocolitiasis antes de ser sometidos a CPRE. (4)

Blum et al (2024) utilizó una cohorte retrospectiva donde se incluyeron 222 pacientes de la población de Tasmania desde el 2018 al 2023. Después de recolectar los parámetros demográficos, de laboratorio e imagenológicos del paciente, desarrollo 4 bases de datos, la primera se establecieron variables continuas y se estandarizaron usando el método Z, tal como variables binarias. La segunda, las variables continuas fueron remplazadas por variables binarias. Se desarrolló una Chi cuadrada para identificar los datos significativos estadísticamente, la tercera base las variables continuas que fueron significativas entre los pacientes con o sin coledocolitiasis se identificaron mediante pruebas de T- de dos vías para reunir las con las variables binarias previas y formar la cuarta base de datos, aquí se decidió la eliminación de los valores de bilirrubina debido a que todos los pacientes superaron el valor de corte. Desarrollaron 4 modelos predictivos usando aprendizaje automatizado: la regresión logística, el K-vecino más cercano (KNN), bosque aleatorio (RF) y XGboost. Además, se realizó un modelo de conjunto que combina las salidas de los modelos LR, KNN y RF. Para el cálculo de modelos RF utilizaron el coeficiente de Gini con el que identificaron que predictores eran significativos y para los modelos de regresión logística se obtuvieron los coeficientes B. Todo lo anterior con una $p < 0.05$. Se utilizaron los criterios de ASGE como control para el diagnóstico de coledocolitiasis. En los resultados de predicción se observó que el modelo RF tuvo una funcionalidad más congruente en los 4 bases de datos: una AUROC 0.83 en la primera y

segunda base, AUROC 0.77 para la tercera base y 0.76 para la cuarta base. En la regresión logística las variables con mayor significancia predictiva fue el diámetro del colédoco y la visualización de lito con un coeficiente logit de 4 según estima la gráfica presentada. La lipasa y la alanina aminotransferasa se relacionaron negativamente con un coeficiente de Logit de -5 presentado en el gráfico de barras apiladas. (14)

Desempeño de modelos predictivos descritos

N. Menezes et al (2000) con su sistema de puntuación diseñado para predecir coledocolitiasis y evitar CPRES innecesarias, usando las siguientes características: edad, sexo, si habían padecido ictericia, si habían presentado colangitis ascendente, niveles de transaminasa normales, elevados o con valores al doble, si hay o no dilatación del colédoco y si se observó lito dentro del colédoco por medio del ultrasonido. La CPRE fue el estudio de comparación para evaluar el desempeño diagnóstico. El punto de cohorte de 3 o más brindó una sensibilidad de 82% y una especificidad del 80% para el diagnóstico de coledocolitiasis. Cabe mencionar que su estudio lo desempeñaron en una población muestra diferente a la que fue desarrollado en un periodo desde abril 1996 a diciembre 1997 con un grupo de 211 pacientes que fueron sometidos a colecistectomía laparoscópica. (9)

Kamath et al (2016) con su estudio observacional prospectivo donde elaboro un análisis multivariable con los datos anteriormente mencionadas. Hizo utilización de los pacientes a los que se les practicó CPRE como grupo control. A la aplicación de su modelo encontró una precisión de predicción de 98.3%. Los datos recolectados mostraron un valor predictivo positivo (VPP) y valor predictivo negativo (VPN) de cada una de las variables que impactan positivamente en la regresión logística, distribuidos de la siguiente manera: transaminasa glutámica oxalacética ($p < 0.001$, VPP 55.8% y VPN 70%), fosfatasa alcalina ($p < 0.001$, VPP 82% y VPN 95.2%), dilatación del conducto en ultrasonido ($p < 0.000$, VPP 77.9% y VPN 97%) y niveles de bilirrubina totales ($p < 0.000$, VPP 87.5% y VPN 99.4%). Aplicaron su puntuación desarrollada en pacientes con cólico biliar asignando un valor por cada punto positivo que presentaron. Determinaron que 218 pacientes tuvieron una puntuación de 0 que cataloga a un paciente sin coledocolitiasis con un VPN del 100% sugiriendo por los autores en un algoritmo realizado por ellos la realización de la colecistectomía laparoscópica. 108 pacientes se posicionaron con una puntuación de 1-2 con un VPN

de 95.4% y 55.6% respectivamente. Ante este escenario sugieren CRM o ultrasonido endoscópico como medida diagnóstica de coledocolitiasis y posteriormente decidir el abordaje diagnóstico con colecistectomía laparoscópica o CPRE. En 3 pacientes se obtuvo una puntuación de 3-4 que documento a 26 pacientes de la muestra con coledocolitiasis lo que presenta un VPP de 96.29%. Los autores sugieren a estos pacientes indicación de CPRE, y en caso de no ser exitosa valorar realizar exploración de vías biliares. Con esto los autores indican que este sistema de puntuación puede evitar la realización de CPRE innecesaria, complicaciones asociadas a este procedimiento, disminuye costos de atención y tiempo en el abordaje. (10)

En el trabajo realizado por Dalai et al (2021) midió el rendimiento de los criterios de riesgo establecidos en las guías ASGE 2019 contra un modelo de aprendizaje automatizado de bosque aleatorizado. Con las guías ASGE midieron una precisión diagnóstica de 80.8%, sensibilidad del 90.9%, especificidad del 25.0%, un VPP 87% y VPN de 33.3%. Comparado contra el modelo de bosque aleatorizado que arrojó en su punto óptimo de la curva (0.825) la precisión diagnóstica fue de 76.9%, sensibilidad de 77.3% especificidad de 75.0%, VPP 94.4% y VPN 37.5%. Esto refleja que el modelo predictivo al tener una mayor especificidad (75.0% vs 25.0%) nos apoya en el diagnóstico de pacientes que probablemente no tendrán coledocolitiasis y por lo tanto sería innecesario someterlos a procedimiento intervencionistas. Se comparó el desempeño del este modelo contra las guías ASGE 2019, usando como gold estándar la CPRE. como resultado secundario se encontró que el modelo presenta una sensibilidad menor a la de las guías ASGE (77.3% vs 90.9%) pero una mayor especificidad (75.0% vs 25.0%). (4)

En la cohorte retrospectiva de Tunruttanakul et al (2022) utilizó su modelo predictivo con 621 pacientes tailandeses, mismo grupo de pacientes con el que lo creo, con el cual encontró aumento en la probabilidad de diagnóstico en mayor medida en el riesgo bajo (66.3%) y riesgo alto (83.8%) el riesgo intermedio continua en cifras similares (49.8%) haciendo uso de la CPRE, la colangiografía y la exploración de vías biliares además de colangiografía magnética como gold estándar de diagnóstico. Esto refleja que el modelo presenta una mayor capacidad diagnóstica (valor predictivo positivo de las guías ASGE 59.9% contra la probabilidad diagnóstica del modelo en el grupo de alto riesgo fue de 83.9%) es aproximadamente un 20% más alta que el valor

TESIS TESIS TESIS TESIS TESIS

previo a la prueba. Para valorar el desempeño del modelo se midió el área bajo la curva (AUC = 0.80, con IC 95%), lo que fue superior a lo estimado con las guías ASGE (AUC=0.67; IC95%) y guías ESGE (AUC =0.67; IC 95%). (13)

Blum et al (2024) utilizó una cohorte retrospectiva con 222 pacientes de la población de Tasmania para la realización de los 4 modelos desarrollados (regresión logística (RL), el K-vecino más cercano (KNN), bosque aleatorio (RF) y XGboost). Encontró que el modelo de bosque aleatorizado presentaba mayor consistencia durante sus pruebas. Se midió la sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, la puntuación F1 y la precisión diagnóstica de las Guías ASGE, así como de los modelos realizados, mostrando superioridad de los modelos sobre las guías ASGE en las siguientes magnitudes: sensibilidad (ASGE= 0.88, RF=0.83), especificidad (ASGE= 0.39, RF=0.69), valor predictivo positivo (ASGE= 0.6, RF=0.74), valor predictivo negativo (ASGE= 0.75, RF=0.79), F1 (ASGE= 0.71, RF=0.78) y precisión diagnóstica (ASGE= 0.64, RF=0.76). Para la evaluación del modelo, utilizaron los mismos datos previamente recopilados dividiéndolos en un grupo de entrenamiento de 70 y prueba de 30. Se encontró que el modelo de bosque aleatorizado arrojó un área bajo la curva más constante y superior comparado con el resto de los modelos (AUROC de 0.83), la sensibilidad fue de 0.94, la precisión de 0.79, especificidad del 0.79, el valor predictivo negativo 0.92, el valor predictivo positivo de 0.77 y la puntuación F1 de 0.81. El rendimiento lo compararon con la categoría de alto riesgo de coledocolitiasis según ASGE. (14)

En lo que respecta a nivel experimental, no se halló estudios en esta búsqueda.

Justificación

El tratamiento de la coledocolitiasis depende el tamaño del lito y la situación en la vía biliar para determinar su manejo, así como de los recursos con los que se cuente en el hospital. En Aguascalientes, el Instituto Mexicano del Seguro Social cuenta con 3 hospitales generales de zona. En el año 2023 se registraron 203 casos de coledocolitiasis y se indicaron 162 CPRES para su resolución. De los 3 hospitales solo 2 cuentan con endoscopistas y solo 1 hospital con el equipo y material necesario para la realización de CPRE, el resto realiza la solicitud de este procedimiento de manera subrogada, lo que implica un tiempo de espera promedio de 3 a 5 días hábiles, para aquellos que se solicitan de manera ordinaria, y 24 h para todos aquellos solicitados como urgencia. Esto se contrapone con los conocimientos actuales en cuanto a la resolución quirúrgica temprana mejora el índice de morbilidad y mortalidad de esta patología.

La recomendación actual es que todos los litos detectados en la vía biliar común deben ser extraídos en cuanto se realiza el diagnóstico, sin embargo, esto se vuelve complicado en nuestra localidad al momento de la elección del tratamiento óptimo para el tipo de paciente entre las opciones diagnósticas y terapéuticas que existen, tomando en cuenta que no todas se encuentran disponibles en el mismo hospital.

Entre las 3 clínicas de atención médica se cuenta distribuidos 1 resonador, 1 equipo de colangiopancreatografía retrógrada endoscópica, 3 equipos de ultrasonido, 2 fluoroscopios y 2 equipos de exploración de vías biliares para exploración quirúrgica abierta. Con estas opciones cada cirujano, con base en su experiencia y dominio de técnicas resuelve los cuadros clínicos con alta sospecha de coledocolitiasis que se le presentan.

Se ha observado una diferencia entre el número de CPRES que se subrogan en cada una de las clínicas del IMSS en Aguascalientes a pesar de manejar una población equitativa para cada uno, no tenemos identificada la causa de esto y será un área de oportunidad para futuras investigaciones. Pero en este momento creemos que los cirujanos, pacientes e instituto se beneficiarán de una herramienta que apoye al diagnóstico oportuno y tratamiento adecuado con los menores riesgos posibles.

En esta época (intensificado con el desarrollo de la inteligencia artificial en varios ámbitos cotidianos, no dejando fuera el mercado de la salud) se han establecido modelos de predicción que apoyen al resultado que podremos pronosticar ante cierto

conjunto de datos conocidos. Una forma de desarrollo de estos modelos en el área médica es mediante análisis estadísticos de regresión que su propósito es explicar un estado de salud o patológico vinculándolo con una variable (matemáticamente representado por Y, variable dependiente) teniendo en cuenta varios datos conjuntamente (representado por X, variables independientes)

Por lo que consideramos pertinente la creación de un modelo predictivo de coledocolitiasis que funja como herramienta adicional para apoyar el juicio médico y la toma de decisiones debido a que los cirujanos de las clínicas no cuentan con todas las opciones terapéuticas en un mismo lugar. Esto contribuiría a disminuir el tiempo en la toma de decisión y ejecución del procedimiento terapéutico a los pacientes que ingresan con sospecha de coledocolitiasis, reflejándose en la disminución de la estancia intrahospitalaria, disminución en costos de atención médica y evitar ejecutar procedimientos innecesarios y algunas ocasiones invasivas en los que los pacientes se pueden aumentar su morbilidad.

Planteamiento del problema

La prevalencia de coledocolitiasis descrita en la bibliografía se reporta en 18.9%. (13) En el 2023, se registraron 203 casos de coledocolitiasis entre las 3 clínicas del Órgano de Operación Administrativa Desconcentrada de Aguascalientes, de las cuales 162 requirieron CPRE. Al no contar con este recurso de manera ordinaria dentro de la institución, debe realizarse de manera subrogada. La CPRE simple con asistencia de anestesia tiene un costo para el instituto de \$1,500.00 pesos mexicanos, aumentando el costo si se agrega colocación de endoprótesis, litotripsia o toma biopsia. Si sumamos que el servicio subrogado se lleva de 3 a 5 días hábiles para su realización esto incrementa aún más los costos de atención y prolonga la estancia intrahospitalaria ya que el paciente debe esperar para recibir un diagnóstico certero y posteriormente culminar con colecistectomía.

Se conoce por la revisión de la literatura y las guías ASGE que la atención de la coledocolitiasis debe ser inmediata para evitar complicaciones y aumento de la morbilidad. Las últimas directrices ASGE 2019, establecen 3 categorías con sus respectivas probabilidades diagnósticas: bajo riesgo (10%), intermedio (10- 50%) y alto riesgo (50%). (6) Sin embargo, se ha encontrado un diagnóstico subóptimo con estos criterios. (4) El principal grupo de riesgo expuesto son las mujeres con una

TESIS TESIS TESIS TESIS TESIS

frecuencia desde el 50.9% (10) hasta el 72.35%. Se ha reportado en estudios como el de Ovalle et al., (2022) con población mexicana en el estado de Nuevo León parámetros en los pacientes que tiene un fuerte factor predictivo para sospechar de esta patología, (volviéndose útil en centros como en este dónde se desarrolló este estudio) donde no se cuenta con colangiografía magnética ni ultrasonido endoscópico, estas fueron cifras fuera de rangos normales de las enzimas hepáticas en un 90.63%, colédoco mayor a 6 mm con bilirrubinas totales mayores a 4 mg/dl (54.55%) y un colédoco medido de 84.94%. Reportó que la prevalencia de edad fue menor a 55 años en un 72.88% (11), mientras que en una población de Tailandia Tunruttanakul et al (2022) mostró prevalencia de los mayores de 55 años en un 73.7%. A su vez, en su análisis encontró que la pancreatitis y la cirrosis fueron variables menos frecuentes en los pacientes con cálculos en el conducto biliar común (CBD). Por otro lado, la fosfatasa alcalina en niveles mayores a 250 mg/dl obtuvo un odds ratio más alto (3,35; IC 95%:) y antagónicamente, la ALP tiene una mínima importancia en las guías actuales. (12)

Dalai et al (2021) encontró que la diabetes mellitus fue una variable que se presentó en 10.3% en pacientes con cuadro de coledocolitiasis, mucho menos frecuente contra 40.7% observado en los pacientes sin coledocolitiasis. (4)

Los modelos predictivos con regresión logística han resultado beneficiosos en el ámbito médico por ser más económicos que otros estudios de imagen y se pueden aplicar con mayor rapidez sobre todo en lugares donde no se encuentra con todos los recursos diagnósticos disponibles de manera inmediata, siendo una ayuda al médico para la toma de decisiones de manera más rápida y con mayor precisión diagnóstica.

Meneses et al (2000) realizó una puntuación con regresión logística, usando la edad, sexo, presencia de ictericia, colangitis ascendente, niveles de transaminasas, dilatación del conducto colédoco, lito en el colédoco por ultrasonido y determinando una sensibilidad 82% y especificidad de 80%. Kamath et al (2016) también realizó una puntuación y propuso un algoritmo de manejo con base a él, lo construyó con el uso de transaminasa glutámica oxalacética, fosfatasa alcalina, milímetros del conducto biliar común y en nivel total de bilirrubina estimando diferentes valores predictivos de positividad o negatividad según la puntuación, refiere que su modelo es útil para la identificación de aquellos que no padecen coledocolitiasis. (9)

Dalai et al (2021) usaron aprendizaje automatizado para realizar un bosque

aleatorizado el cual mostró alta precisión diagnóstica (76.9%) (4). Blum et al (2024) de igual manera realizó el mismo tipo de modelo en su población identificando una precisión diagnóstica parecida de 76.0%. (14)

Tunruttanakul et al (2022) diseño con un modelo de regresión logística multivariable una puntuación para valorar su desempeño contralas guías ASGE según las categorías de riesgo, encontrando tener una mayor capacidad diagnóstica de 83.8% contra las guías ASGE que fue de 59.9% en esta muestra. (12)

Por lo tanto, se ha observado una inconsistencia en las prevalencias de factores predictores de coledocolitiasis posiblemente derivado de la etnia y datos sociodemográficos de los pacientes. Debido a que de ellos nace el análisis estadístico para la creación de un modelo predictor, requerimos conocer en nuestra población cuales con los factores con mayor nivel de asociación y crear un modelo predictivo por medio de la regresión logística múltiple que funja como herramienta diagnóstica para la toma de decisiones de manera más oportuna, disminuir estancia intrahospitalaria y morbilidad del paciente.

Pregunta de investigación

¿Un nuevo modelo predictivo diagnóstico para coledocolitiasis basado en los datos clínicos y epidemiológicos del Órgano de Operación Administrativa Desconcentrada en Aguascalientes tiene mayor fuerza de asociación con respecto a las Guías Internacionales ASGE 2019?

Objetivo primario

- a. Generar un nuevo modelo predictivo diagnóstico para coledocolitiasis en adultos de las unidades del Órgano de Operación Administrativa Desconcentrada en Aguascalientes y compararlo con los criterios internacionales actuales

Objetivos secundarios

- a. Recolectar de variables predictoras de los expedientes para el diagnóstico de coledocolitiasis.
- b. Obtener de los coeficientes B de cada una de las variables para identificar las que

presentan mayor valor predictivo para el modelo.

- c. Generar un nuevo modelo predictivo por medio de regresión logística con base en los datos clínicos y epidemiológicos obtenidos
- d. Calcular la sensibilidad, especificidad, el valor predictivo positivo, valor predictivo negativo, razón de verosimilitud positiva y negativa, así como OR ajustada de las guías ASGE 2019 para el diagnóstico de coledocolitiasis.
- e. Calcular la sensibilidad, especificidad, el valor predictivo positivo, valor predictivo negativo razón de verosimilitud positiva y negativa, así como OR ajustada del modelo predictivo creado.
- f. Comparar las OR ajustadas obtenidas

Hipótesis

H0 El nuevo modelo predictivo no tiene una OR ajustada mayor a 0.5 con respecto a la OR ajustada de los criterios ASGE

H1 El nuevo modelo predictivo tiene una OR ajustada mayor a 0.5 con respecto a la OR ajustada de los criterios ASGE

Material y métodos diseño de investigación

Se realizó un estudio a nivel analítico, observacional, retrospectivo incluyendo a pacientes que se les practicó CPRE o colangiografía transoperatoria por coledocolitiasis en las 3 clínicas del IMSS de Órgano de Operación Administrativa Desconcentrada de Aguascalientes, desde enero 2021 hasta diciembre 2023. Se identificaron en el expediente físico y electrónico los valores de edad, sexo, IMC, si es portador de diabetes mellitus, si padece hipertensión arterial, hábito tabáquico y consumo de alcohol, el diagnóstico de envió a CPRE o exploración de vías biliares, el grado de coledocolitiasis según la ASGE, niveles de fosfatasa alcalina y la dilatación del colédoco en milímetros, visualización de lito en conducto biliar común, niveles de bilirrubina total, niveles de AST, ALT, GGT, si presentó cuadro de colangitis y/o pancreatitis biliar. Cada una de las variables se tomó del expediente físico, electrónico del hospital (PHEDS y ECE), las cifras de laboratorio se buscaron en el sistema institucional (ILABWEB), los parámetros imagenológicos se buscaron en el reporte emitido por radiólogo, en caso donde no se contó con reporte, se realizó la medición en el corte transversal de la imagen usando el sistema institucional (RAYOSX). En casos donde no se contó con registro físico o electrónico de las variables, el paciente no se incluyó en el estudio.

Universo de trabajo

Lugar donde se desarrolló en estudio

Las 3 clínicas del Instituto Mexicano del Seguro Social del Órgano de Operación Administrativa Desconcentrada de Aguascalientes.

Universo de estudio

Pacientes mayores de 18 años que ingresaron al servicio de cirugía con riesgo intermedio y alto de coledocolitiasis por criterios clínicos, laboratoriales y de imagen establecidos por la ASGE 2019 durante el periodo de enero 2021 a diciembre 2023.

Población de estudio

Participaron en el estudio pacientes a partir de los 18 años de ambos sexos, con diagnóstico de riesgo intermedio y alto de coledocolitiasis que posteriormente fueron sometidos al tratamiento indicado y según la disponibilidad colangiopancreatografía endoscópica retrograda o colangiografía intraoperatoria con derechohabencia al Instituto Mexicano del Seguro Social con adscripción las 3 clínicas de la delegación de Aguascalientes.

Unidad de observación

Pacientes que acudieron a atención médica y se les estableció el diagnóstico de riesgo intermedio y alto de coledocolitiasis sometidos a colangiopancreatografía endoscópica retrograda o colangiografía intraoperatoria desde enero 2023 a diciembre 2023 indicada por cirujanos con derechohabencia al Instituto Mexicano del Seguro Social con adscripción a cualquiera de las 3 clínicas delegación de Aguascalientes.

Unidad de análisis

Posterior a la lista de verificación se vaciaron los resultados obtenidos en una base de datos en el programa Excel para el registro de las variables elegidas de cada paciente que fueron diagnosticados con riesgo intermedio y alto de coledocolitiasis y sometidos a CPRE o colangiografía intraoperatoria en pacientes adscritos a cualquiera de las 3 clínicas del Órgano de Operación Administrativa Desconcentrada Aguascalientes, corroborando con esto el diagnóstico, siendo estos 2 procedimientos el estándar de oro.

Criterios de inclusión

Los criterios de inclusión para el protocolo fueron acorde a lo revisado en la búsqueda bibliográfica:

- a. Los pacientes que ingresaron con el diagnóstico de riesgo intermedio y alto de coledocolitiasis y se les prescribió algún procedimiento diagnóstico-terapéutico como CPRE (siendo únicamente la 1era vez) por el diagnóstico de coledocolitiasis o colangiografía transoperatoria.

- b. Los pacientes que se les realizó exploración de vías biliares (abierta o por laparoscopia) sin CPRE previa
- c. Pacientes que ingresaron con el diagnóstico de pancreatitis biliar y se resolvió mediante CPRE.
- d. Pacientes que ingresaron con el diagnóstico de colangitis que se resolvió por CPRE o exploración de vías biliares.
- e. Pacientes mayores de 18 años
- f. Derechohabientes de OOAD Aguascalientes

Criterios de no inclusión

Se excluyeron del estudio los pacientes con alguna de las siguientes características:

- a) Instrumentación previa de la vía biliar***

Criterios de exclusión

- a. Pacientes con registro incompleto en las variables en expediente***
- b. Pacientes con bajo riesgo de coledocolitiasis***
- c. Pacientes a los que se les encontró patología maligna en el reporte de la CPRE.***
- d. Reporte de estenosis del conducto biliar por estudios de imagen***

Tamaño de muestra

Para el cálculo de muestra se utilizó el programa Epidat en su versión 4.2, en el apartado de casos y controles con grupos independientes. Según la literatura internacional, se encontró que pacientes expuestos presentaron una proporción de éxito (positivo para el diagnóstico) del 59.9%. (12). Extrapolando estos datos realizamos nuestro cálculo de muestra con un IC del 95% y una potencia del 80% con una odds ratio a detectar del 2.25.

Datos:

Proporción de casos expuestos:59,900%

Proporción de controles expuestos:39,900%

Odds ratio a detectar:2,250

Número de controles por caso:1

Nivel de confianza:95,0%

Resultados:

Proporción de casos expuestos:59,900%

Proporción de controles expuestos:39,900%

Odds ratio a detectar:2,250

Número de controles por caso:1

Nivel de confianza:95,0%

Potencia (%)	Tamaño de la muestra*		
	Casos	Controles	Total
80,0	97	97	194

Resultados:

Ilustración 1. Calculo de muestra

Se calculó una muestra de 194. Esta muestra se obtuvo de las 3 clínicas de atención de manera proporcional a los casos que se presentaron en 1 año en cada unidad. Por lo que el tamaño muestral en cada unidad fue el siguiente:

Muestreo aleatorio estratificado

Entrada manual: Número de estratos: 3

Entrada automática: ☒ Abrir datos individuales ☐ Abrir datos agregados

Reparto de la muestra: Reparto proporcional al tamaño de los estratos

Estrato	Tamaño del estrato	Tamaño de la muestra
1	79	76
2	67	64
3	57	54

Tamaño de la muestra: 194 **Distribuir**

Datos: ☐ Ordenar la muestra

Presentar resultados: ☒ En pantalla ☐ En fichero

Ocultar Calcular Limpiar Cerrar

Ilustración 2.Distribución del tamaño de muestra entre los 3 hospitales.

Se obtuvieron 76 expedientes de la clínica N° 1, 64 expedientes de la clínica N° 2 y 54 expedientes de la clínica N° 3. Para la recolección de expedientes en cada unidad se utilizó el muestreo aleatorizado simple.

Operacionalización de las variables

Se empleó las variables que han resultado significativamente asociadas en la

literatura. Se hizo un registro empleando una base de datos de la edad, el sexo, IMC, portador de diabetes mellitus, hipertensión arterial, hábito tabáquico y consumo de alcohol, el diagnóstico de envió a CPRE o exploración de vías biliares, niveles de fosfatasa alcalina y la dilatación del colédoco en milímetros, visualización de lito en conducto biliar común, niveles de bilirrubina total, niveles de AST, ALT, GGT, si presentó cuadro de colangitis y/o pancreatitis biliar.

VARIABLE	DEFINICIÓN CONCEPTUAL	OBTENCIÓN OPERACIONAL	INDICADOR	VALOR O ESCALA	TIPO DE VARIABLE	CODIFICACIÓN
Edad	Espacio de años que han corrido de un tiempo a otro. (22)	Intervalo de años desde la fecha de nacimiento registrada en su CURP	Numérico (años)	Años	Cuantitativa continua	El valor medido
Sexo	Condición orgánica, masculina o femenina. (22)	Clasificación sexual del paciente conforme a sus características sexuales secundarios.	Descrito en expediente	Femenino Masculino	Cualitativa nominal dicotómica	Femenino = 0 Masculino =1
IMC	Índice de Quetelet es un indicador antropométrico que se calcula dividiendo el peso en kilogramos entre la estatura en metros elevada al cuadrado. Permite interpretar con mayor precisión el peso con relación a la estatura. (23)	Cociente que resulta entre la división del peso del paciente y su estatura.	Numérico (kg/m ²)	Coeficiente de los kilos del paciente entre la altura en metros	Cuantitativa continua	El valor medido
Diabetes mellitus	La diabetes tipo 2 es un tipo muy heterogéneo enfermedad con edad variable en el momento de inicio, relacionada grado de obesidad, resistencia a la insulina y tendencia a desarrollar complicaciones. (24)	Paciente que se conozca con diagnóstico de diabetes mellitus, con algún registro en el expediente de su enfermedad.	Descrito en expediente	Diabético No diabético	Cualitativa nominal dicotómica	No Diabético =0 Diabético=1
Hipertensión arterial	Valores repetidos de presión arterial sistólica en el consultorio de 140 mmHg y/o presión arterial diastólica de 90 mmHg. (25)	Paciente que se conozca con diagnóstico de este padecimiento e ingiera medicamentos para ello, con algún registro En el expediente de su enfermedad.	Descrito en expediente	Hipertenso No hipertenso	Cualitativa nominal dicotómica	No hipertenso = 0 Hipertenso = 1
Tabaquismo	Intoxicación producida por el abuso del tabaco. (26)	Consumo de registrado en la historia clínica.	Descrito en expediente	Positivo Negativo	Cualitativa nominal	Negativo=0 Positivo= 1
Alcoholismo	Abuso en el consumo de bebidas alcohólicas. Muchos accidentes de tráfico se deben al alcoholismo. (27)	Consumo de alcohol registrado en la historia clínica.	Descrito en expediente	Positivo Negativo	Cualitativa nominal	Negativo=0 Positivo= 1
Diagnostico pre-CPRE o pre-CIO	Determinación de la naturaleza de una enfermedad mediante la observación de	Indicación de realización de procedimiento debido a	Descrito en expediente	Coledocolitiasis Colangitis Pancreatitis	Cualitativa nominal	Coledocolitiasis =0 Colangitis=1

	sus síntomas. (28)	coledocolitiasis.		biliar		Pancreatitis biliar =2
Riesgo de coledocolitiasis según ASGE	Clasificación de recomendaciones, la metodología de evaluación, desarrollo y evaluación para la coledocolitiasis. (6)	Clasificación a la que pertenece cada paciente según la ASGE	Descrito en expediente	Medio Alto	Cualitativa ordinal	Bajo =0 Medio = 1 Alto =2
Fosfatasa alcalina (FA)	La prueba ALP se usa para detectar y vigilar las enfermedades óseas o hepáticas. Se encuentra en muchos tejidos, las concentraciones más altas se pueden localizar en el hígado, el epitelio del tracto biliar y el hueso. (29)	Indicador de la función hepática, los rangos de corte en laboratorio del hospital son: 38-126u/l	Descrito en el sistema de laboratorio	Unidades entre 1 litro	Cuantitativa continua	El valor arrojado por el reporte de laboratorio
Bilirrubina total (BT)	Es la suma de la bilirrubina conjugada (directa) y la no conjugada (indirecta). (29)	Indicador de la función hepática, los rangos de corte en laboratorio del hospital son: 0.20 a 1.3 mg/dl	Descrito en el sistema de laboratorio	Los miligramos contenidos en un 1 decilitro.	Cuantitativa continua	El valor arrojado por el reporte de laboratorio
Alanino aminotransferasa (ALT)	Enzima que se encuentra de manera en el hígado y, en menores cantidades, en riñones, corazón y músculo esquelético. (29)	Indicador de la función hepática, los rangos de corte en laboratorio del hospital son: Varones menores a 50 U/L Mujeres: menor a 35 U/L.	Descrito en el sistema de laboratorio	Las unidades de enzima encontradas en un litro.	Cuantitativa continua	El valor arrojado por el reporte de laboratorio
Aspartato aminotransferasa (AST)	Enzima se encuentra en concentraciones muy elevadas dentro de tejidos altamente metabólicos, como el músculo cardíaco, Células hepáticas, células musculoesqueléticas y en menor grado, riñones, páncreas. En presencia de enfermedad o lesión que afecte las células de estos tejidos, las células se lisan. (29)	Indicador de la función hepática, los rangos de corte en laboratorio del hospital son: Adultos. 15-46 U/L Mujeres: 14-36 U/L Varones: 17-59 U/L.	Descrito en el sistema de laboratorio	Las unidades de enzima encontradas en un litro.	Cuantitativa continua	El valor arrojado por el reporte de laboratorio
Gama glutamiltransferasa (GGT)	La enzima GGTP participa en la transferencia de aminoácidos y péptidos a través de la membrana celular y tal vez participe en el metabolismo del glutatión. (29)	Indicador de la función hepática, los rangos de corte en laboratorio del hospital son: 12 a 43 U/L	Descrito en el sistema de laboratorio	Las unidades de enzima encontradas en un litro.	Cuantitativa continua	El valor arrojado por el reporte de laboratorio
Grosor de colédoco	Formación tubular por arriba de la rama derecha portal. (30)	Milímetros que abarca el calibre del colédoco por ultrasonido en su porción media.	Descrito en el reporte de ultrasonido/ medido de imágenes corte transversal	Milímetros medidos por programa electrónico de ultrasonido	Cuantitativa continua	El valor reportado por radiólogo en reporte oficial o medición en software en vista sagital en porción media del colédoco.
Lito visible dentro de la vía biliar	Imagen hiperecogénica en un fondo hipoecoico / anecoico que deja Sombra acústica	Visualización de una hiperecogenicidad que se encuentre dentro de la vía biliar.	Descrito en el reporte de ultrasonido/ visualizado en imágenes corte	Se observa o reporte lito No se observa ni	Cualitativa nominal	Sin lito = 0 Con lito =1

	posterior. (30)		transversal	reporta lito		
Pancreatitis biliar	Establecimos el diagnóstico de pancreatitis conforme a las guías de Atlanta 2012 que definen el diagnóstico cuando el paciente cumple 2 de las siguientes características: 1) dolor abdominal (persistente, severo en epigastrio con irradiación a la espalda), 2) niveles de lipasa (o amilasa) al menos 3 veces más del límite normal y 3) características imagenológicas en tomografía abdominal contrastada, resonancia magnética o ultrasonido abdominal. (31) Para asegurar su origen biliar, el estudio de imagen debería reportar litos en el conducto biliar, una dilatación mayor a 6 mm o colelitiasis.	establecimos el diagnóstico de pancreatitis conforme a las guías de Atlanta 2012.	Descrito en expediente	Pancreatitis Sin pancreatitis	Cualitativa nominal	Sin pancreatitis =0 Pancreatitis =1
Colangitis aguda	acorde a las guías Tokyo del 2018 según estas guías se define el diagnóstico de colangitis aguda al presentar 1 criterio A (fiebre, escalofríos o evidencia laboratorial de respuesta inflamatoria sistémica; 1 criterio B (ictericia o enzimas hepáticas anormales y un criterio C (que en estudios de imagen se observe dilatación del conducto biliar o evidencia de otra etiología como lito, estenosis, portador de Stent, entre otros). (32)	la definimos acorde a las guías Tokyo del 2018.	Descrito en expediente	Colangitis Sin colangitis	Cualitativa nominal	Sin colangitis = 0 Colangitis =1

Tabla 1. Se muestran las 18 variables, su significado y las unidades en las que serán medidas

Procedimientos para recolección de datos

Posterior al contar con el registro del protocolo en la plataforma SIRELCIS. Se solicitó la aprobación del proyecto a los jefes del departamento de Cirugía General de las 3 clínicas del IMSS en Aguascalientes, así como a los jefes de enseñanza de cada una de las clínicas. Se realizó la solicitud de participación de investigador principal y asociados.

Posteriormente por medio de la plataforma institucional SIRELCIS, el protocolo se sometió a valoración por el Comité de Ética en Investigación y el Comité Nacional de Investigación en Salud y tras su aprobación se solicitó la autorización a los directivos de las 3 clínicas del IMSS para la revisión de expedientes.

Se designó como mínimo 3 días por semana en días hábiles para revisión de expedientes (1 día en cada una de las unidades) en turno matutino y vespertino dependiendo de la indicación de cada uno de los jefes de departamento de cirugía para la revisión de los expedientes. Se creó formatos de solicitud de expedientes, en otras clínicas se usó el establecido en la clínica con la firma del jefe de departamento de cirugía para su extracción del departamento de archivo. Cada vez que se solicitó un expediente se brindó el formato de solicitud (Anexo A), se regresó los expedientes al término de su revisión o al final de la semana manteniendo su resguardo en el departamento de jefatura. Al devolver el expediente a archivo se solicitó el vale como prueba de la devolución de dicho documento.

Se realizó una revisión exhaustiva del expediente que aseguró contenga todas las variables que se incluirán en el estudio. Al hacer la recolección de datos, este se supervisó por 2 investigadores. Así mismo, la base de datos se compartió por todos los investigadores participantes de manera que todos pudieron tener el mismo registro a la vez, validaron el llenado y la información obtenida. Con cada uno de los expedientes seleccionados se realizó el llenado de una base de datos en el programa Excel (Anexo B). Este archivo se compartió con todos los investigadores que se encargaron de la recolección de datos para el registro simultáneo de las variables, una doble verificación del llenado correcto y aseguraron el respaldo del contenido en caso de pérdida por alguno de los investigadores.

No se contó con un contacto directo con el paciente. Se garantizó con base a la ley de protección de datos personales la confidencialidad de sus datos personales y médicos. Estos fueron usados únicamente para fines de investigación y se mantuvieron el anonimato de los pacientes en el reporte del estudio.

Los investigadores nos comprometimos éticamente a realizar el muestreo completo de los 305 pacientes que nuestro cálculo arroja con la veracidad que el expediente nos brinda. Al contar con la base de datos completa para la obtención de los resultados primarios se realizó una asociación de las variables recolectadas con Software R y la elección de las variables con mayor significancia estadística y por otro lado las que muestran una significancia positiva para el diagnóstico de la enfermedad.

Al contar con estas variables organizadas sistemáticamente, se acudió al Departamento de Estadística en el edificio 30 de la Universidad Autónoma de Aguascalientes donde el doctor en ciencias con orientación en Ciencias de la

Computación capacitó y guio con el desarrollo del modelo predictivo y su instrumentación con base a la medición de su desempeño.

Instrumentos a utilizar

Para la recolección de datos: se utilizó el expediente físico del paciente resguardado en el departamento de archivo de cada una de las clínicas del Órgano de Operación Administrativa Desconcentrada de Aguascalientes. La recolección de los parámetros de laboratorio se llevó a cabo en la base de datos institucional. Cada una de las variables se extrajo con las siguientes normas establecidas por nosotros:

Edad: se midió el intervalo de tiempo en años desde la fecha de nacimiento marcada por el C.U.R.P. que se encuentra registrado en los documentos del expediente.

Sexo: se extrajo con base a lo redactado en la nota médica de ingreso.

IMC: Se tomó con base al peso y talla especificado en la valoración preoperatoria de medicina interna y/o anestesiología. Se realizó la división con 2 decimales sin aplicar redondeo.

Diabetes mellitus: se tomó este diagnóstico de la nota de valoración preoperatoria de medicina interna y/o anestesiología.

Hipertensión arterial: se tomó este diagnóstico de la nota de valoración preoperatoria de medicina interna y/o anestesiología.

Tabaquismo: se tomó este registro de la nota de valoración preoperatoria de medicina interna, anestesiología y/o notas médicas.

Alcoholismo: se tomó este registro de la nota de valoración preoperatoria de medicina interna, anestesiología y/o notas médicas.

Diagnóstico pre-CPRE o pre-CIO: se tomó de las notas médicas de cirugía general y/o formato de solicitud subrogada de CPRE.

Fosfatasa alcalina (FA): el registro se buscó los laboratorios del paciente en el sistema institucional ILABWEB y/o notas de expediente donde se especificó la fecha de toma de laboratorios.

Bilirrubina total (BT): se buscó en los laboratorios del paciente en el sistema institucional ILABWEB y/o notas de expediente donde se especifique fecha de toma de laboratorios.

Alaninoaminotransferasa (ALT): el registro se buscó los laboratorios del paciente en el sistema institucional ILABWEB y/o notas de expediente donde se especifique

fecha de toma de laboratorios.

Aspartato aminotransferasa (AST): se tomó el registro buscando los laboratorios del paciente en el sistema institucional ILABWEB y/o notas de expediente donde se especifique fecha de toma de laboratorios.

Gama glutamiltransferasa (GGT): se tomó el registro buscando los laboratorios del paciente en el sistema institucional ILABWEB y/o notas de expediente donde se especifique fecha de toma de laboratorios.

Grosor de colédoco: la medida del grosor del colédoco en milímetros del reporte realizado por el radiólogo. En los casos de no se contó con el reporte, se midió por uno de los investigadores desde la parte media del colédoco, con el programa institucional RAYOSX.

Lito visible dentro de la vía biliar: se tomó este parámetro del reporte realizado por el radiólogo. En los casos donde no se contó con el reporte, se visualizó por uno de los investigadores el corte transversal del ultrasonido (o estudios de imagen con los que cuente como colangiografía magnética o tomografía) usando el programa institucional RAYOSX.

Pancreatitis biliar: se tomó de las notas médicas de cirugía general y/o formato de solicitud subrogada de CPRE.

Colangitis aguda: se tomó de las notas médicas de cirugía general y/o formato de solicitud subrogada de CPRE.

Para el registro de los datos: se utilizó una tabla realizada exclusivamente para estos fines en el programa Excel.

Para el análisis estadístico: se utilizó el software R exportando la base de datos creada en Excel para el análisis descriptivo y las métricas que determinaran la significancia de nuestras variables. Se realizó el desarrollo del modelo con los comandos correspondientes y por último se medirá el desempeño del modelo.

Métodos para el control y calidad de los datos

Para garantizar la verosimilitud de los datos, los investigadores realizaron el registro de los datos inmediatamente hacia la base de datos de Excel. Posteriormente, como control de calidad y precisión, el investigador principal verificó los datos de la recolección de datos y corroboró la veracidad y correcta captura de información que el investigador asociado hizo en la hoja de Excel.

Análisis estadístico de los datos

Se utilizó el programa R (R software) debido a que se trata de un programa

estable, maduro con una trascendencia de 30 años, observado un incremento de la cantidad de usuarios. Es un software “free”, este adjetivo significa que es un programa que puede ser utilizado en varios sistemas operativos (es multiplataforma) además de estar accesible a todo público.

Se importó la base de datos creada en Excel al programa R con sus variables expuestas según su codificación en la tabla previa.

Se dividió a los pacientes en 2 grupos: aquellos que tuvieron un diagnóstico confirmatorio de coledocolitiasis por CPRE o colangiografía transoperatoria y los que se les descartó por el mismo método. Surgieron 2 grupos: pacientes con coledocolitiasis y sin coledocolitiasis.

Con el grupo de pacientes con coledocolitiasis se identificó 2 tipos de variables: las bivariadas y las multivariadas. Con cada una de ellas se estimó los odds ratio ajustados de cada variable para demostrar su poder de diagnóstico positivo y controlar la confusión. Se consideró una $p < 0.05$ y se realizó el cálculo de los niveles de confianza. Se utilizó nivel de significancia para las variables bivariadas y multivariadas. Se identificó y apartó del resto las variables que portaron valor significativo ($p < 0.05$) para el diagnóstico lo que representó su potencial positivo para la enfermedad.

Realizamos un catálogo de comandos específicos para realización del modelo predictivo desde *read.csv*, para la apertura de los documentos dentro del programa, *lm (formula)* para el diseño del modelo predictivo e *hist (fórmula)* para la realización de histogramas, por mencionar algunos. Los códigos que fueron utilizados se pueden consultar al final de este documento. (Anexo D)

Las variables cuantitativas continuas fueron 8, se expresaron con los valores exactos a 2 decimales, estas fueron: edad, índice de masa corporal, niveles de fosfatasa alcalina, bilirrubinas totales, alaninoaminotransferasa, aspartato aminotransferasa, gama glutamiltransferasa y grosor de colédoco.

Las variables dicotómicas fueron 8, se organizaron en código binario definidas como positivas 1 y negativas 0. Estas variables fueron: sexo (femenino = 0, masculino = 1), diabetes mellitus (no diabético = 0, diabético = 1), hipertensión arterial (hipertenso = 0, no hipertenso = 1), tabaquismo (negativo = 0, positivo = 1), alcoholismo (negativo = 0, positivo = 1), lito visible dentro de la vía biliar (sin lito = 0, con lito = 1), pancreatitis biliar (sin pancreatitis = 0 pancreatitis = 1) y colangitis aguda (sin colangitis = 0, colangitis = 1).

Las variables politómicas fueron 2, se organizaron en lenguaje binario dependiendo

de su categoría: diagnóstico inicial antes de la realización de CPRE o colangiografía (coledocolitiasis= 0, colangitis=1, pancreatitis biliar =2), riesgo de coledocolitiasis según ASGE (medio = 1, alto =2)

Se realizó una obtención de métricas de ajuste del modelo a los datos como es el coeficiente B y un análisis de componentes principales, se usó el software R con la finalidad de solucionar la multicolinealidad. Se construyó el modelo predictivo usando una regresión logística múltiple con las variables seleccionadas en base a estas métricas mencionadas.

El grupo de comparación para nuestro modelo predictivo fueron los pacientes a los que se les practicó CPRE (estándar de oro), exploración de vías biliares o colangiografía y resultó positiva para coledocolitiasis. Se tomó la probabilidad preprueba establecida por la literatura del 15%.

Se obtuvo el diseño del modelo, con la misma población seleccionada se utilizó la estrategia de bootstrap al menos en 20 ocasiones, esto se apegó a lo establecido en el estatuto TRIPOD, se realizó la medición del OR en la muestra aplicando los criterios establecidos por ASGE 2019 y aplicando el modelo creado. Se procedió a la comparación entre las OR y el área bajo la curva (AUC) de cada prueba diagnóstica. Se obtuvo las métricas de sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, accuracy, precisión (por su traducción en inglés referido como “accuracy” en algunas bibliografías, aunque este término llegue a denotar la veracidad y precisión), razón de verosimilitud positiva y negativa.

Aspectos éticos

No se tuvo contacto directo con los pacientes. Se garantizó en todo momento la discreción, resguardo de la información por 5 años, respeto de la identidad de los pacientes como de sus datos conforme a la Ley General de Datos personales.

Los procedimientos que se emplearon cumplieron los establecido por la Ley General de Salud en materia de Investigación para la Salud y la declaración de Helsinki y de la Asociación Médica Mundial.

Acorde a lo referido en el Procedimiento para la evaluación, registro, seguimiento, enmienda y cancelación de protocolos de investigación en salud presentados ante los Comités Locales de investigación en Salud y los comités de ética en Investigación 2810-003-002 dentro del Instituto Mexicano del Seguro Social se consideró una investigación **sin riesgo** debido a que no utilizamos intervenciones con los pacientes. Nuestro estudio fue observacional y no influyó de ninguna manera el tratamiento que pudo recibir los pacientes en hospitalizaciones futuras.

Debido a que se trata de un proyecto que involucró las 3 clínicas del Órgano de Operación Administrativa Desconcentrada de Aguascalientes, el proyecto será sometido para su autorización por el comité nacional del Instituto Mexicano del Seguro Social.

El presente estudio cumplió con los requisitos mencionados en los artículos del título quinto de la Ley Federal de Salud dedicados a la investigación en Salud citando a los artículos 96, 97, 98, 99, 100, 101 y 102.

Con este estudio se pretendió obtener una herramienta diagnóstica sin costo válida para la población con sospecha de coledocolitiasis de la ciudad de Aguascalientes. Se verán beneficiados los pacientes al no ser sometidos estudios innecesarios, atención terapéutica oportuna y disminución de la morbilidad en esta patología. Como institución nos beneficiaremos en contar con una herramienta útil y disponible para su aplicación en cualquier clínica del IMSS de la OOAD Aguascalientes, se pretende disminuir el costo a la atención que les brinda a los pacientes por una disminución de la estancia intrahospitalaria y disminuir los riesgos en la morbilidad que impliquen otra directriz terapéutica.

Aspectos de bioseguridad

Para este estudio no utilizamos material biológico, transporte de muestras, cepas de virus, bacterias, hongos u otro tipo de microorganismos. Uso de animales ni sustancias tóxicas que pudiera poner en riesgo a los investigadores o personal de la unidad.

Recursos, financiamiento y factibilidad

Recursos materiales: bolígrafos, libreta, carpetas de almacenamiento.

Recursos digitales: internet portátil, programa Excel, programa Epidat 4.2, R software.

Recursos humanos:

Investigador principal: Dra. Ariadna Beltrán Rico

Tutora: Dra. Ivonne González Espinosa

Tutor asociado: Dr. José Luis Bizueto Monroy

Tutor asociado: Dr. Rogelio Salinas Gutiérrez

Recursos financieros: Los recursos materiales fueron los mismo utilizados de manera personal. Los recursos digitales con acceso libre en la red y otros con

acceso libre en la plataforma de recursos utilizando el correo institucional de la Universidad Autónoma de Aguascalientes.

Resultados

Se realizó la recolección de datos de 305 pacientes con riesgo intermedio y alto de coledocolitiasis en las 3 clínicas de la OOAD Aguascalientes integrándolos con código binario en una matriz de datos utilizando el programa Excel versión 2502. La matriz de datos es guardada en formato separado por comas (CSV) requerido para su importación en software R studio versión 2024.12.1.

Se recolectaron 18 variables: edad, sexo, índice de masa corporal (IMC), diabetes mellitus (DM), hipertensión arterial (HTA), tabaquismo, alcoholismo, diagnóstico pre- CPRE o pre colangiografía intraoperatoria, el riesgo ASGE clasificado a su ingreso hospitalario, niveles de fosfatasa alcalina (FA), niveles de bilirrubina total (BT), directa (BD) e indirecta (BI), alaninoaminotransferasa (ALT), aspartato aminotransferasa (AST), gama glutamiltransferasa (GGT), grosor de colédoco en estudio de imagen y si el lito es visible o no por este medio, si el paciente presentó pancreatitis biliar y/o colangitis durante la hospitalización.

Se realizó la importación de la matriz en el software. De los 305 pacientes se excluyeron los que presentaron los siguientes diagnósticos: estenosis de colédoco (7 pacientes), cáncer de páncreas (9 pacientes), antecedente de realización de CPRE (24 pacientes, de los cuales: 16 pacientes por colelitiasis previas, 7 por lesión iatrogénica de vía biliar, 1 por lesión traumática de vía biliar), coledocolitiasis residual (4 pacientes), instrumentación previa de la vía biliar (realización de cirugías relacionadas a la vía biliar, 14 pacientes), pacientes con ictericia de etiología indeterminada a los que se les solicitó CPRE por otros servicios diferentes a cirugía general (2 pacientes) sospecha de tumor de la vía biliar (9 pacientes de los cuales: 6 con sospecha de tumor indeterminado, linfoma en vía biliar (1 paciente), tumor pseudopapilar de células claras (1 paciente), tumor de duodeno (1 paciente), pacientes menores de 18 años (4 pacientes), solicitud de alta voluntaria sin concluir el protocolo diagnóstico (2 pacientes), no se realiza CPRE, no es posible concluir la CIO (34 pacientes), quistes hepáticos (1 paciente) y metástasis hepáticas (1 paciente).

De los 305 pacientes se excluyeron 111 pacientes conformando nuestra base de datos con 194 pacientes. De los cuales 75 pacientes pertenecen al hospital 1, 63 pacientes al hospital 2, y 56 pacientes al hospital 3. Los 3 grupos pertenecientes a la OOAD Aguascalientes.

De los 194 pacientes con sospecha de coledocolitiasis, el 56.18% (109 pacientes) se comprobó el diagnóstico de coledocolitiasis, el 43.81% (85 pacientes) no presentaron evidencia de lito en colédoco al realizar el procedimiento diagnóstico-terapéutico (CPRE ó colangiografía transoperatoria). El sexo femenino conformó la muestra en un 67.01% (130 mujeres) y masculino un 32.98% (64 hombres). El diagnóstico médico por el cual el cirujano decidió la realización de CPRE y/o CIO fueron sospecha de coledocolitiasis en 142 pacientes (73.19%), clasificados con riesgo intermedio 90 pacientes (46.39%) y en riesgo alto 104 pacientes (53.60%); colangitis en 30 pacientes (15.46%) y pancreatitis en 22 pacientes (11.34%).

De los 109 pacientes con coledocolitiasis el promedio global de edad fue de 53.03 años, esto en base al sexo, en las mujeres, la edad promedio de presentación que encontramos en nuestra población fue de 49.89 años, y en hombres de 59.96 años. El sexo femenino fue más prevalente en 68.80% (75 mujeres) comparado con el 31.19% de hombres (34 pacientes masculinos).

El índice de masa corporal (IMC) promedio fue de 27.69 kg/m², en cambio al grupo de pacientes a los que no se les corroboró coledocolitiasis presentaron un IMC promedio 29.51 kg/m², con respecto al género, el promedio de IMC en el sexo femenino fue de 28.36 kg/m² (sd= 4.86), y en hombres el 26.21 kg/m² (sd=4.80). La diabetes mellitus fue una enfermedad concomitante en el 22.01 % (24 pacientes) y la hipertensión arterial en el 43.11% (47 pacientes) de los pacientes con presencia de coledocolitiasis. El 27.52 % (30 pacientes) practicaban el hábito tabáquico y el consumo de alcohol al grado de catalogarse como alcoholismo fue en 14.67% de los pacientes (16 pacientes).

En el grupo de los pacientes a los que se les corroboró coledocolitiasis. El diagnóstico presuntivo con la que se indicó el tratamiento (CPRE y/o colangiografía transoperatoria) fue la propia coledocolitiasis en 78 pacientes (71.55 %). De estos, 41 pacientes (37.61%) fueron catalogados con riesgo intermedio y 68 pacientes (62.38%) con riesgo alto según guías ASGE 2019; colangitis en 22 pacientes (20.18%) y pancreatitis en 9 pacientes (8.25%).

Respecto a los parámetros bioquímicos, se observó que el promedio de fosfatasa alcalina fue de 363.41 U/L (sd=258.81) en mujeres de 342.79 U/L (sd=253.35) y en hombres de 408.91 U/L (sd=268.68). El promedio de bilirrubina total fue de 5.18 U/L (sd=5.19), en mujeres de 4.30 U/L (sd=3.36), y en hombres de 7.13 U/L (sd=

7.56). Los niveles de alaninoaminotransferada (ALT) promedio fue de 303.83 U/L (sd=285.43) mientras que los de aspartatoaminotransferada (AST) de 308.36 U/L (sd=363.89). Los niveles de GGT promedio fueron de 633.24 U/L (sd=837.63).

La medición del colédoco promedio de los pacientes con coledocolitiasis fue de 9.68 mm (sd=6.28), en mujeres 8.72 mm (sd=3.65) en hombres fue de 11.81 mm (sd=9.63). El lito dentro del colédoco se observó solo en 31 pacientes (28.44%) a pesar de que posteriormente se corroboró la litiasis intraductal.

Durante su hospitalización previo al abordaje terapéutico la pancreatitis biliar se desarrolló en 25 pacientes (22.93%) y colangitis en 23 pacientes (21.10%)

Procesamiento de los resultados

A todas las variables cuantitativa se le aplicó la fórmula de Shapiro-Wilk para contrastar las normalidades de las variables. Las variables que mostraron una distribución anormal fueron: edad ($p=0.005633$), esta distribución inclinada más por el sexo femenino donde mostró una ($p=0.002738$), no así para los hombres ($p = 0.05163$) fosfatasa alcalina ($p < 2.992e-09$), bilirrubinas totales ($p < 1.057e-12$), ALT ($p < 3.478e-09$), AST ($p < 2.625e-14$), GGT ($p < 2.2e-16$) y el grosor del colédoco ($p < 3.272e-13$).

Para las variables cualitativas se utilizó la prueba “Chi cuadrada”. Con este método, las que presentaron una distribución anormal fueron riesgo ASGE ($p < 0.0085$) y lito visible por ultrasonido ($p < 0.04586$). El resto tuvo una distribución normal.

Las variables numéricas que presentaron una distribución normal se les aplicó la prueba T (T-student) y aquellas con una distribución anormal la prueba U de Man Whitney.

A continuación, se muestra la tabla con todas las variables y sus valores P finales obtenidos:

Variables predictoras	Sospecha coledocolitiasis (194)		Valor P
	Presente = 109	Ausente = 85	
Edad	53.03 (19.50)	49.25 (18.56)	0.217
Mujeres	46.86 (16.17)	46.24 (18.33)	0.342
Hombres	56.66 (18.65)	54.77 (18.00)	0.2621

Sexo			
Mujeres (%)	75 (68.80)	55 (64.70)	0.653
Hombres (%)	34 (31.19)	30 (35.29)	
IMC	27.69 (4.92)	29.51 (6.86)	0.04097
Mujeres	28.36 (4.86)	30.62 (7.49)	0.05422
Hombres	26.21 (4.80)	27.48 (5.03)	0.310
Diabetes mellitus			
Portador (%)	24 (22.01)	25 (29.41)	0.3128
No portador (%)	85 (77.98)	60 (70.58)	
Hipertensión arterial			
Portador (%)	47 (43.11)	28 (32.94)	0.195
No portador (%)	62 (56.88)	57 (67.05)	
Tabaquismo			
Activo (%)	30 (27.52)	24 (28.23)	1
No activo (%)	79 (72.47)	61 (71.76)	
Alcoholismo			
Padece (%)	16 (14.67)	16 (18.82)	0.5641
No padece (%)	93 (85.32)	69 (81.17)	
Diagnostico pre-CPRE			
Coledocolitiasis n= 142 (%)	78 (71.55)	64 (75.29)	
Colangitis n= 30 (%)	22 (20.18)	8 (9.41)	0.056
Pancreatitis biliar = 22 (%)	9 (8.25)	13 (15.29)	
Riesgo ASGE			
Intermedio (%)	41 (37.61)	49 (57.64)	0.0085
Alto (%)	68 (62.38)	36 (42.35)	
Fosfatasa alcalina	363.41 (258.81)	277.83 (259.13)	0.001
Mujeres	342.79 (253.35)	234.00 (146.92)	0.004
Hombres	408.91 (268.68)	358.18 (379.45)	0.059
Bilirrubina total	5.18 (5.19)	4.07 (3.80)	0.150
Mujeres	4.30 (3.36)	3.82 (3.94)	0.232
Hombres	7.13 (7.56)	4.53 (3.54)	0.303
Alanino	303.83 (285.43)	327.11 (473.72)	0.789

aminotrasferasa			
Mujeres	317.18 (285.81)	310.94 (354.72)	0.681
Hombres	274.37 (286.61)	356.76 (643.67)	0.861
Aspartato aminotransferasa	308.36 (363.89)	288.99 (398.23)	0.294
Mujeres	346.07 (410.25)	288.49 (377.32)	0.287
Hombres	225.16 (213.50)	289.90 (440.75)	0.798
Gama glutamiltransferasa	633.24 (837.63)	410.15 (336.33)	0.011
Mujeres	575.91 (924.49)	367.76 (310.99)	0.034
Hombres	759.70 (596.39)	487.89 (371.34)	0.073
Grosor de colédoco	9.68 (6.28)	8.13 (3.67)	0.118
Mujeres	8.72 (3.65)	7.98 (3.60)	0.254
Hombres	11.81 (9.63)	8.41 (3.84)	0.203
Lito visible			
Se observa (%)	31 (28.44)	13 (15.29)	0.0458
No se observa (%)	78 (71.55)	72 (84.70)	
Pancreatitis biliar			
Padece (%)	25 (22.93)	27 (31.76)	0.224
No padece (%)	84 (77.06)	58 (68.23)	
Colangitis aguda			
Padece (%)	23 (21.10)	13 (15.29)	0.397
No padece (%)	86 (78.89)	72 (84.70)	

Tabla 2. Muestra las 18 variables incluidas en el estudio. La columna PRESENTE se refiere a los pacientes con diagnóstico establecido de coledocolitiasis y AUSCENTE que tuvieron sospecha, pero se descartó. Los valores P en negritas indican las variables que presentaron significancia estadísticamente respecto a una $p < 0.050$

De las 18 variables únicamente 5 mostraron significancia estadística fueron índice de masa corporal (IMC) ($p < 0.04097$), riesgo ASGE ($p < 0.0085$), fosfatasa alcalina (FA) ($p < 0.001$), gamaglutamil transferasa (GGT) ($p < 0.011$) y la evidencia de lito visible por ultrasonido en colédoco ($p < 0.0458$).

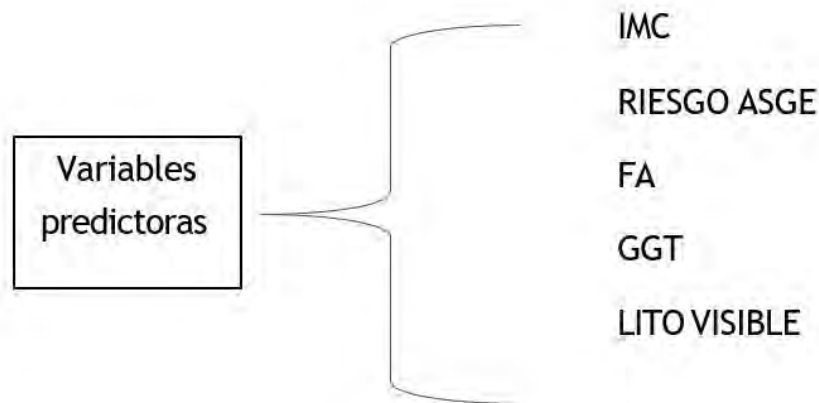


Ilustración 3. Variables que mostraron poder predictivo y fueron usadas para la elaboración del modelo.

Construcción del modelo: regresión logística múltiple

Se realizó una regresión logística múltiple con cada una de las variables predictoras con la finalidad de obtener la razón de momios, el intervalo de confianza, el P calculado dependiente de la regresión y el coeficiente B de cada variable. El coeficiente B representa cuanto aumenta el odds ratio por cada incremento de unidad y por lo tanto aumenta la significancia de la variable en cuanto al resultado final, es decir, la probabilidad de que el paciente tenga o no coledocolitiasis.

La fórmula de regresión logística que es la siguiente:

$$\ln\left(\frac{\pi}{1-\pi}\right) = B_0 + B_{1x} \dots$$

x = variable independiente

$\pi(x)$ = es la probabilidad de coledocolitiasis

$1 - \pi(x)$ = es la probabilidad de no tener coledocolitiasis

B = el grado en el que aumentará el odds ratio por cada incremento de unidad de la variable predictora x

Si despejamos la probabilidad de tener coledocolitiasis la fórmula se transforma en lo siguiente:

$$\pi(x) = \frac{\exp(B_0 + B_1x)}{1 + \exp(B_0 + B_1x)}$$

Si reemplazamos los factores por los conceptos médicos representativos, podemos interpretar la fórmula de la siguiente manera:

$$\ln\left(\frac{\text{probabilidad de coledocolitiasis}}{\text{probabilidad de no tener coledocolitiasis}}\right) = \text{interceptor de variable predictora} + \text{coeficiente } B \text{ de variable predictora}$$

El despeje de la fórmula es el siguiente:

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{\exp(\text{interceptor de variable predictora} + \text{coeficiente } B \text{ de variable predictora})}{1 + \exp(\text{interceptor de variable predictora} + \text{coeficiente } B \text{ de variable predictora})} \right)$$

Modelo de regresión logística con la variable “IMC”

En el software R se introdujo la función **glm()** con lo que nos arroja los siguientes datos:

Interceptor= 1.76747

Coeficiente B de la variable IMC= -0.05316

Valor P= 0.0356

Usando la fórmula:

$$\ln\left(\frac{\text{probabilidad de coledocolitiasis}}{\text{probabilidad de no tener coledocolitiasis}}\right) = \text{interceptor} + \text{coeficiente B IMC}$$

Si despejamos:

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{\exp(\text{interceptor} + \text{coeficiente B}_{IMC})}{1 + \exp(\text{interceptor} + \text{coeficiente B}_{IMC})} \right)$$

Sustituimos:

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{\exp(1.76747 - 0.05316)}{1 + \exp(1.76747 - 0.05316)} \right)$$

$$\text{Probabilidad de coledocolitiasis} = 0.8473345$$

Coledocolitiasis=0.8473345, esto representa la probabilidad de tener coledocolitiasis respecto y únicamente con el valor de IMC.

Obtenemos el odds ratio ajustado:

$$\text{Odds ratio adj} = \left(\frac{P}{(1 - P)} \right)$$

$$Odds\ ratio\ adj = \left(\frac{0.8473945}{(1 - 0.8473945)} \right)$$

$$Odds\ ratio\ ajustado = 5.552843$$

Con los valores obtenidos podemos deducir que

Valor de la hipótesis nula = interceptor de la variable IMC = 1.76747

Hipótesis alternativa = coeficiente B de la variable IMC = -0.05316

Valor P de la variable IMC = 0.0356

Dado que el valor P es menor a 0.05, rechazamos la hipótesis nula. Podemos afirmar que la variable IMC tiene poder explicativo en contener litos intraductales.

Modelo de regresión logística con la variable “Riesgo ASGE”

En el software R se introdujo la función **glm ()** con lo que nos arroja los siguientes datos:

Interceptor = 1.13368

Coeficiente B de la variable IMC = -0.04458

Coeficiente B de la variable RIESGO ASGE = 0.74205

Valor P de la variable RIESGO ASGE = 0.0133

Con la fórmula:

$$\ln \left(\frac{\text{probabilidad de coledocolitiasis}}{\text{probabilidad de no tener coledocolitiasis}} \right) = \text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{riesgo\ ASGE}$$

Despejamos:

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{\exp(\text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{riesgo\ ASGE})}{1 + \exp(\text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{riesgo\ ASGE})} \right)$$

Sustituimos:

$$\text{Probabilidad de colecolitiasis} = \left(\frac{\exp(1.13368 - 0.04458 + 0.74205)}{1 + \exp(1.13368 - 0.04458 + 0.74205)} \right)$$

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{6.24106}{7.24106} \right)$$

$$\text{Probabilidad de coledocolitiasis} = 0.8618987$$

Coledocolitiasis = 0.8618987, esta es la probabilidad de tener coledocolitiasis es de con respecto a 2 variables simultaneas IMC y el riesgo ASGE clasificado.

Obtenemos el OR ajustado:

$$\text{Odds ratio adj} = \left(\frac{P}{(1 - P)} \right)$$

$$\text{Odds ratio adj} = \left(\frac{0.8618987}{(1 - 0.8618987)} \right)$$

$$\text{Odds ratio adj} = 6.24106$$

Hipótesis nula = 1.13368

Hipótesis alternativa = 0.74205

Valor P de la variable RIESGO ASGE = 0.0133

Debido a que el valor P de la variable RIESGO ASGE es menor a 0.05 se descarta la hipótesis nula e inferimos que la clasificación que se otorga en los pacientes si tiene una asociación positiva para la presencia de coledocolitiasis.

Modelo de regresión logística con la variable "Fosfatasa alcalina (FA)"

En el software R se introdujo la función **glm ()** con las variables: **IMC+RIESGO ASGE+FA** lo que nos arroja los siguientes datos:

Interceptor= 0.6203412

Coeficiente B de la variable IMC= -0.0389987

Coeficiente B de la variable RIESGO ASGE= 0.7021587

Coeficiente B de la variable FA= 0.0011850

Valor P de la variable FA= 0.0724

Con la fórmula:

$$\ln\left(\frac{\text{probabilidad de coledocolitiasis}}{\text{probabilidad de no tener coledocolitiasis}}\right) = \text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{riesgo ASGE} + \text{coeficiente } B_{FA}$$

Despejamos:

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{\exp(\text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{riesgo ASGE} + \text{coeficiente } B_{FA})}{1 + \exp(\text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{riesgo ASGE} + \text{coeficiente } B_{FA})} \right)$$

Sustituimos:

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{\exp(0.6203412 - 0.0389987 + 0.7021587 + 0.0011850)}{1 + (0.6203412 - 0.0389987 + 0.7021587 + 0.0011850)} \right)$$

$$\frac{\text{Probabilidad de coledocolitiasis}}{\text{Probabilidad de no tener coledocolitiasis}} \equiv \left(\frac{1.284686}{0.7832464} \right)$$

Coledocolitiasis = 0.7832464, esta es la probabilidad de tener coledocolitiasis con

TESIS TESIS TESIS TESIS TESIS

respecto a las 3 variables simultaneas IMC, riesgo ASGE y las cifras de fosfatasa alcalina que presente el paciente.

Obtenemos el OR ajustado:

$$Odds\ ratio\ adj = \left(\frac{P}{(1 - P)} \right)$$

$$Odds\ ratio\ adj = \left(\frac{0.7832464}{(1 - 0.7832464)} \right)$$

$$Odds\ ratio\ adj = 3.613534$$

Hipótesis nula = 0.62034

Hipótesis alternativa = 0.00118

Valor P de la variable FA= 0.0724

El valor P de la variable es mayor a 0.5 por lo que no se puede descartar la hipótesis nula e infiere que la fosfatasa alcalina no tiene poder explicativo en la enfermedad de coledocolitiasis.

Modelo de regresión logística con la variable “gama glutamiltransferasa (GGT)”

En el software R se introdujo la función **glm ()** con las variables **IMC+RIESGO ASGE+FA+GGT** lo que nos arroja los siguientes datos:

Interceptor= 0.4399598

Coeficiente B de la variable IMC= -0.0414324

Coeficiente B de la variable RIESGO ASGE= 0.7582962

Coeficiente B de la variable FA= 0.0005316

Coeficiente B de la variable GGT = 0.0008773

Valor P de la variable GGT= 0.0444

Con la fórmula:

$$\ln\left(\frac{\text{probabilidad de coledocolitiasis}}{\text{probabilidad de no tener coledocolitiasis}}\right) = \text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{riesgo ASGE} + \text{coeficiente } B_{FA} + \text{coeficiente } B_{GGT}$$

Despejamos:

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{\exp(\text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{riesgo ASGE} + \text{coeficiente } B_{FA} + \text{coeficiente } B_{GGT})}{1 + \exp(\text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{riesgo ASGE} + \text{coeficiente } B_{FA} + \text{coeficiente } B_{GGT})} \right)$$

Sustituimos:

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{\exp(0.4399598 - 0.0414324 + 0.7582962 + 0.0005316 + 0.0008773 + .0444)}{1 + \exp(0.4399598 - 0.0414324 + 0.7582962 + 0.0005316 + 0.0008773 + .0444)} \right)$$

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{3.184301}{4.184301} \right)$$

$$\text{Probabilidad de coledocolitiasis} = 0.7610115$$

La probabilidad de **coledocolitiasis es de 0.7610115** tomando en cuenta las variables de IMC, el riesgo ASGE, los niveles de FA y de GGT del paciente simultáneamente. Obtenemos el OR ajustado

$$\text{Odds ratio adj} = \left(\frac{P}{(1 - P)} \right)$$

$$\text{Odds ratio adj} = \left(\frac{0.7610115}{1 - 0.7610115} \right)$$

$$\text{Odds ratio adj} = 3.184301$$

Hipótesis nula = 0.4399598

Hipótesis alternativa = 0.0008773

Valor P de la variable GGT = 0.0444

Debido a que el valor P de la variable GGT es menor a 0.5, podemos rechazar la hipótesis nula, lo que significa que GGT tiene un valor estadísticamente significativo para la probabilidad de coledocolitiasis.

Modelo de regresión logística con la variable "Lito"

En el software R se introdujo la función *glm ()* con las variables **IMC+RIESGO ASGE+FA + GGT+ LITO** lo que nos arroja los siguientes datos:

Interceptor= 0.3703030

Coeficiente B de la variable IMC= -0.0408966

Coeficiente B de la variable RIESGO ASGE= 0.5246047

Coeficiente B de la variable FA= 0.0006079

Coeficiente B de la variable GGT = 0.0009280

Coeficiente B de la variable LITO = 0.5974993

Valor P de la variable LITO = 0.1674

Con la fórmula:

$$\ln\left(\frac{\text{probabilidad de coledocolitiasis}}{\text{probabilidad de no tener coledocolitiasis}}\right) = \text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{\text{riesgo ASGE}} + \text{coeficiente } B_{FA} + \text{coeficiente } B_{GGT} + \text{coeficiente } B_{LITO}$$

Despejamos:

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{\exp(\text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{\text{riesgo ASGE}} + \text{coeficiente } B_{FA} + \text{coeficiente } B_{GGT} + \text{coeficiente } B_{LITO})}{1 + \exp(\text{interceptor} + \text{coeficiente } B_{IMC} + \text{coeficiente } B_{\text{riesgo ASGE}} + \text{coeficiente } B_{FA} + \text{coeficiente } B_{GGT} + \text{coeficiente } B_{LITO})} \right)$$

Sustituimos:

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{\exp(0.3703030 - 0.0408966 + 0.5246047 + 0.0006079 + 0.0009280 + 0.5974993)}{1 + \exp(0.3703030 - 0.0408966 + 0.5246047 + 0.0006079 + 0.0009280 + 0.5974993)} \right)$$

$$\text{Probabilidad de coledocolitiasis} = \left(\frac{4.276121}{5.276121} \right)$$

$$\text{Probabilidad de coledocolitiasis} = 0.8104668$$

La probabilidad de **coledocolitiasis es de 0.8104668**, al incluir simultáneamente las cifras de las variables IMC, el riesgo ASGE, valores de FA, GGT y la evidencia de visualización de lito en el ultrasonido.

Obtenemos el odds ratio ajustado con la siguiente fórmula:

$$\text{Odds ratio adj} = \left(\frac{P}{(1 - P)} \right)$$

$$\text{Odds ratio adj} = \left(\frac{0.8104668}{1 - 0.8104668} \right)$$

$$\text{Odds ratio adj} = 4.276121$$

La hipótesis nula es = 0.370303

La hipótesis alternativa = 0.5974993

El valor P de la variable LITO = 0.1674

Debido a que el valor P de la variable LITO es mayor a 0.05 no se puede rechazar la hipótesis nula e inferimos que el observar el lito durante la realización del ultrasonido no tiene el suficiente poder explicativo estadísticamente para ser significativo.

Para una mejor visualización de los datos, los valores se representan en la siguiente tabla de una manera exponencial (Tabla 2).

Variable	Odds ratio	IC 95%	Valor P	Coefficiente B
IMC	5.5528	0.90117 0.99564	0.035	0.9482
Riesgo ASGE	6.2410	1.17082 3.79976	0.013	2.1002
Fosfatasa alcalina	3.6135	0.9999 1.0025	0.072	1.0011
Gamaglutamil transferasa	3.184	1.0000 1.0017	0.044	1.000
Lito visible	4.2761	0.7868 4.3320	0.1674	1.8175

Tabla 3. Muestra el odds ratio, los intervalos de confianza con un error alfa de 5%, el valor P de cada variable y los coeficientes B. Estos últimos se ilustran exponencialmente para una mejor interpretación lectora.

Selección del modelo predictivo

Existen varias formas de hacer un pre procesamiento de los datos obtenidos para su optimización y emplear las variables que nos serán más representativas. En Machine Learning una de las herramientas para encontrar el mejor modelo es el criterio de información de Akaike (AIC), se describe como la siguiente formula:

$$AIC = 2K - 2X \ln(L)$$

K = número de parámetros del modelo

$\ln(L)$ = es la función de Log-verosimilitud para el modelo estadístico

En el software R introducimos la función `step()` usando el modelo de regresión logística con las 5 variables que se muestran en la tabla 2. El modelo predictivo con el valor AIC menor es el siguiente:

```
glm(formula = LITO_COLANGIO ~ IMC + GGT + LITO, family = "binomial", data = DATOS)
```

AIC: 258.6

Las variables: IMC, GGT y LITO forman parte del modelo elegido con base a los criterios de Akaike, significa que representa el mejor modelo.

Desempeño del modelo predictivo

Para medir el desempeño del modelo se obtuvieron las métricas de sensibilidad, especificidad, el valor predictivo positivo, el valor predictivo negativo, razón de verosimilitud positiva y negativa; el valor odds ratio ajustada del modelo predictivo creado.

Para el desempeño del modelo utilizamos el método Bootstrap con 20 repeticiones. Primero separamos nuestra muestra en 2 grupos en una proporción 70:30. El grupo del 70% se utilizó para entrenar el modelo y el grupo del 30% para medir su desempeño.

El entrenamiento del modelo se realiza mediante el software con la función `glm()` y la base de datos a emplear se sustituye por la de entrenamiento. Posteriormente utilizando este mismo modelo se realiza la medición de su desempeño con la función `predict.glm()` **que** predecirá los resultados usando el otro grupo de pacientes destinado para la prueba.

Con este método logramos un grupo representativo para una medición confiable. Los datos obtenidos por este modelo se presentan en la tabla 3:

Bootstrap	Se	Sp	VPP	VPN	A	P	OR adj	AUC	RV+	RV-
1	0.45714	0.78787	0.69565	0.57777	0.61764	0.69565	3.12781	0.62510	2.15510	0.68901
2	0.52500	0.72413	0.72413	0.52500	0.60869	0.72413	2.90131	0.68620	1.90312	0.65595
3	0.43902	0.63636	0.69230	0.37837	0.50793	0.69230	1.36956	0.59201	1.20731	0.88153
4	0.43902	0.89655	0.85714	0.53061	0.62857	0.85714	6.78260	0.72918	4.24390	0.62570
5	0.30232	0.85185	0.76470	0.43396	0.51428	0.76470	2.49166	0.61584	2.04069	0.81900
6	0.42105	0.86666	0.80000	0.54166	0.61764	0.80000	4.72727	0.72456	3.15789	0.66801
7	0.35714	0.75862	0.68181	0.44897	0.52112	0.68181	1.74603	0.60344	1.47959	0.84740
8	0.48837	0.67741	0.67741	0.48837	0.56756	0.67741	2.00454	0.58289	1.51395	0.75526
9	0.37500	0.83333	0.71428	0.54545	0.59210	0.71428	3.00000	0.65208	2.25000	0.75000
10	0.51351	0.85294	0.79166	0.61702	0.67605	0.79166	6.12222	0.68521	3.49189	0.57036
11	0.35714	0.75862	0.68181	0.44897	0.52112	0.68181	1.74603	0.57963	1.47959	0.84740
12	0.46511	0.73333	0.71428	0.48888	0.57534	0.71428	2.39130	0.63178	1.74418	0.72938
13	0.42857	0.91428	0.85714	0.57142	0.64935	0.85714	8.00000	0.66190	5.0000	0.62500
14	0.37209	0.65384	0.64000	0.38636	0.47826	0.64000	1.11934	0.53041	1.07493	0.96032
15	0.42105	0.72727	0.64000	0.52173	0.56338	0.64000	1.93939	0.60127	1.54385	0.79605
16	0.41666	0.54545	0.50000	0.46153	0.47826	0.50000	0.85714	0.56902	0.91666	1.06944
17	0.55813	0.78787	0.77419	0.57777	0.65789	0.77419	4.69172	0.66666	2.63122	0.56082
18	0.42857	0.70000	0.66666	0.46666	0.54166	0.66666	1.75000	0.65476	1.42857	0.81632
19	0.57894	0.61290	0.64705	0.54285	0.59420	0.64705	2.17708	0.65449	1.49561	0.68698
20	0.51063	0.75000	0.75000	0.51063	0.60759	0.75000	3.13043	0.62433	2.04255	0.65248

Tabla 4. Muestra las predicciones realizadas con el modelo en cada uno de los 20 remuestreos realizados. Se (sensibilidad), Sp (especificidad), VPP (valor predictivo positivo), VPN (valor predictivo negativo), A (acurracy), P (precisión), OR adj (razón de momios ajustada), AUC (área bajo la curva), RV+ (razón de verosimilitud positiva) y RV- (razón de verosimilitud negativa)

Realizamos un promedio de las 20 repeticiones en cada una de las columnas de las variables con lo que obtenemos las siguientes métricas globales:

Métricas	Resultado global	Desempeño
Sensibilidad (Se)	0.4427	44.27%
Especificidad (Sp)	0.7534	75.34%
Valor predictivo positivo (VPP)	0.7135	71.35%
Valor predictivo negativo (VPN)	0.5032	50.32%
Acuracy (A)	0.5759	57.59%
Precisión (P)	0.7135	71.35%
Odds ratio ajustado (OR adj)	3.1037	3.10
Área bajo la curva (AUC)	0.6335	0.63
Razón de verosimilitud positiva (RV+)	2.1400	2.14
Razón de verosimilitud negativa (RV-)	0.7503	0.75

Tabla 5. Muestra los promedios de cada una de las métricas para estimar una medición global de la prueba diagnóstica. Se (sensibilidad), Sp (especificidad), VPP (valor predictivo positivo), VPN (valor predictivo negativo), A (acuracy), P (precisión), OR adj (razón de momios ajustada), AUC (área bajo la curva), RV+ (razón de verosimilitud positiva) y RV- (razón de verosimilitud negativa)

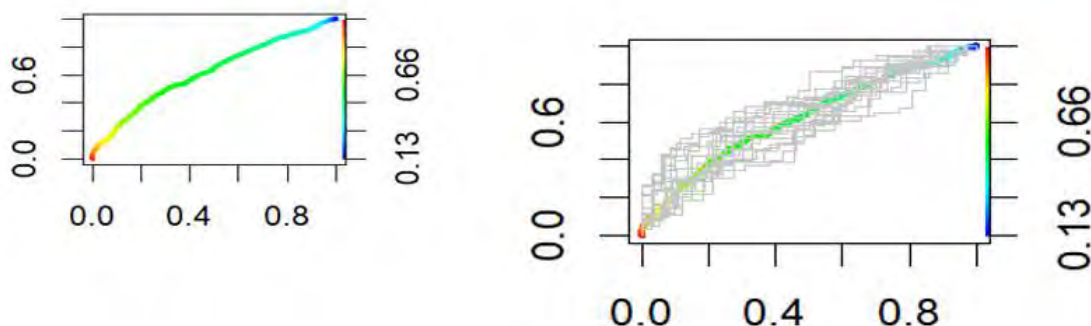


Ilustración 4. Curva ROC del modelo. La gráfica de la izquierda muestra la curva con los desempeños globales. La gráfica de la derecha además de la línea previa, muestra el resto de las curvas creadas con los 20 remuestreos.

Otros modelos alternativos

Durante el desarrollo del modelo nos dimos a la tarea de averiguar otras formas de procesar los datos y realizar diferentes modelos buscando si alguno pudiera representar una mejor herramienta de uso por representar mayor precisión diagnóstica y/o mayor facilidad de aplicación. Al realizar esta investigación surgieron 2 propuestas de modelos:

Modelo secundario: averiguamos como sería el comportamiento del desempeño del modelo tomando en cuenta las 18 variables. Con la función, `glm()` se colocó la fórmula con las 18 variables mostrando los siguientes valores de cada una de las variables:

```
glm(LITO_COLANGIO~EDAD+SEXO+IMC+DM2+HTA+TAB+ALCOHOL+DX_PRE+RIESGO_ASGE+FA+BT+AST_TGO+ALT_TGP+GGT+CO+LITO+PANCREATITIS+COLANGITIS, data=DATOS, family="binomial")
```

De la misma manera que el 1er modelo, se realizó entrenamiento con la función `glm()` y la medición del desempeño con la función `predict.glm()`

El desempeño de este modelo secundario se muestra en la siguiente tabla:

Bootstrap	Se	Sp	VPP	VPN	A	P	OR adj	AUC	RV+	RV-
1	0.54285	0.66666	0.63333	0.57894	0.60294	0.63333	2.37500	0.68398	1.6285	0.68571
2	0.50000	0.62068	0.64516	0.47368	0.55072	0.64516	1.63636	0.57758	1.3181	0.80555
3	0.43902	0.45454	0.60000	0.30303	0.44444	0.60000	0.65217	0.51108	0.8048	1.23414
4	0.48780	0.65517	0.66666	0.47500	0.55714	0.66666	1.80952	0.59629	1.4146	0.78177
5	0.53488	0.62962	0.69696	0.45945	0.57142	0.69696	1.95500	0.58311	1.4441	0.73871
6	0.47368	0.63333	0.62068	0.48717	0.54411	0.62068	1.55454	0.57982	1.2918	0.83102
7	0.35714	0.72413	0.65217	0.43750	0.50704	0.65217	1.45833	0.56568	1.2946	0.88775
8	0.46511	0.70967	0.68965	0.48888	0.56756	0.68965	2.12560	0.65041	1.6020	0.75369
9	0.42500	0.69444	0.60714	0.52083	0.55263	0.60714	1.67984	0.64166	1.3909	0.82800
10	0.70270	0.67647	0.70270	0.67647	0.69014	0.70270	4.94214	0.69316	2.1719	0.43948
11	0.54761	0.62068	0.67647	0.48648	0.57746	0.67647	1.98086	0.61576	1.4437	0.72883
12	0.39534	0.63333	0.60714	0.42222	0.49315	0.60714	1.12937	0.53798	1.0782	0.95471
13	0.42857	0.77142	0.69230	0.52941	0.58441	0.69230	2.53125	0.52176	1.8750	0.74074
14	0.55813	0.53846	0.66666	0.42424	0.55072	0.66666	1.47368	0.58407	1.2093	0.82059
15	0.52631	0.60606	0.60606	0.52631	0.56338	0.60606	1.70940	0.57496	1.3360	0.78157

16

0.58333	0.51515	0.56756	0.53125	0.55072	0.56756	1.48750	0.56902	1.2031	0.80882
0.67441	0.45454	0.61702	0.51724	0.57894	0.61702	1.72619	0.56941	1.2364	0.71627
0.42857	0.73333	0.69230	0.47826	0.55555	0.69230	2.06250	0.69761	1.6071	0.77922
0.65789	0.61290	0.67567	0.59375	0.63768	0.67567	3.04487	0.63242	1.6995	0.55817
0.53191	0.71875	0.73529	0.51111	0.60759	0.73529	2.90404	0.62965	1.8912	0.65124

17

18

19

20

Tabla 6. Muestra las predicciones realizadas con el modelo secundario que incluye las 18 variables. Se muestran en cada fila cada uno de los 20 remuestreos realizados. Se (sensibilidad), Sp (especificidad), VPP (valor predictivo positivo), VPN (valor predictivo negativo, A (acurracy), P (precisión), OR adj (razón de momios ajustada), AUC (área bajo la curva), RV+ (razón de verosimilitud positiva) y RV- (razón de verosimilitud negativa)

Realizamos un promedio de las 20 repeticiones en cada una de las columnas de las variables con lo que obtenemos las siguientes métricas globales:

Métricas	Resultado global	Desempeño
Sensibilidad (Se)	0.5130	51.30%
Especificidad (Sp)	0.6334	63.34%
Valor predictivo positivo (VPP)	0.6525	65.25%
Valor predictivo negativo (VPN)	0.4960	49.60%
Acuracy (A)	0.5643	56.43%
Precisión (P)	0.6525	65.25%
Odds ratio ajustado (OR adj)	2.0119	2.01
Área bajo la curva (AUC)	0.6007	0.60
Razón de verosimilitud positiva (RV+)	1.4470	1.44
Razón de verosimilitud negativa (RV-)	0.7763	0.77

Tabla 7. Muestra los promedios de cada una de las métricas para estimar una medición global del modelo con 18 variables. Se (sensibilidad), Sp (especificidad), VPP (valor predictivo positivo), VPN (valor predictivo negativo, A (acurracy), P (precisión), OR adj (razón de momios ajustada), AUC (área bajo la curva), RV+ (razón de verosimilitud positiva) y RV- (razón de verosimilitud negativa)

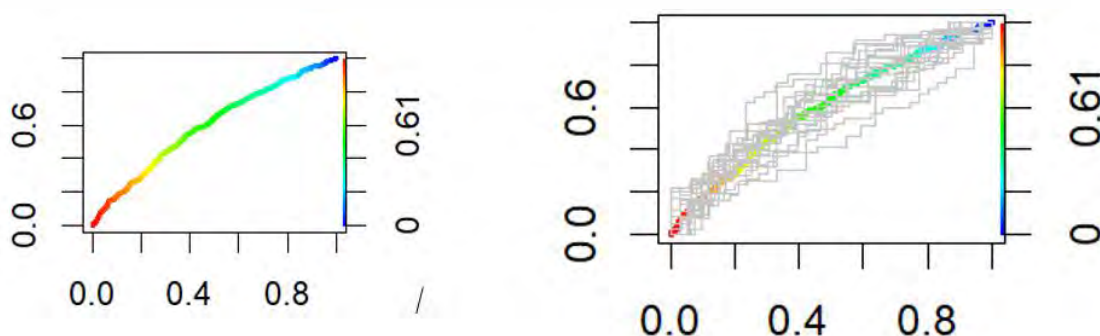


Ilustración 5. Curva ROC del modelo del modelo con 18 variables. La gráfica de la izquierda muestra la curva con los desempeños globales. La gráfica de la derecha además de la línea previa, muestra el resto de las curvas creadas con los 20 remuestreos.

Modelo terciario: para el diseño de este modelo colocamos las 18 variables a la función `glm()` y posteriormente aplicamos los criterios de Akaike (AIC) con lo cual el software hace la elección del mejor modelo que resulta con 5 variables como se muestra en la siguiente fórmula:

```
glm(formula = LITO_COLANGIO ~ DM2 + HTA + DX_PRE + GGT + LITO, family = "binomial", data = DATOS)
```

Igualmente, se realizó entrenamiento con la función `glm()` y la medición del desempeño con la función `predict.glm()`

El desempeño de este modelo terciario se muestra en la siguiente tabla:

Bootstrap	Se	Sp	VPP	VPN	A	P	OR adj	AUC	RV+	RV-
1	0.51428	0.75757	0.69230	0.59523	0.63235	0.69230	3.30882	0.73333	2.12142	0.64114
2	0.62500	0.75862	0.78125	0.59459	0.68115	0.78125	5.23809	0.71810	2.58928	0.49431
3	0.51219	0.63636	0.72413	0.41176	0.55555	0.72413	1.83750	0.65299	1.40853	0.76655
4	0.60975	0.79310	0.80645	0.58974	0.68571	0.80645	5.98958	0.79226	2.94715	0.49204
5	0.58139	0.85185	0.86206	0.56097	0.68571	0.86206	7.98611	0.80361	3.92441	0.49140
6	0.55263	0.86666	0.84000	0.60465	0.69117	0.84000	8.02941	0.83508	4.14473	0.51619
7	0.38095	0.75862	0.69565	0.45833	0.53521	0.69565	1.93406	0.66009	1.57823	0.81601
8	0.53488	0.74193	0.74193	0.53488	0.62162	0.74193	3.30625	0.69692	2.07267	0.62689
9	0.42500	0.83333	0.73913	0.56603	0.61842	0.73913	3.69565	0.74305	2.55000	0.69000
10	0.56756	0.85294	0.80769	0.64444	0.70422	0.80769	7.61250	0.77424	3.85945	0.50698
11	0.57142	0.68965	0.72727	0.52631	0.61971	0.72727	2.96296	0.69211	1.84127	0.62142
12	0.60465	0.83333	0.83870	0.59523	0.69863	0.83870	7.64705	0.79457	3.62790	0.47441
13	0.50000	0.80000	0.75000	0.57142	0.63636	0.75000	4.00000	0.72857	2.50000	0.62500
14	0.51162	0.65384	0.70967	0.44736	0.56521	0.70967	1.97883	0.61449	1.47803	0.74692
15	0.50000	0.81818	0.76000	0.58695	0.64788	0.76000	4.50000	0.71610	2.75000	0.61111
16	0.55555	0.63636	0.62500	0.56756	0.59420	0.62500	2.18750	0.64309	1.52777	0.69841
17	0.72093	0.72727	0.77500	0.66666	0.72368	0.77500	6.88888	0.76250	2.64341	0.38372
18	0.47619	0.83333	0.80000	0.53191	0.62500	0.80000	4.54545	0.72539	2.85714	0.62857
19	0.57894	0.77419	0.75862	0.60000	0.66666	0.75862	4.71428	0.75466	2.56391	0.54385
20	0.55319	0.71875	0.74285	0.52272	0.62025	0.74285	3.16402	0.69747	1.96690	0.62164

Tabla 8. Muestra las predicciones realizadas con el modelo terciario que incluye 5 variables. Se muestran en cada fila cada uno de los 20 remuestreos realizados. Se (sensibilidad), Sp (especificidad), VPP (valor predictivo positivo), VPN (valor predictivo negativo), A (acuracy), P (precisión), OR adj (razón de momios ajustada), AUC (área bajo la curva), RV+ (razón de verosimilitud positiva) y RV- (razón de verosimilitud negativa)

Realizamos un promedio de las 20 repeticiones en cada una de las columnas de las

variables con lo que obtenemos las siguientes métricas globales:

Métricas	Resultado global	Desempeño
Sensibilidad (Se)	0.5438	54.38%
Especificidad (Sp)	0.7667	76.67%
Valor predictivo positivo (VPP)	0.7588	75.88%
Valor predictivo negativo (VPN)	0.5588	55.88%
Acuracy (A)	0.6404	64.04%
Precisión (P)	0.7588	75.88%
Odds ratio ajustado (OR adj)	4.5763	4.57
Área bajo la curva (AUC)	0.7269	0.72
Razón de verosimilitud positiva (RV+)	2.5476	2.54
Razón de verosimilitud negativa (RV-)	0.5998	0.59

Tabla 9. Muestra los promedios de cada una de las métricas para estimar una medición global del modelo con 5 variables. Se (sensibilidad), Sp (especificidad), VPP (valor predictivo positivo), VPN (valor predictivo negativo), A (acuracy), P (precisión), OR adj (razón de momios ajustada), AUC (área bajo la curva), RV+ (razón de verosimilitud positiva) y RV- (razón de verosimilitud negativa)

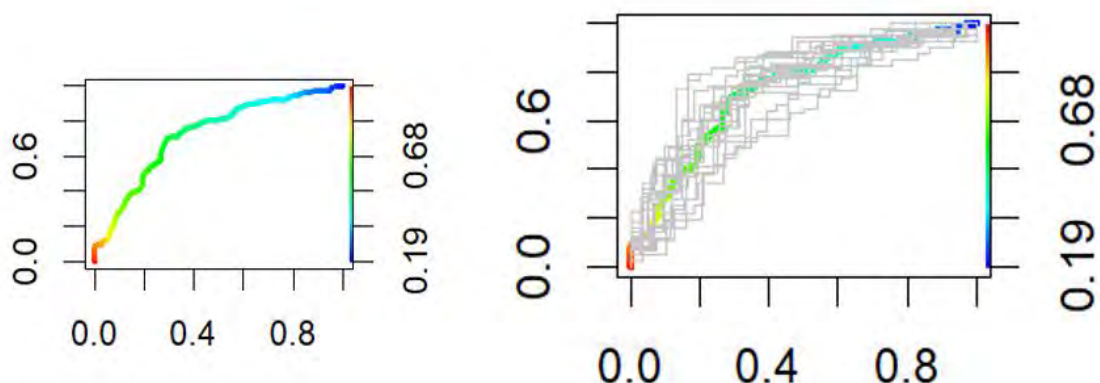


Ilustración 6. Curva ROC del modelo del modelo con 5 variables. La gráfica de la izquierda muestra la curva con los desempeños globales. La gráfica de la derecha además de la línea previa, muestra el resto de las curvas creadas con los 20 remuestreos.

La importancia de los umbrales en las pruebas diagnósticas

Para la obtención del desempeño global del modelo optamos por usar un umbral de 0.60. En learn machine el umbral va desde 0 hasta 1 y la elección de la cifra puede variar dependiendo su aplicación clínica. Nos dimos a la tarea de realizar predicciones con varios umbrales para observar que variabilidad se presentaba en

las métricas de rendimiento con respecto a los modelos secundario y terciario:

Con un umbral de 0.21 observamos que se observa un beneficio en el valor predictivo y negativo de la prueba, haciendo este ajuste el valor predictivo positivo pico arrojado es de 0.7407, con un valor promedio de 0.5877

El valor predictivo negativo pico registrado es de 0.75 y un valor promedio de 0.5634.

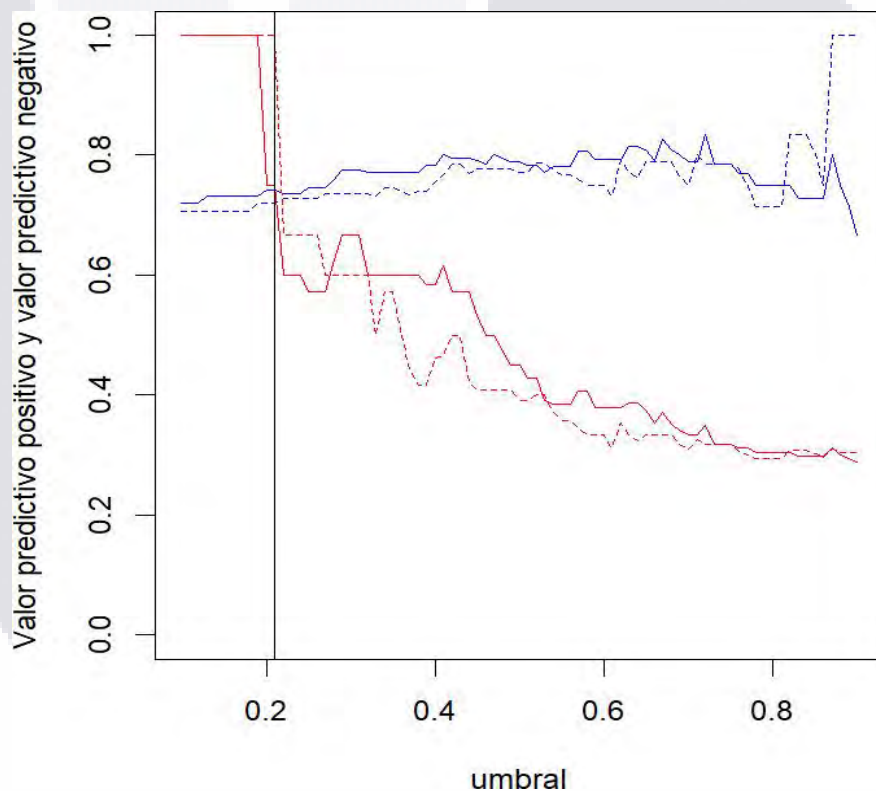


Ilustración 7. Las líneas azules representan el valor predictivo positivo y las líneas rojas representan el valor predictivo negativo. La línea continua evalúa el modelo con 18 variables y las líneas discontinuas el modelo con 5 variables. Obsérvese como las curvas siguen una dirección inversamente proporcional, pero en el umbral 0.2 (marcado con la línea vertical en negro) hay un punto de intersección entre las 2, lo que representa que con este valor de umbral obtendríamos el mayor poder predictivo tanto negativo como positivo para la prueba.

En cambio, si realizamos un ajuste del umbral a 0.60, observamos una optimización

en el valor de la especificidad motivo por el cual escogimos este valor para el desempeño del modelo. Como se muestra en la gráfica. El punto de intercepción de ambos valores corresponde a 0.55, sin embargo, deseamos optimizar el modelo aumentando su especificidad.

En este modelo, se muestra una sensibilidad pico de 0.5121 (51.21%) y una especificidad pico de 0.5882 (58.88%). Esto con respecto al modelo terciario.

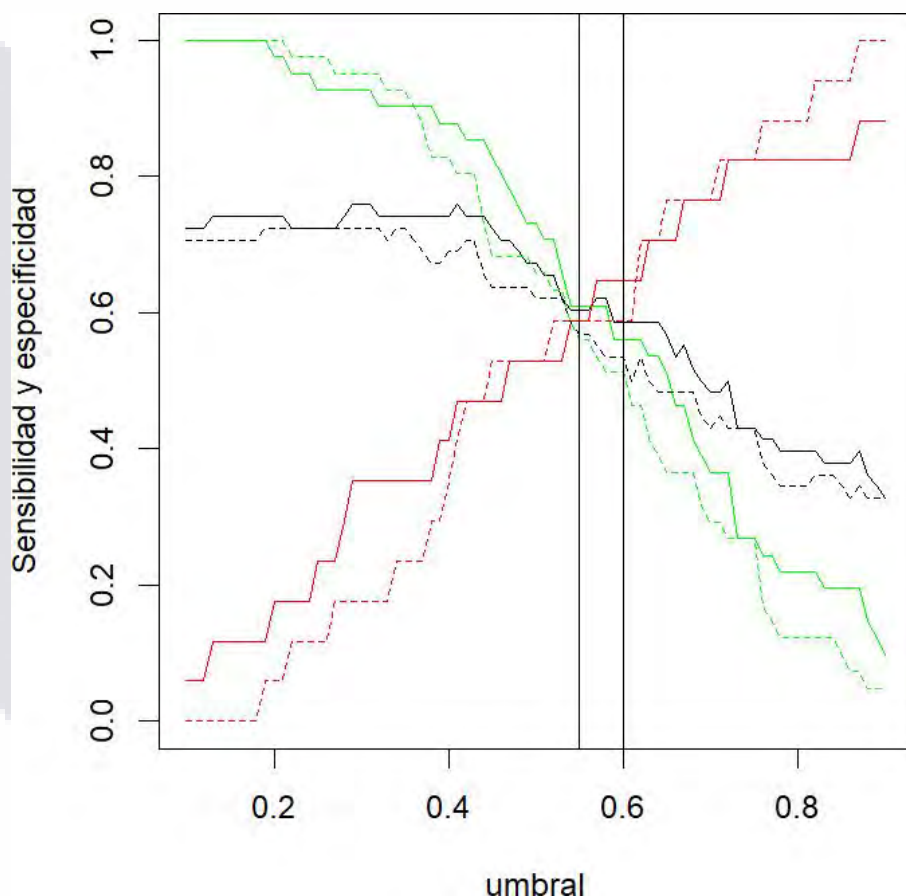


Ilustración 8. Las líneas verdes representan la sensibilidad, las líneas rojas representan la especificidad y la línea negra representa la acuracy. Las líneas continuas corresponden al modelo con 18 variables y las líneas discontinuas al modelo de 5 variables. Nótese la relación inversamente proporcional de la especificidad y sensibilidad. El punto de intersección de todas las líneas es en un umbral de 0.55. Sin embargo, al desear aumentar el desempeño para algún fin específico de la prueba se moviliza el punto de corta a 0.60 (2da línea negra vertical) donde se ve favorecida la especificidad de la prueba diagnóstica.

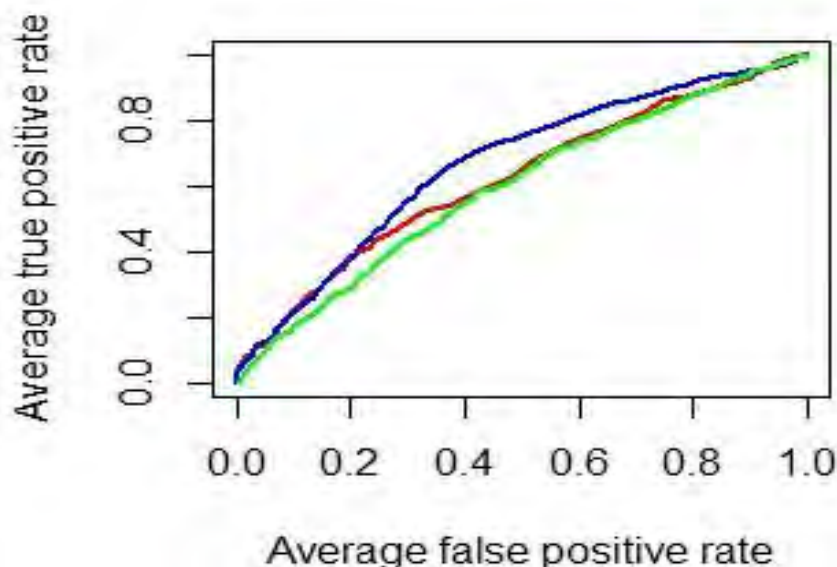


Ilustración 9. Cada una de las líneas representa la curva ROC de cada uno de los modelos creados. La línea azul representa el modelo secundario con un AUC= 0.7269, la curva verde representa el modelo primario con un AUC = 0.6335 y la curva roja el modelo terciario con un AUC = 0.6007.

Desempeño de los criterios ASGE 2019

Se realizó una matriz de confusión con los pacientes de la muestra que tuvieron coledocolitiasis respecto a los criterios ASGE 2019 contra los que se les corroboró este diagnóstico. Se clasificaron como positivos a coledocolitiasis solo los pacientes con riesgo alto, ya que a estos pacientes implicaría referirlos inmediatamente a CPRE.

CRITERIOS ASGE 2019	ENFERMEDAD PRESENTE	ENFERMEDAD AUSENTE
COLEDOCOLITIASIS (RIESGO ALTO)	a = 68	b= 36
NO COLEDOCOLITIASIS (RIESGO MODERADO)	c=41	d= 49

Tabla 10. Matriz de confusión que muestra los resultados obtenidos en la muestra, el total de pacientes clasificados con riesgo alto de coledocolitiasis fueron la suma de a + b (104) y con riesgo moderado c + d (90), sin embargo, solamente a 68 pacientes y 41 paciente de cada categoría se les corroboró el diagnóstico.

Mediante el ordenamiento de esta información se obtienen las siguientes métricas:

Métricas	Resultado global	Desempeño
Sensibilidad (Se)	0.6238	62.38%
Especificidad (Sp)	0.5764	57.64%
Valor predictivo positivo (VPP)	0.6538	65.38%
Valor predictivo negativo (VPN)	0.5444	54.44%
Acuracy (A)	0.6030	60.30%
Precisión (P)	0.6538	65.38%
Odds ratio ajustado (OR adj)	2.3219	2.32
Área bajo la curva (AUC)	0.5993	0.599
Razón de verosimilitud positiva (RV+)	1.47	1.47
Razón de verosimilitud negativa (RV-)	0.655	0.65

Tabla 11. Muestra el desempeño de los criterios ASGE 2019 en la muestra estudiada. Se (sensibilidad), Sp (especificidad), VPP (valor predictivo positivo), VPN (valor predictivo negativo), A (accuracy), P (precisión), OR adj (razón de momios ajustada), AUC (área bajo la curva), RV+ (razón de verosimilitud positiva) y RV- (razón de verosimilitud negativa)

Discusión

Los 194 pacientes que completan la muestra de este estudio y que fueron sometidos a CPRE y/o CIO, solo el 56.18% de ellos se les pudo corroborar la presencia de coledocolitiasis. Esto representa una prevalencia de la enfermedad (probabilidad pre- prueba) de 56.18 %.

El sexo femenino representó la mayoría de la población con coledocolitiasis (68.80%) Esta proporción es acorde con lo reportado en la literatura mexicana de 88.4%, reforzando el acrónimo tan mencionado de las 4 F, siendo mujer (female) uno de los factores predictores importantes. El promedio de edad de presentación es en la 5ta década de la vida. Específicamente en las mujeres la edad promedio fue de 49.8 años y en hombres de 59.96 años. Nuevamente observamos la similitud con el acrónimo 4 F donde otra de las F es edad de 40 años (Forty).

En contraste a lo que crearíamos, encontramos que el índice de masa corporal promedio en los pacientes con coledocolitiasis fue de 27.69 kg/m² (sobrepeso), además el IMC representó una correlación significativamente estadística ($p < 0.04097$), pero el promedio de presentación no llega a los niveles de obesidad. Esto lleva a reflexionar sobre la asociación que hemos realizado en el campo médico al intuir que entre más sea el grado de sobrepeso u obesidad están correlacionados proporcionalmente con la presencia de litiasis, quizá lo esté, más

TESIS TESIS TESIS TESIS TESIS

esto no determina que la litiasis presente concomitantemente coledocolitiasis o lo predisponga. Otro factor que pudo haber influido en esta correlación fue la captura de datos, ya que se observó que pocos médicos realizaban fielmente la captura del peso y talla del paciente, medidas necesarias para el cálculo de este valor. Esto también pudo haber influido en los resultados obtenidos. Habría que hacer un estudio con mayor detalle para reforzar esta teoría.

La diabetes fue observada en un 22.01% y la hipertensión en el 43.11%. No presencia no determino una asociación positiva con la presencia de litos intraductales. El tener hábito tabáquico y alcoholismo tampoco se relaciona con la presencia de coledocolitiasis.

Los parámetros bioquímicos tampoco representaron una correlación con coledocolitiasis, excepto la GGT, que mostró una fuerte correlación con la presencia de litiasis intraductal ($p < 0.011$).

Durante la construcción del modelo nos percatamos como el IMC muestra un OR = 5.55, sin embargo, la probabilidad logit es negativa, lo que indica una disminución en la probabilidad de obtener un éxito. Evidentemente el IMC tiene una significancia estadística respecto a presentación de litos intraductales (valor $p < 0.035$) pero con una asociación negativa ($B = -0.05316$).

Al hacer la agregación de la variable riesgo ASGE, presenta un coeficiente B positivo, lo que denota una relación significativa, junto con un OR = 6.24, denota una fuerte probabilidad de contribución al diagnóstico de coledocolitiasis ($B = 0.74205$), esta fue la variable que presentó el coeficiente B más alto. Lo cual era esperado ya que con base a este criterio de tomó la decisión de inclusión en el estudio.

La fosfatasa alcalina (FA) presentó un OR = 3.18 y un coeficiente Beta positivo ($B = 0.0011850$), pero un valor $P < 0.072$ en conjunto con el resto de las variables. Por este motivo no explica un valor estadísticamente significativo. Sin embargo, el poder de asociación de riesgo ASGE se mantiene fuertemente positivo y el coeficiente logit de IMC negativo.

La gama glutamiltransferasa (GGT) presentó el OR más bajo (OR = 3.184) un coeficiente beta positivo ($B = 0.0008773$) y un valor P con poder explicativo para coledocolitiasis. ($p < 0.044$)

La visualización de lito intraductal presentó un coeficiente B positivo ($B = 0.5974993$) pero una p sin significancia estadística. ($p < 0.1674$)

Con el análisis de regresión logística multiple encontramos que la visualización de lito dentro del colédoco y la fosfatasa alcalina no tiene significancia estadística con

la presentación de coledocolitiasis. Por otro lado, el riesgo ASGE calculado y la GGT tiene un poder estadísticamente explicativo para el diagnóstico, mientras que el IMC también lo tiene, pero como un valor predictor negativo.

Durante la medición del desempeño del modelo, logramos una mayor especificidad del modelo ($Sp=0.7534, 75.34\%$), comparándolo contra los criterios ASGE 2019 ($Sp=0.5764, 57.64\%$), con una sensibilidad menor ($Se=0.4427, 44.27\%$ vs $Se=0.6238, 62.38\%$), el resto de los parámetros de desempeño también fueron superiores comparado con los criterios ASGE: valor predictivo positivo (71.35% vs 65.38%), la precisión ($0.7135, 71.35\%$ vs $0.6538, 65.38\%$), el área bajo la curva ($AUC=0.6335$ vs $AUC=0.5993$), la razón de verosimilitud positiva ($RV+ = 2.14$ vs $RV+=1.47$) y la razón de verosimilitud negativa ($RV- =0.75$ vs 0.65).

La sensibilidad, el valor predictivo negativo ($VPN = 54.44\%$) y accuracy ($A= 60.3\%$) fueron mayores con los criterios ASGE.

El área bajo la curva es mayor comparado con la obtenida con los criterios ASGE 2019 (0.63 vs 0.59) esto le confiere una superioridad diagnóstica a nuestro modelo.

La mayor especificidad encontrada en este modelo proporciona mayor utilidad para detectar a los pacientes con coledocolitiasis y evitar someter a los pacientes con falsos positivos a los riesgos que conlleva la realización de CPRE y sus posibles complicaciones. También reduce el costo hospitalario si se aplica el modelo inicialmente podemos aumentar la probabilidad del diagnóstico hasta en un 71.35% , lo cual nos parece adecuado, incluso se podría optar la opción de revalorar el manejo intrahospitalario en vez de permanecer hospitalizados hasta que la colangiografía pueda ser realizada, lo que condiciona una espera de mínimo 72 h.

La accuracy fue ligeramente menor comparado con los criterios ASGE (57.59% vs 60.30%), pero la precisión fue mayor (71.35% vs 65.38%), aterrizando en el contexto de que un falso positivo es una decisión que se desea evitar al realizar el diagnóstico favorece la precisión mostrada contra la accuracy, ya que nos va a permitir conocer de los pacientes que tienen diagnóstico de coledocolitiasis, cuantos portan este diagnóstico correctamente.

La sensibilidad y especificidad no son suficientes para dar valor diagnóstico a nuestra prueba ya que es solo el paso inicial para hacer el diagnóstico y poder orientarnos sobre la probabilidad de padecer la enfermedad. Las métricas que mayor valor aportan para decidirnos sobre esta prueba es el valor predictivo positivo, en nuestro modelo este fue de 71.35% y la razón de verosimilitud positiva que fue de 2.14. Ambos nos aportarán que tan probable es que el paciente con un resultado positivo (en este caso siendo los criterios ASGE 2019) certeramente

TESIS TESIS TESIS TESIS TESIS

padezca la enfermedad.

Según la literatura la razón de verosimilitud positiva entre 2.1 a 5.0 proporciona resultados precisos, nuestro modelo arrojó una RV + de 2.14. Lo que la coloca en una opción confiable para apoyar el diagnostico de coledocolitiasis. (33)

Respecto al modelo alternos otros modelos encontrados durante el diseño observamos que el modelo con 5 variables mostró la mayor cantidad de medidas de rendimiento superiores comparado con riesgo ASGE y el modelo primario únicamente siendo inferior en la sensibilidad (54.38 % vs criterios ASGE 62.38%) y en la razón de verosimilitud negativa. (0.59 vs criterios ASGE 0.65).

Al comparar las 3 áreas bajo la curva de los 3 modelos observamos una mayor eficacia diagnostica con el modelo secundario con un AUC= 0.7269, seguido de nuestro modelo con una AUC = 0.6335. Los criterios ASGE mostraron un AUC Inferior a todos los modelos creados. (AUC= 0.599).

Esto es un hallazgo interesante y podría ser un parteaguas para próximos estudios.

Conclusiones

El modelo predictivo proporciona una herramienta precisa, y sin costo en su aplicación, con los datos obtenidos no consideramos que sea suficiente para sustituir los criterios ASGE 2019, pero si es una herramienta de apoyo que aporta un valor predictivo positivo, una razón de verosimilitud positiva y un área bajo la curva considerablemente a favor para la toma de decisión terapéutica del paciente.

Debería tomarse en cuenta el valor de la GGT que tiene correlación significativa con coledocolitiasis, así como evitar pensar que por más grado de obesidad presente el paciente tendrá mayor probabilidad de presentar coledocolitiasis. Así mismo se presentaron diferencias significativas de estas variables respecto al sexo, lo que pudiera indicar que los criterios diagnostico podría realizarse dependiendo el sexo del paciente.

El modelo alternativo con 5 variables obtenido por Learn Machine mostró una superioridad frente a al resto de los modelos y respecto a los criterios ASGE 2019, se requiere más estudios de prueba de desempeño para determinar si es una herramienta diagnóstica superior a los criterios.

El Learn Machine nos ayuda a crear herramientas que tienden a la individualidad del paciente. Esto es beneficioso en el área médica ya que aumentará la probabilidad de certeza en los diagnósticos.

Glosario

Predicción: predicción o proyección de la naturaleza de problemas futuros o condiciones existentes basadas en la extrapolación o interpretación de datos científicos existentes o por aplicación de metodología científica. (15)

Coledocolitiasis: “La presencia de un cálculo o cálculos dentro del conducto biliar común (CDB) se conoce como coledocolitiasis.” (16) “Los cálculos en el conducto biliar (coledocolitiasis) son el resultado más frecuente de la migración de cálculos biliares desde la vesícula biliar hacia el árbol biliar.” (6)

ASGE: (por sus siglas en inglés) Sociedad Americana Endoscopia Gastrointestinal.

ESGE: (por sus siglas en inglés) Sociedad Europea de Endoscopia Gastrointestinal.

Análisis multivariable: “se define como el análisis estadístico que incluye y analiza tres o más variables simultáneamente. Los análisis multivariados más habituales son los modelos de regresión (múltiple, logística, Cox y Poisson) La regresión logística la variable dependiente no tiene un sentido numérico en sí misma, sino que es el logaritmo neperiano ($\ln$) de la probabilidad (p) de que ocurra un suceso, dividido por la probabilidad de que no ocurra ($1-p$).” (17)

Modelo predictivo “estima el valor de un atributo de destino determinado y existen datos de entrenamiento de muestra para los que se conocen los valores de ese atributo. Otra técnica empleada en el modelado predictivo es el análisis de regresión, que se puede utilizar cuando el atributo objetivo es un valor numérico y el objetivo es predecir ese valor para nuevos datos.” (18)

Colangiopancreatografía magnética (CRM): “Se basa en la utilización de secuencias altamente potenciadas en T2, en las que la señal de las estructuras llenas de líquido estático (bilis) es muy alta en comparación con las del resto de las estructuras. Pueden adquirirse imágenes en los 3 planos del espacio y obtener representación anatómica del árbol biliar.” (19)

Colangiopancreatografía retrógrada endoscópica (CPRE): “es una técnica mixta, endoscópica y radiológica, introducida por McCune et al en 1968. La técnica nace con un objetivo fundamentalmente diagnóstico, de tal manera que conseguían de manera retrógrada, a través de la papila, rellenar con contraste los conductos biliar y pancreático. La función terapéutica de la CPRE se reconoció y desarrollo después de la introducción de la esfinterotomía en 1972 por Classen y Demling de Alemania y Kawai et al. de Japón.” (20)

Colangiografía intraoperatoria (CIO): “requiere la identificación del conducto cístico en la unión con el cuello de la vesícula biliar durante la cirugía

TESIS TESIS TESIS TESIS TESIS

(colecistectomía). Se realiza una pequeña incisión en el conducto a través de la cual se inserta un catéter y luego se inyecta un contraste para obtener imágenes fluoroscópicas.” (20)

TRIPOD: “es la Declaración para la Transparencia del Reporte de un Modelo Predictivo Multivariable para el Diagnóstico y Pronóstico en Individuos y la lista de verificación correspondiente se desarrollaron con el objetivo de mejorar la transparencia de la notificación de modelos predictivos, evaluar el posible sesgo de los estudios de modelado y evaluar la utilidad de los modelos de predicción clínica antes de su implementación en la práctica clínica habitual.” (21)



TESIS TESIS TESIS TESIS TESIS

Referencias bibliográficas

1. Gonzales Pérez Luis Gabriel, Zaldívar Ramírez Felipe Rafael, Tapia Contla Brenda Ruth, Díaz Contreras Piedras Carlos Manuel, Arellano López Paul Ricardo, Hurtado López Luis Mauricio. Factores de riesgo de la coledocolitiasis asintomática; experiencia en el Hospital General de México. 2018. 2018 Sep;40(3):164–8.
2. Lin H, Zhou X, Zhang Z. The Diagnostic Value of GGT-Based Biochemical Indicators for Choledocholithiasis with Negative Imaging Results of Magnetic Resonance Cholangiopancreatography. Teekaraman Y, editor. Contrast Media & Molecular Imaging. 2022 jun 27; 2022:1–10.
3. Marín López José de Jesús, Flores Álvarez Efrén, Gómez Arámbula Ramiro, Cruz de la Torre Gonzáles José, Chávez Fernández Daniel, Rodríguez Osuna José Augusto, et al. Experiencia de cinco años en el manejo de pacientes con alto riesgo de coledocolitiasis. 2020. 2020 Mar;21(1):26.
4. Dalai C, Azizian JM, Trieu H, Rajan A, Chen FC, Dong T, et al. Machine learning models compared to existing criteria for noninvasive prediction of endoscopic retrograde cholangiopancreatography-confirmed choledocholithiasis. Liver Research. 2021 Dec;5(4):224–31.
5. Miguelena Bobadilla José Ma. Cirugía para el grado de medicina. In: Cirugía para el grado de medicina [Internet]. 2nd ed. Madrid, España: Editorial Medica Panamericana; 2019 [cited 2023 Apr 27]. Available from: <https://mieureka-medicapanamericana-com.pbidi.unam.mx:2443/viewer/cirugia-libro-electronico/iv>
6. Buxbaum JL, Abbas Fehmi SM, Sultan S, Fishman DS, Qumseya BJ, Cortessis VK, et al. ASGE guideline on the role of endoscopy in the evaluation and management of choledocholithiasis. Gastrointestinal Endoscopy. 2019 jun;89(6):1075-1105.e15.
7. Silva Aycaguer LC. Excursión a la regresión logística en ciencias de la salud. 2007 [cited 2024 Sep 6]; Available from: <https://elibro.net/es/lc/uaa/titulos/52889>
8. Stoltzfus JC. Logistic Regression: A Brief Primer. Academic Emergency Medicine. 2011;18(10):1099–104.
9. Menezes N, Marson LP, deBeaux AC, Muir IM, Auld CD. Prospective analysis of a scoring system to predict choledocholithiasis. British Journal of Surgery. 2002 Dec 6;87(9):1176–81.
10. Kamath SU, Dharap SB, Kumar V. Scoring system to preoperatively predict choledocholithiasis. Indian J Gastroenterol. 2016 May;35(3):173–8.
11. Khoury T, Kadah A, Mari A, Kalisky I, Katz L, Mahamid M, et al. A validated score predicting common bile duct stone in patients hospitalized with acute calculus cholecystitis: a multi-center retrospective study. Surg Endosc. 2021 Jul;35(7):3709–15.
12. Ovalle-Chao C, Guajardo-Nieto DA, Elizondo-Pereo RA. 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. Revista de Gastroenterología de México (English Edition). 2023 Oct;88(4):322–32.
13. Tunruttanakul S, Verasmith K, Patumanond J, Mingmalairak C. Development of a Predictive Model for Common Bile Duct Stones in Patients With Clinical Suspicion of Choledocholithiasis: A Cohort Study. Gastroenterol Res. 2022 Oct;15(5):240–52.
14. Blum J, Hunn S, Smith J, Chan FY, Turner R. Using artificial intelligence to predict

choledocholithiasis: can machine learning models abate the use of MRCP in patients with biliary dysfunction? ANZ Journal of Surgery. 2024 Mar 25; ans.18950.

15. Búsqueda | Portal Regional de la BVS [Internet]. [cited 2024 Sep 4]. Available from: https://pesquisa.bvsalud.org/portal/decs-locator/?lang=es&mode=&tree_id=I01.320

16. Molvar C, Glaenger B. Choledocholithiasis: Evaluation, Treatment, and Outcomes. Semin intervent Radiol. 2016 oct 31;33(04):268–76.

17. Miguel Ángel Martínez González, Sánchez Villegas Almudena, Toledo Atucha Estefanía, Faulin Fajardo Javier. Bioestadística Amigable. 4a edición. España: Elsevier; 2020. 552 p.

18. Encyclopedia Britannica [Internet]. [cited 2024 Apr 23]. Data-minig-Britannica Digital Learning. Available from: <https://academic-eb-com.pbidi.unam.mx:2443/levels/collegiate/article/data-mining/437561#281963.toc>

19. José Luis del Cura Rodríguez, Salvador Pedraza Gutiérrez, Ángel Gayete Cara, Álex Rovira Cañellas. Radiología Esencial. In: Radiología Esencial. Madrid, España: Editorial Medica Panamericana; 2018. (2).

20. Pascual Parilla Paricio, Eduardo García-Grandero Ximenez, Elena Martín Pérez, Salvador Morales-Conde, Salvador Navarro Soto, Eduardo María Targarona Soler. Cirugía AEC. Manual de la Asociación Española de Cirujanos. In: Cirugía AEC Manual de la Asociación Española de Cirujanos. 3era edición. Madrid, España: Editorial Medica Panamericana; 2022.

21. Patzer RE, Kaji AH, Fong Y. TRIPOD Reporting Guidelines for Diagnostic and Prognostic Studies. JAMA Surg. 2021 jul 1;156(7):675.

22. ASALE R, RAE. «Diccionario de la lengua española» - Edición del Tricentenario. [cited 2024 Apr 27]. edad | Diccionario de la lengua española. Available from: <https://dle.rae.es/edad>

23. López ER, López NLN. EL PESO CORPORAL SALUDABLE: DEFINICIÓN Y CÁLCULO EN DIFERENTES GRUPOS DE EDAD.

24. Davies MJ, Aroda VR, Collins BS, Gabbay RA, Green J, Maruthur NM, et al. Management of Hyperglycemia in Type 2 Diabetes, 2022. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care. 2022 nov 1;45(11):2753–86.

25. Mancia G, Kreutz R, Brunström M, Burnier M, Grassi G, Januszewicz A, et al. 2023 ESH Guidelines for the management of arterial hypertension The Task Force for the management of arterial hypertension of the European Society of Hypertension: Endorsed by the International Society of Hypertension (ISH) and the European Renal Association (ERA). Journal of Hypertension. 2023 Dec;41(12):1874–2071.

26. ASALE R, RAE. «Diccionario de la lengua española» - Edición del Tricentenario. [cited 2024 Apr 27]. tabaquismo | Diccionario de la lengua española. Available from: <https://dle.rae.es/tabaquismo>

27. ASALE R, RAE. «Diccionario de la lengua española» - Edición del Tricentenario. [cited 2024 Apr 27]. alcoholismo | Diccionario de la lengua española. Available from: <https://dle.rae.es/alcoholismo>

28. ASALE R, RAE. «Diccionario de la lengua española» - Edición del Tricentenario. [cited 2024 Apr 29]. diagnóstico, diagnóstica | Diccionario de la lengua española. Available from: <https://dle.rae.es/diagnóstico>

29. Laboratorio clínico [Internet]. [cited 2024 Apr 27]. Available from: [https://aprendeenlinea-manualmoderno-com.pbidi.unam.mx:2443/reader/275437/&returnUrl%3DaHR0cHM6Ly9hcHJlbnRlZW5saW5lYS1tYW51YWxtb2Rlcm5vLWNvbS5wYmlkaS51bmFtLm14OjI0NDMvZXhwbG9yZTtzZWFiY2hUZXh0PWxhYm9yYXRvcmlvJTlwY2xpJUNDJTgxbmljbztwaHJhc2VNYXRjaD0wO3RoZW1ITmFtZT1EZWZhdWx0LVRoZW1lL3Byb2R1Y3QtZGV0YWlscy8yNzU0Mzc%3D?eapub=https%3A%2F%2Freader-apsouth.ipublishcentral.net%2Fd02e391d66cf08e8810703c566b318ea%2Fe369853df766fa44e1ed0ff613f563bd%2F35d5c32d9aa31a989e6422661450235e%2F7f46165474d11ee5836777d85df2cdab%2F&goto=epubcfi\(/6/28!/4/82/1:0\)&productType=ebook&themeName=Default-Theme](https://aprendeenlinea-manualmoderno-com.pbidi.unam.mx:2443/reader/275437/&returnUrl%3DaHR0cHM6Ly9hcHJlbnRlZW5saW5lYS1tYW51YWxtb2Rlcm5vLWNvbS5wYmlkaS51bmFtLm14OjI0NDMvZXhwbG9yZTtzZWFiY2hUZXh0PWxhYm9yYXRvcmlvJTlwY2xpJUNDJTgxbmljbztwaHJhc2VNYXRjaD0wO3RoZW1ITmFtZT1EZWZhdWx0LVRoZW1lL3Byb2R1Y3QtZGV0YWlscy8yNzU0Mzc%3D?eapub=https%3A%2F%2Freader-apsouth.ipublishcentral.net%2Fd02e391d66cf08e8810703c566b318ea%2Fe369853df766fa44e1ed0ff613f563bd%2F35d5c32d9aa31a989e6422661450235e%2F7f46165474d11ee5836777d85df2cdab%2F&goto=epubcfi(/6/28!/4/82/1:0)&productType=ebook&themeName=Default-Theme)
30. Eureka [Internet]. [cited 2024 Apr 27]. Available from: <https://mieureka-medicapanamericana-com.pbidi.unam.mx:2443/viewer/abc-de-la-ecografia-abdominal/54>
31. Banks PA, Bollen TL, Dervenis C, Gooszen HG, Johnson CD, Sarr MG, et al. Classification of acute pancreatitis—2012: revision of the Atlanta classification and definitions by international consensus. *Gut*. 2013 Jan;62(1):102–11.
32. Kiriyama 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 Hepato Biliary Pancreat*. 2018 Jan;25(1):17–30.
33. Fletcher, G. S. (2021). *Epidemiología clínica* (W. Kluwer, Ed.). Wolters Kluwer.

Anexos

Anexo A. Formato de solicitud de expedientes

FORMATO DE SOLICITUD DE EXPEDIENTE
Vale por expediente clínico

Nombre del paciente: _____
NSS: _____
Fecha de solicitud: _____
Hospital en el que se solicita: _____

Firma de quien solicita expediente principal

Firma de investigador

Firma de jefe de departamento de cirugía

A la devolución del expediente deberá ser firmado por quien entrega y recibe dicho documento:
Fecha de entrega: _____

Nombre y firma de quien entrega expediente

Nombre y firma de quien recibe

Objetivos: con este formato de realizará la solicitud de expedientes al área de archivo de cada una de las unidades para la revisión de los datos de los pacientes seleccionados. Deberá contar con la firma del tesista, el investigador adjunto y el jefe del departamento de cirugía correspondiente a cada unidad. Cada uno de los vales quedará en resguardo por archivo hasta la devolución del expediente a su área donde se hará una devolución de este al investigador principal como comprobante de resguardo adecuado del expediente físico.

Lugar de recolección de datos: oficina del departamento de cirugía de cada unidad, biblioteca del hospital y cuartos asignados al trabajo de médicos.

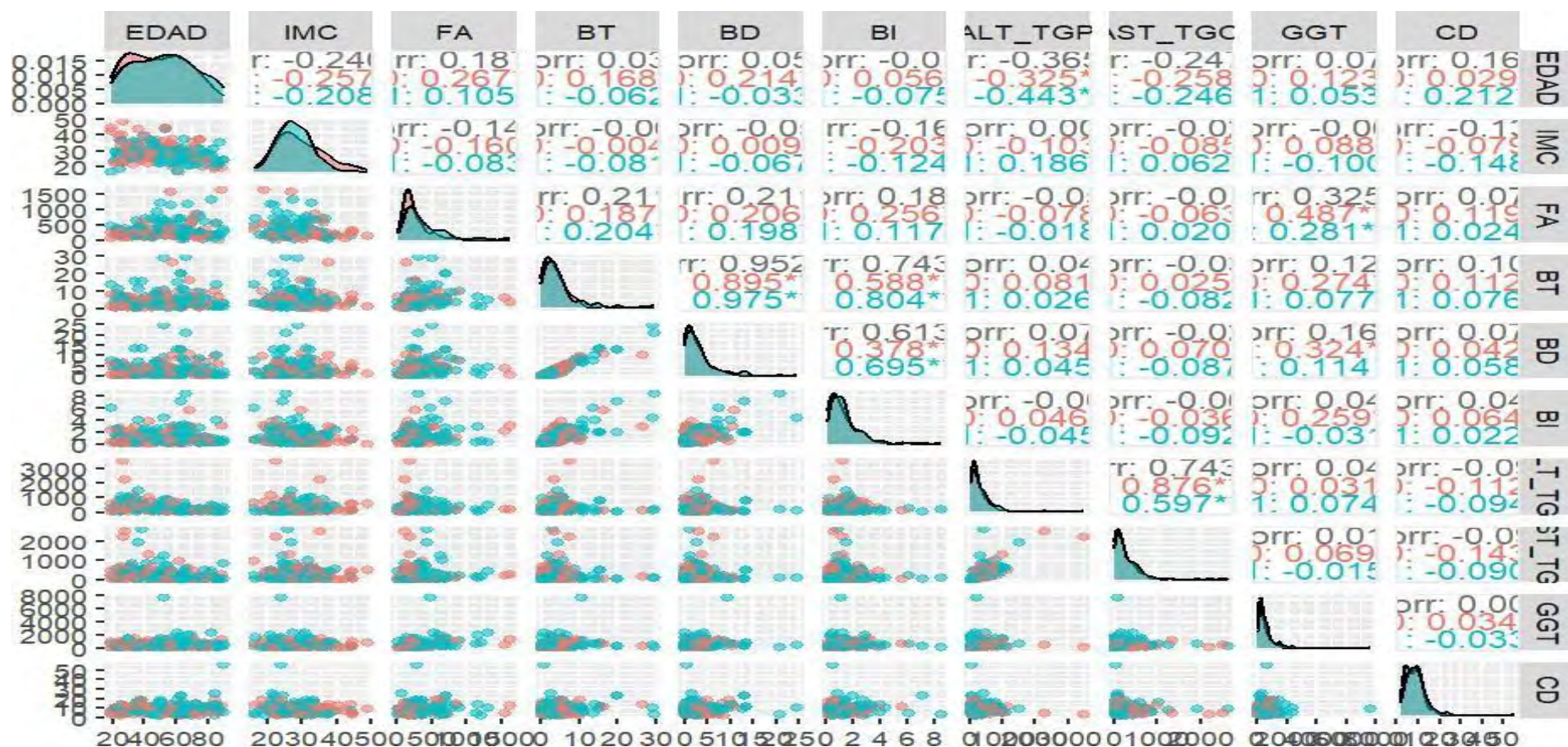
Anexo B. Base de datos para captura de variables

ID	NOMBRE	RPS	DIAGNOSTICO	RIESGO	FECHA DE ESTUDIO	TIPO DE ESTUDIO	RESULTADO CPRE	EDAD	SEXO	IMC	BNC	HTA	TAB	OH	FA	BT	AST	ALT	GGT	TOLEDOOL	LITOS	AMOREATITE	COLANGITIS
1																							
2																							
3																							
4																							
5																							
6																							
7																							
8																							
9																							
10																							
11																							
12																							
13																							
14																							
15																							
16																							
17																							
18																							
19																							
20																							
21																							
22																							
23																							
24																							
25																							
26																							
27																							
28																							
29																							
30																							
31																							
32																							
33																							
34																							
35																							
36																							
37																							
38																							
39																							
40																							
41																							
42																							
43																							
44																							
45																							
46																							

Objetivos: Con esta base de datos realizaremos la recolección de cada una de las variables. La base de datos se encontrará en la nube compartida con los investigadores para que todos cuenten con la misma base de datos actualizada en tiempo real y con respaldo de la información. De manera que cada uno pueda verificar el proceso de captura de datos.

Lugar de recolección de datos: oficina del departamento de cirugía de cada unidad, biblioteca del hospital y cuartos asignados al trabajo de médicos.

Anexo C. Coeficiente de correlación de las variables continuas



C. El anexo muestra la gráfica con las variables cuantitativas que muestran el índice de correlación respecto a la probabilidad de padecer coledocolitiasis. El color azul representa los pacientes que no tienen coledocolitiasis y la curva roja los pacientes con presencia de coledocolitiasis. Entre más sobrepuestas se encuentren ambas curvas significa que las variables no discriminan en gran medida los pacientes positivos de los negativos.

Anexo D. Scripts del software R utilizados

```
#1.IMPORTACION DE BASE DE DATOS.....
MUESTRA <- read.csv(file.choose(),header=TRUE)

#2.FILTRO DE DATOS DE INCLUSIÓN.....
with(MUESTRA,table(EXCLUSION))
DATOS<-subset(MUESTRA,EXCLUSION==0)
table(DATOS$RIESGO_ASGE)
riesgo.bajo <- which(DATOS$RIESGO_ASGE==0)
DATOS <- DATOS[-riesgo.bajo, ]

#3.CALCULO DE PROMEDIO Y DESVIACIONES ESTANDAR DE LAS VARIABLES.....
###EDAD###
DATOS$EDAD <- as.numeric(DATOS$EDAD)
aggregate(EDAD~LITO_COLANGIO,data=DATOS,FUN=mean)
aggregate(EDAD~LITO_COLANGIO+SEXO,data=DATOS,FUN=mean)
aggregate(EDAD~LITO_COLANGIO,data=DATOS,FUN=sd)
aggregate(EDAD~LITO_COLANGIO+SEXO,data=DATOS,FUN=sd)

###IMC###
DATOS$IMC <- as.numeric(DATOS$IMC)
aggregate(IMC~LITO_COLANGIO,data=DATOS,FUN=mean)
aggregate(IMC~LITO_COLANGIO,data=DATOS,FUN=sd)
aggregate(IMC~LITO_COLANGIO+SEXO,data=DATOS,FUN=mean)
aggregate(IMC~LITO_COLANGIO+SEXO,data=DATOS,FUN=sd)

###FA###
DATOS$FA <- as.numeric(DATOS$FA)
aggregate(FA~LITO_COLANGIO,data=DATOS,FUN=mean)
aggregate(FA~LITO_COLANGIO,data=DATOS,FUN=sd)
aggregate(FA~LITO_COLANGIO+SEXO,data=DATOS,FUN=mean)
aggregate(FA~LITO_COLANGIO+SEXO,data=DATOS,FUN=sd)

###BT###
DATOS$BT <- as.numeric(DATOS$BT)
aggregate(BT~LITO_COLANGIO,data=DATOS,FUN=mean)
aggregate(BT~LITO_COLANGIO,data=DATOS,FUN=sd)
aggregate(BT~LITO_COLANGIO+SEXO,data=DATOS,FUN=mean)
aggregate(BT~LITO_COLANGIO+SEXO,data=DATOS,FUN=sd)
aggregate(BD~LITO_COLANGIO,data=DATOS,FUN=mean)

###ALANINOAMINOTRAASFERASA###
DATOS$ALT_TGP <- as.numeric(DATOS$ALT_TGP)
aggregate(ALT_TGP~LITO_COLANGIO,data=DATOS,FUN=mean)
aggregate(ALT_TGP~LITO_COLANGIO,data=DATOS,FUN=sd)
aggregate(ALT_TGP~LITO_COLANGIO+SEXO,data=DATOS,FUN=mean)
aggregate(ALT_TGP~LITO_COLANGIO+SEXO,data=DATOS,FUN=sd)

###ASPARTATOAMINOTRANSFERASA###
DATOS$AST_TGO <- as.numeric(DATOS$AST_TGO)
aggregate(AST_TGO~LITO_COLANGIO,data=DATOS,FUN=mean)
aggregate(AST_TGO~LITO_COLANGIO,data=DATOS,FUN=sd)
aggregate(AST_TGO~LITO_COLANGIO+SEXO,data=DATOS,FUN=mean)
aggregate(AST_TGO~LITO_COLANGIO+SEXO,data=DATOS,FUN=sd)

###GAMAGLUTAMILTRANSFERASA###
DATOS$GGT <- as.numeric(DATOS$GGT)
aggregate(GGT~LITO_COLANGIO,data=DATOS,FUN=mean)
aggregate(GGT~LITO_COLANGIO,data=DATOS,FUN=sd)
aggregate(GGT~LITO_COLANGIO+SEXO,data=DATOS,FUN=mean)
aggregate(GGT~LITO_COLANGIO+SEXO,data=DATOS,FUN=sd)

###TAMAÑO DE COLEDOCO###
DATOS$CD <- as.numeric(DATOS$CD)
aggregate(CD~LITO_COLANGIO,data=DATOS,FUN=mean)
aggregate(CD~LITO_COLANGIO,data=DATOS,FUN=sd)
aggregate(CD~LITO_COLANGIO+SEXO,data=DATOS,FUN=mean)
aggregate(CD~LITO_COLANGIO+SEXO,data=DATOS,FUN=sd)
```

#4.CALCULO DE FRECUENCIAS DE VARIABLES.....

```
###RIESGO_ASGE###
tabla_frecuencias <- table(DATOS$RIESGO_ASGE)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=NULL)*100

###PRESENCIA_DE_LITO###
with(DATOS,table(LITO_COLANGIO))
tabla_frecuencias <- table(DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias, margin=NULL)*100

###PRESENCIA_DE_LITO_RESPECTO AL HOSPITAL###
tabla_frecuencias <- table(DATOS$HOSPITAL, DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias, margin=2)

###SEXO_DE_PACIENTES###
tabla_frecuencias <- table(DATOS$SEXO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=NULL)*100

###SEXO_RESPECTO_A_LA PRESENCIA_DE_LITO###
tabla_frecuencias <- table(DATOS$SEXO, DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=2)
prop.table(tabla_frecuencias,margin=2)*100

###DIABETES_RESPECTO_A_LA PRESENCIA_DE_LITO###
tabla_frecuencias <- table(DATOS$DM, DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=2)
prop.table(tabla_frecuencias,margin=2)*100

###HIPERTENSION_RESPECTO_A_LA PRESENCIA_DE_LITO###
tabla_frecuencias <- table(DATOS$HTA, DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=2)
prop.table(tabla_frecuencias,margin=2)*100

###FUMADORES_RESPECTO_A_LA PRESENCIA_DE_LITO###
tabla_frecuencias <- table(DATOS$TAB, DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=2)
prop.table(tabla_frecuencias,margin=2)*100

###CONSUMO_DE_ALCOHOL_RESPECTO_A_LA PRESENCIA_DE_LITO###
tabla_frecuencias <- table(DATOS$ALCOHOL, DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=2)
prop.table(tabla_frecuencias,margin=2)*100

###DIAGNOSTICOS_PREVIOS_A_COLANGIOGRAFIA###
tabla_frecuencias <-table(DATOS$DX_PRE)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=NULL)*100

###DIAGNOSTICOS_PREVIOS_A_COLANGIOGRAFIA_RESPECTO_A_LA PRESENCIA_DE_LITO###
tabla_frecuencias <- table(DATOS$DX_PRE, DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=2)
prop.table(tabla_frecuencias,margin=2)*100

###RIESGO_ASGE_RESPECTO_A_LA PRESENCIA_DE_LITO###
tabla_frecuencias <- table(DATOS$RIESGO_ASGE, DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=2)
prop.table(tabla_frecuencias,margin=2)*100

###OBSERVACION_DE_LITO INTRALUMINAL_RESPECTO_A_LA PRESENCIA_REAL_DE_LITO###
tabla_frecuencias <- table(DATOS$LITO, DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=2)
prop.table(tabla_frecuencias,margin=2)*100

###PRESENCIA_DE_PANCREATITIS_RESPECTO_A_LA PRESENCIA_DE_LITO###
tabla_frecuencias <- table(DATOS$PANCREATITIS, DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=2)
prop.table(tabla_frecuencias,margin=2)*100

###PRESENCIA_DE_COLANGITIS_RESPECTO_A_LA PRESENCIA_DE_LITO###
tabla_frecuencias <- table(DATOS$COLANGITIS, DATOS$LITO_COLANGIO)
print(tabla_frecuencias)
prop.table(tabla_frecuencias,margin=2)
prop.table(tabla_frecuencias,margin=2)*100
```

```

#5.APLICACION DE PRUEBA SHAPIRO-WILK.....

###VARIABLE EDAD###
DATOS$EDAD <- as.numeric(DATOS$EDAD)
###EDAD Y LITO PRESENTE###
shapiro.test(DATOS$EDAD[DATOS$LITO_COLANGIO==1])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==1])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==1], xlab = "PRESENTE", ylab = "EDAD", main = "EDAD EN COLEDOCOLITIASIS ")

###EDAD Y LITO AUSENTE###
shapiro.test(DATOS$EDAD[DATOS$LITO_COLANGIO==0])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==0])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==0], xlab = "AUSENTE", ylab = "EDAD", main = "EDAD SIN COLEDOCOLITIASIS ")

###EDAD EN MUJERES Y LITO PRESENTE###
shapiro.test(DATOS$EDAD[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0], xlab = "PRESENTE", ylab = "EDAD MUJERES", main = "EDAD DE MUJERES CON COLEDOCOLITIASIS ")

###EDAD EN MUJERES Y LITO AUSENTE###
shapiro.test(DATOS$EDAD[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0], xlab = "AUSENTE", ylab = "EDAD MUJERES", main = "EDAD DE MUJERES SIN COLEDOCOLITIASIS ")

###EDAD EN HOMBRRES Y LITO PRESENTE###
shapiro.test(DATOS$EDAD[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1], xlab = "PRESENTE", ylab = "EDAD HOMBRRES", main = "EDAD DE HOMBRRES CON COLEDOCOLITIASIS ")

###EDAD EN HOMBRRES Y LITO AUSENTE###
shapiro.test(DATOS$EDAD[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])
hist(DATOS$EDAD[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1], xlab = "AUSENTE", ylab = "EDAD HOMBRRES", main = "EDAD DE HOMBRRES SIN COLEDOCOLITIASIS ")

###VARIABLE IMC###
DATOS$IMC <- as.numeric(DATOS$IMC)

###IMC Y LITO PRESENTE###
shapiro.test(DATOS$IMC[DATOS$LITO_COLANGIO==1])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==1])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==1], xlab = "PRESENTE", ylab = "IMC", main = "IMC CON COLEDOCOLITIASIS ")

###IMC Y LITO AUSENTE###
shapiro.test(DATOS$IMC[DATOS$LITO_COLANGIO==0])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==0])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==0], xlab = "AUSENTE", ylab = "IMC", main = "IMC SIN COLEDOCOLITIASIS ")

###IMC EN MUJERES Y LITO PRESENTE###
shapiro.test(DATOS$IMC[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0], xlab = "PRESENTE", ylab = "IMC MUJERES", main = "IMC MUJERES CON COLEDOCOLITIASIS ")

###IMC EN MUJERES Y LITO AUSENTE###
shapiro.test(DATOS$IMC[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0], xlab = "AUSENTE", ylab = "IMC MUJERES", main = "IMC MUJERES SIN COLEDOCOLITIASIS ")

###IMC EN HOMBRRES Y LITO PRESENTE###
shapiro.test(DATOS$IMC[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1], xlab = "PRESENTE", ylab = "IMC HOMBRRES", main = "IMC HOMBRRES CON COLEDOCOLITIASIS ")

###IMC EN HOMBRRES Y LITO AUSENTE###
shapiro.test(DATOS$IMC[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])
hist(DATOS$IMC[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1], xlab = "AUSENTE", ylab = "IMC HOMBRRES", main = "IMC HOMBRRES SIN COLEDOCOLITIASIS ")

```

```
###VARIABLE FOSFATASA ALCALINA###
DATOS$FA <- as.numeric(DATOS$FA)

###FOSFATASA ALCALINA Y LITO PRESENTE###
shapiro.test(DATOS$FA[DATOS$LITO_COLANGIO==1])
hist(DATOS$FA[DATOS$LITO_COLANGIO==1])
hist(DATOS$FA[DATOS$LITO_COLANGIO==1], xlab = "PRESENTE", ylab = "FOSFATASA", main = "FOSFATASA EN COLEDOCOLITIASIS ")

###FOSFATASA ALCALINA Y LITO AUSENTE###
shapiro.test(DATOS$FA[DATOS$LITO_COLANGIO==0])
hist(DATOS$FA[DATOS$LITO_COLANGIO==0])

###FOSFATASA ALCALINA EN MUJERES CON LITO PRESENTE###
shapiro.test(DATOS$FA[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])
hist(DATOS$FA[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])

###FOSFATASA ALCALINA EN MUJERES CON LITO AUSENTE###
shapiro.test(DATOS$FA[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])
hist(DATOS$FA[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])

###FOSFATASA ALCALINA EN HOMBRES Y LITO PRESENTE###
shapiro.test(DATOS$FA[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])
hist(DATOS$FA[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])

###FOSFATASA ALCALINA EN HOMBRES Y LITO AUSENTE###
shapiro.test(DATOS$FA[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])
hist(DATOS$FA[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])

###VARIABLE BILIRRUBINA TOTAL###
DATOS$BT <- as.numeric(DATOS$BT)

###BILIRRUBINA TOTAL Y LITO PRESENTE###
shapiro.test(DATOS$BT[DATOS$LITO_COLANGIO==1])
hist(DATOS$BT[DATOS$LITO_COLANGIO==1])

###BILIRRUBINA TOTAL Y LITO AUSENTE###
shapiro.test(DATOS$BT[DATOS$LITO_COLANGIO==0])
hist(DATOS$BT[DATOS$LITO_COLANGIO==0])

###BILIRRUBINA TOTAL EN MUJERES CON LITO PRESENTE###
shapiro.test(DATOS$BT[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])
hist(DATOS$BT[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])

###BILIRRUBINA TOTAL EN MUJERES CON LITO AUSENTE###
shapiro.test(DATOS$BT[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])
hist(DATOS$BT[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])

###BILIRRUBINA TOTAL EN HOMBRES CON LITO PRESENTE###
shapiro.test(DATOS$BT[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])
hist(DATOS$BT[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])

###BILIRRUBINA TOTAL EN HOMBRES CON LITO AUSENTE###
shapiro.test(DATOS$BT[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])
hist(DATOS$BT[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])

###VARIABLE ALANINOAMINOTRANSFERASA###
DATOS$ALT_TGP <- as.numeric(DATOS$ALT_TGP)

###ALANINOAMINOTRANSFERASA Y LITO PRESENTE###
shapiro.test(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==1])
hist(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==1])

###ALANINOAMINOTRANSFERASA Y LITO AUSENTE###
shapiro.test(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==0])
hist(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==0])

###ALANINOAMINOTRANSFERASA EN MUJERES Y LITO PRESENTE###
shapiro.test(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])
hist(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])

###ALANINOAMINOTRANSFERASA EN MUJERES Y LITO AUSENTE###
shapiro.test(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])
hist(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])

###ALANINOAMINOTRANSFERASA EN HOMBRES Y LITO PRESENTE###
shapiro.test(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])
hist(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])

###ALANINOAMINOTRANSFERASA EN HOMBRES Y LITO AUSENTE###
shapiro.test(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])
hist(DATOS$ALT_TGP[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])
```

```

###VARIABLE ASPARTATO AMINOTRANSFERASA###
DATOS$AST_TGO <- as.numeric(DATOS$AST_TGO)

###ASPARTATO AMINOTRANSFERASA Y LITO PRESENTE###
shapiro.test(DATOS$AST_TGO[DATOS$LITO_COLANGIO==1])
hist(DATOS$AST_TGO[DATOS$LITO_COLANGIO==1])

###ASPARTATO AMINOTRANSFERASA Y LITO AUSENTE###
shapiro.test(DATOS$AST_TGO[DATOS$LITO_COLANGIO==0])
hist(DATOS$AST_TGO[DATOS$LITO_COLANGIO==0])

###ASPARTATO AMINOTRANSFERASA EN MUJERES CON LITO PRESENTE###
shapiro.test(DATOS$AST_TGO[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])
hist(DATOS$AST_TGO[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])

###ASPARTATO AMINOTRANSFERASA EN MUJERES CON LITO AUSENTE###
shapiro.test(DATOS$AST_TGO[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])
hist(DATOS$AST_TGO[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])

###ASPARTATO AMINOTRANSFERASA EN HOMBRERES CON LITO PRESENTE###
shapiro.test(DATOS$AST_TGO[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])
hist(DATOS$AST_TGO[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])

###ASPARTATO AMINOTRANSFERASA EN HOMBRERES CON LITO AUSENTE###
shapiro.test(DATOS$AST_TGO[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])
hist(DATOS$AST_TGO[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])

###VARIABLE GAMA GLUTAMILTRANSFERASA###
DATOS$GGT <- as.numeric(DATOS$GGT)

###GAMA GLUTAMILTRANSFERASA Y LITO PRESENTE###
shapiro.test(DATOS$GGT[DATOS$LITO_COLANGIO==1])
hist(DATOS$GGT[DATOS$LITO_COLANGIO==1])

###GAMA GLUTAMILTRANSFERASA Y LITO AUSENTE###
shapiro.test(DATOS$GGT[DATOS$LITO_COLANGIO==0])
hist(DATOS$GGT[DATOS$LITO_COLANGIO==0])

###GAMA GLUTAMILTRANSFERASA EN MUJERES CON LITO PRESENTE###
shapiro.test(DATOS$GGT[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])
hist(DATOS$GGT[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])

###GAMA GLUTAMILTRANSFERASA EN MUJERES CON LITO AUSENTE###
shapiro.test(DATOS$GGT[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])
hist(DATOS$GGT[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])

###GAMA GLUTAMILTRANSFERASA EN HOMBRERES CON LITO PRESENTE###
shapiro.test(DATOS$GGT[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])
hist(DATOS$GGT[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])

###GAMA GLUTAMILTRANSFERASA EN HOMBRERES CON LITO AUSENTE###
shapiro.test(DATOS$GGT[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])
hist(DATOS$GGT[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])

###VARIABLE COLEDOCO###
DATOS$CD <- as.numeric(DATOS$CD)

###TAMAÑO DE COLEDOCO CON LITO PRESENTE###
shapiro.test(DATOS$CD[DATOS$LITO_COLANGIO==1])
hist(DATOS$CD[DATOS$LITO_COLANGIO==1])

###TAMAÑO DE COLEDOCO CON LITO AUSENTE###
shapiro.test(DATOS$CD[DATOS$LITO_COLANGIO==0])
hist(DATOS$CD[DATOS$LITO_COLANGIO==0])

###TAMAÑO DE COLEDOCO EN MUJERES CON LITO PRESENTE###
shapiro.test(DATOS$CD[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])
hist(DATOS$CD[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==0])

###TAMAÑO DE COLEDOCO EN MUJERES CON LITO AUSENTE###
shapiro.test(DATOS$CD[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])
hist(DATOS$CD[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==0])

###TAMAÑO DE COLEDOCO EN HOMBRERES CON LITO PRESENTE###
shapiro.test(DATOS$CD[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])
hist(DATOS$CD[DATOS$LITO_COLANGIO==1 & DATOS$SEXO==1])

###TAMAÑO DE COLEDOCO EN HOMBRERES CON LITO AUSENTE###
shapiro.test(DATOS$CD[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])
hist(DATOS$CD[DATOS$LITO_COLANGIO==0 & DATOS$SEXO==1])

```

```
#6.APLICACION DE PRUEBA U DE MANN-WHITNEY.....

####VARIABLE EDAD###
DATOS$EDAD<- as.numeric(DATOS$EDAD)
wilcox.test(EDAD~LITO_COLANGIO, alternative="two.sided", data=DATOS)

####EDAD EN MUJERES###
DATOSEDA_MUJERES<-subset(DATOS,SEXO==0)
wilcox.test(EDAD~LITO_COLANGIO, alternative="two.sided", data=DATOSEDA_MUJERES)

#### VARIABLE FOSFATASA###
DATOS$FA <- as.numeric(DATOS$FA)
wilcox.test(FA~LITO_COLANGIO, alternative="two.sided", data=DATOS)

#### FOSFATASA EN MUJERES###
DATOSFA_MUJERES<-subset(DATOS,SEXO==0)
wilcox.test(FA~LITO_COLANGIO, alternative="two.sided", data=DATOSFA_MUJERES)

#### FOSFATASA EN HOMBRES###
DATOSFA_HOMBRES<-subset(DATOS,SEXO==1)
wilcox.test(FA~LITO_COLANGIO, alternative="two.sided", data=DATOSFA_HOMBRES)

####VARIABLE BILIRRUBINTA TOTAL###
DATOS$BT <- as.numeric(DATOS$BT)
wilcox.test(BT~LITO_COLANGIO, alternative="two.sided", data=DATOS)

####BILIRRUBINTA TOTAL EN MUJERES###
DATOSBT_MUJERES<-subset(DATOS,SEXO==0)
wilcox.test(BT~LITO_COLANGIO, alternative="two.sided", data=DATOSBT_MUJERES)

####BILIRRUBINTA TOTAL EN HOMBRES###
DATOSBT_HOMBRES<-subset(DATOS,SEXO==1)
wilcox.test(BT~LITO_COLANGIO, alternative="two.sided", data=DATOSBT_HOMBRES)

####VARIABLE ALANINOAMINOTRASFERASA###
DATOS$ALT_TGP <- as.numeric(DATOS$ALT_TGP)
wilcox.test(ALT_TGP~LITO_COLANGIO, alternative="two.sided", data=DATOS)

####VARIABLE ALANINOAMINOTRASFERASA EN MUJERES###
DATOSALT_TGP_MUJERES<-subset(DATOS,SEXO==0)
wilcox.test(ALT_TGP~LITO_COLANGIO, alternative="two.sided", data=DATOSALT_TGP_MUJERES)

####VARIABLE ALANINOAMINOTRASFERASA EN HOMBRES###
DATOSALT_TGP_HOMBRES<-subset(DATOS,SEXO==1)
wilcox.test(ALT_TGP~LITO_COLANGIO, alternative="two.sided", data=DATOSALT_TGP_HOMBRES)

####VARIABLE ASPARTATO AMINOTRASFERASA###
DATOS$AST_TGO <- as.numeric(DATOS$AST_TGO)
wilcox.test(AST_TGO~LITO_COLANGIO, alternative="two.sided", data=DATOS)

####ASPARTATO AMINOTRASFERASA EN MUJERES###
DATOSAST_TGO_MUJERES<-subset(DATOS,SEXO==0)
wilcox.test(AST_TGO~LITO_COLANGIO, alternative="two.sided", data=DATOSAST_TGO_MUJERES)

####ASPARTATO AMINOTRASFERASA EN HOMBRES###
DATOSAST_TGO_HOMBRES<-subset(DATOS,SEXO==1)
wilcox.test(AST_TGO~LITO_COLANGIO, alternative="two.sided", data=DATOSAST_TGO_HOMBRES)

####VARIABLE GAMAGLUTAMIL TRANSFERASA###
DATOS$GGT <- as.numeric(DATOS$GGT)
wilcox.test(GGT~LITO_COLANGIO, alternative="two.sided", data=DATOS)

####GAMAGLUTAMIL TRANSFERASA EN MUJERES###
DATOSGGT_MUJERES<-subset(DATOS,SEXO==0)
wilcox.test(GGT~LITO_COLANGIO, alternative="two.sided", data=DATOSGGT_MUJERES)

####GAMAGLUTAMIL TRANSFERASA EN HOMBRES###
DATOSGGT_HOMBRES<-subset(DATOS,SEXO==1)
wilcox.test(GGT~LITO_COLANGIO, alternative="two.sided", data=DATOSGGT_HOMBRES)

####VARIABLE TAMAÑO DE COLEDOCO###
DATOS$CD <- as.numeric(DATOS$CD)
wilcox.test(CD~LITO_COLANGIO, alternative="two.sided", data=DATOS)

####TAMAÑO DE COLEDOCO EN HOMBRES###
DATOSCD_HOMBRES<-subset(DATOS,SEXO==1)
wilcox.test(CD~LITO_COLANGIO, alternative="two.sided", data=DATOSCD_HOMBRES)

#7.APLICACIÓN DE PRUEBA T STUDENT.....

####VARIABLE EDAD EN HOMBRES###
DATOS$EDAD <- as.numeric(DATOS$EDAD)
DATOSEDA_HOMBRES<-subset(DATOS,SEXO==1)
t.test(EDAD ~ LITO_COLANGIO, data = DATOSEDA_HOMBRES)

####VARIABLE IMC###
DATOS$IMC <- as.numeric(DATOS$IMC)
t.test(IMC~LITO_COLANGIO, data=DATOS)

####VARIABLE IMC EN MUJERES###
DATOSIMC_MUJERES<-subset(DATOS,SEXO==0)
t.test(IMC ~ LITO_COLANGIO, data = DATOSIMC_MUJERES)

####VARIABLE IMC EN HOMBRES###
DATOSIMC_HOMBRES<-subset(DATOS,SEXO==1)
t.test(IMC ~ LITO_COLANGIO, data = DATOSIMC_HOMBRES)

#VARIABLE TAMAÑO COLEDOCO###
DATOS$CD <- as.numeric(DATOS$CD)

####TAMAÑO COLEDOCO EN MUJERES###
DATOSCD_MUJERES<-subset(DATOS,SEXO==0)
t.test(CD ~ LITO_COLANGIO, data = DATOSCD_MUJERES)
```

```

#8.APLICACION DE PRUEBA CHI CUADRADA.....

###VARIABLE SEXO###
chisq.test(DATOS$SEXO,DATOS$LITO_COLANGIO, correct = TRUE)

###VARIABLE DM2###
chisq.test(DATOS$DM2,DATOS$LITO_COLANGIO, correct = TRUE)

###VARIABLE HTA###
chisq.test(DATOS$HTA,DATOS$LITO_COLANGIO, correct = TRUE)

###VARIABLE TABAQUISMO###
chisq.test(DATOS$TAB,DATOS$LITO_COLANGIO, correct = TRUE)

### VARIABLE ALCOHOL###
chisq.test(DATOS$ALCOHOL,DATOS$LITO_COLANGIO, correct = TRUE)

###VARIABLE DIAGNOSTICO PREVIO###
chisq.test(DATOS$DX_PRE,DATOS$LITO_COLANGIO, correct = TRUE)

###VARIABLE RIESGO ASGE###
chisq.test(DATOS$RIESGO_ASGE,DATOS$LITO_COLANGIO, correct = TRUE)

###VARIABLE PRESENCIA DE LITO###
chisq.test(DATOS$LITO,DATOS$LITO_COLANGIO, correct = TRUE)

###VARIABLE PANCREATITIS###
chisq.test(DATOS$PANCREATITIS,DATOS$LITO_COLANGIO, correct = TRUE)

###VARIABLE COLANGITIS###
chisq.test(DATOS$COLANGITIS,DATOS$LITO_COLANGIO, correct = TRUE)

#9.REGRESION LOGISTICA POR VARIABLE.....

###MODELO DE REGRESION LOGISTICA CON VARIABLE IMC###
modelologistico.1 <- glm(as.factor(LITO_COLANGIO)~as.numeric(IMC), data=DATOS,family="binomial")
summary(modelologistico.1)

#PROBABILIDAD DE COLEDOCOLITIASIS RESPECTO A IMC###
prob.piedras.IMC<- exp(1.76747-0.05316)/(1+exp(1.76747-0.05316))
print(prob.piedras.IMC)

###ODDS RATIO DE IMC###
ratio.IMC <- prob.piedras.IMC / (1-prob.piedras.IMC)
print(ratio.IMC)

###ODDS RATIO Y SU INTERVALO DE CONFIANZA EXPONENCIADOS###
exp(cbind(coef(modelologistico.1), confint(modelologistico.1)))

#HIPOTESIS NULA = 1.76747
#HIPOTESIS VERDADERA = -0.05316
#VALOR P= 0.0356

###DADO QUE EL VALOR P ES MENOR A 0.05, RECHAZAMOS LA HIPOTESIS NULA,LA VARIABLE IMC TIENE PODER
#EXPLICATIVO EN PADECER COLEDOCOLITIASIS###

###MODELO DE REGRESION LOGISTICA CON VARIABLE RIESGO ASGE###
modelologistico.2 <- glm(as.factor(LITO_COLANGIO)~as.numeric(IMC)+as.factor(RIESGO_ASGE), data=DATOS,family="binomial")
summary(modelologistico.2)

#PROBABILIDAD DE COLEDOCOLITIASIS RESPECTO A RIESGO ASGE###
prob.piedras.RIESGO_ASGE<-exp(1.13368-0.04458+0.74205)/(1+ exp (1.13368-0.04458+0.74205))
print(prob.piedras.RIESGO_ASGE)

#ODDS RATIO DE RIESGO ASGE##
ratio.RIESGO_ASGE<-prob.piedras.RIESGO_ASGE/(1-prob.piedras.RIESGO_ASGE)
print(ratio.RIESGO_ASGE)

###ODDS RATIO Y SU INTERVALO DE CONFIANZA EXPONENCIADOS###
exp(cbind(coef(modelologistico.2), confint(modelologistico.2)))

#HIPOTESIS NULA= 1.13368
#HIPOTESIS VERDADERA = 0.74205
# VALOR P = 0.0133

###DADO QUE EL VALOR P ES MENOR A 0.05, RECHAZAMOS LA HIPOTESIS NULA,LA VARIABLE RIESGO ASGE TIENE PODER
#EXPLICATIVO EN PADECER COLEDOCOLITIASIS###

```

TESIS TESIS TESIS TESIS TESIS



TESIS TESIS TESIS TESIS TESIS

```

####MODELO DE REGRESION LOGISTICA CON VARIABLE FOSFATASA ALCALINA###
modelologistico.3 <- glm(as.factor(LITO_COLANGIO)~as.numeric(IMC)+
  as.factor(RIESGO_ASGE)+ as.numeric (FA), data=DATOS,family="binomial")
summary(modelologistico.3)

#PROBABILIDAD DE COLEDOCOLITIASIS RESPECTO A FOSFATASA ALCALINA###
prob.piedras.FA<- exp(0.6203412-0.0389987+0.7021587+0.0011850)/(1+exp (0.6203412-0.0389987+0.7021587+0.0011850 ))
print(prob.piedras.FA)
ratio.FA<-(prob.piedras.FA)/(1-prob.piedras.FA)
print(ratio.FA)

#ODDS RATIO DE FOSFATASA ALCALINA##
(0.8098448)/(1-0.8098448)

####ODDS RATIO Y SU INTERVALO DE CONFIANZA EXPONENCIADOS###
exp(cbind(coef(modelologistico.3), confInt(modelologistico.3)))

#HIPOTESIS NULA=0.8437
#HIPOTESIS VERDADERA = 0.0010
# VALOR P = 0.0966

####DADO QUE EL VALOR P ES MAYOR A 0.05, NO RECHAZAMOS LA HIPOTESIS NULA, Y LA VARIABLE FOSFATASA ALCALINA
#NO TIENE PODER EXPLICATIVO EN PADECER COLEDOCOLITIASIS###

####MODELO DE REGRESION LOGISTICA CON VARIABLE GAMA GLUTAMIL TRANSFERASA###
modelologistico.4 <- glm(as.factor(LITO_COLANGIO)~as.numeric(IMC)+
  as.factor(RIESGO_ASGE)+ as.numeric (FA)+ as.numeric (GGT), data=DATOS,family="binomial")
summary(modelologistico.4)

#PROBABILIDAD DE COLEDOCOLITIASIS RESPECTO A GAMA GLUTAMIL TRANSFERASA###
prob.piedras.GGT<-exp (0.4399598-0.041432+0.7582962+0.0005316+0.0008773)/(1+exp(0.4399598-0.041432+0.7582962+0.0005316+0.0008773))
print(prob.piedras.GGT)
ratio.GGT<-(prob.piedras.GGT)/(1-prob.piedras.GGT)
print(ratio.GGT)

####ODDS RATIO Y SU INTERVALO DE CONFIANZA EXPONENCIADOS###
exp(cbind(coef(modelologistico.4), confInt(modelologistico.4)))

#HIPOTESIS NULA= 0.9042
#HIPOTESIS VERDADERA = 0.0009
# VALOR P = 0.034

####DADO QUE EL VALOR P ES MENOR A 0.05, RECHAZAMOS LA HIPOTESIS NULA, LA VARIABLE GAMA GLUTAMIL TRANSFERASA TIENE PODER
#EXPLICATIVO EN PADECER COLEDOCOLITIASIS###

####MODELO DE REGRESION LOGISTICA CON VARIABLE PRESENCIA DE LITO###
modelologistico.5 <- glm(as.factor(LITO_COLANGIO)~as.numeric(IMC)+
  as.factor(RIESGO_ASGE)+ as.numeric (FA)+
  as.numeric (GGT)+ as.factor (LITO), data=DATOS,family="binomial")
summary(modelologistico.5)

#PROBABILIDAD DE COLEDOCOLITIASIS RESPECTO A PRESENCIA DE LITO ###
prob.piedras.LITO<-exp(0.3703030-0.0408966+0.5246047+0.0006079+0.0009200+0.5974993)/(1+exp(0.3703030-0.0408966+0.5246047+0.0006079+0.0009200+0.5974993))
print(prob.piedras.LITO)
ratio.LITO<-(prob.piedras.LITO)/(1-prob.piedras.LITO)
print(ratio.LITO)

#ODDS RATIO DE PRESENCIA DE LITO ###
0.8324183/(1-0.8324183)

####ODDS RATIO Y SU INTERVALO DE CONFIANZA EXPONENCIADOS###
exp(cbind(coef(modelologistico.5), confInt(modelologistico.5)))

#HIPOTESIS NULA= 0.8685
#HIPOTESIS VERDADERA = 0.5484
# VALOR P = 0.207

####DADO QUE EL VALOR P ES MAYOR A 0.05, NO RECHAZAMOS LA HIPOTESIS NULA, Y LA VARIABLE PRESENCIA DE LITO
#NO TIENE PODER EXPLICATIVO EN PADECER COLEDOCOLITIASIS###

```

```

#10.SELECCION DEL MODELO.....

DATOS$LITO_COLANGIO <- as.factor(DATOS$LITO_COLANGIO)
DATOS$IMC <- as.numeric(DATOS$IMC)
DATOS$RIESGO_ASGE <- as.factor(DATOS$RIESGO_ASGE)
DATOS$FA <- as.numeric(DATOS$FA)
DATOS$GGT <- as.numeric(DATOS$GGT)
DATOS$LITO <- as.factor(DATOS$LITO)

modelologistico.5.1 <- glm(LITO_COLANGIO~IMC+RIESGO_ASGE+FA+GGT+LITO, data=DATOS,family="binomial")
summary(modelologistico.5.1)

#SE REALIZA LA SELECCION DEL MODELO CON AKAIKE
step(modelologistico.5.1, direction="backward")

#MODELO DEFINITIVO CON 3 VARIABLES (IMC Y GGT, AIC = 258.6)
modelologistico.5.2 <- glm(LITO_COLANGIO ~ IMC + GGT + LITO, family = "binomial",
                           data = DATOS)
summary(modelologistico.5.2)
exp(modelologistico.5.2$coefficients)

#PROBABILIDADES ESTIMADAS POR INDIVIDUO
pred<-predict(modelologistico.5.2,type="response")
pred

#11. MODELO PRIMARIO.....

library("caret")
library(ROCR)
library("caret")

#SE INDICA COMO LEER CADA UNA DE LAS VARIABLES:

DATOS$IMC <- as.numeric(DATOS$IMC)
DATOS$GGT <- as.numeric(DATOS$GGT)
DATOS$LITO <- as.factor(DATOS$LITO)
DATOS$LITO_COLANGIO<-as.factor(DATOS$LITO_COLANGIO)

#MODELO SELECCIONADO
modelo.3 <- glm(formula = LITO_COLANGIO ~ IMC + GGT + LITO,data= DATOS, family = "binomial")
summary(modelo.3)

### IDENTIFICACION DE LOS 2 GRUPOS DE LITO_COLANGIO###
indice.lc.ausente<-which(DATOS$LITO_COLANGIO==0)
indice.lc.presente<-which(DATOS$LITO_COLANGIO==1)

##CORROBORAR LAS SEPARACIONES##
table(DATOS$LITO_COLANGIO[indice.lc.ausente])
table(DATOS$LITO_COLANGIO[indice.lc.presente])

##INDICO BOOTSTRAP ( 20 R)
n_repeticiones <- 20

```

TESIS TESIS TESIS TESIS TESIS



TESIS TESIS TESIS TESIS TESIS

```

#DEFINIR LA TABLA BASE##
metricas <- data.frame(BOOTSTRAP=seq(1,n_repeticiones), SENSITIVITY=0, SPECIFICITY=0,
POS.PRED.VALUE=0, NEG.PRED.VALUE=0, ACCURACY=0,
PRECISION=0, RVP=0, RVN=0, ODDS.RATIO=0, AUC=0)

## SE INICIA EL BOOTSTRAP##
predicciones<-list()
etiquetas<-list()
set.seed(14042025)

for(i in 1:n_repeticiones) {
  ##DEFINIMOS GRUPO DE ENTRENAMIENTO##
  posiciones<-c(sample(x=indice.lc.ausente, size=length(indice.lc.ausente), replace=TRUE),
sample(x=indice.lc.presente, size=length(indice.lc.presente), replace=TRUE))
  ##GRUPO DE ENTRENAMIENTO##
  entrenamiento <- DATOS[posiciones,]
  ##GRUPO DE PRUEBA##
  prueba <- DATOS[-posiciones,]
  modelo.3.e <- glm(formula = LITO_COLANGIO ~ IMC + GGT + LITO, data = entrenamiento,
family = "binomial")
  ##NOTA: se realizan predicciones
  predicciones_prob_boot <- predict(modelo.3.e, newdata = prueba, type = "response")
  predicciones_clase <- ifelse(predicciones_prob_boot > 0.60, 1, 0)

  #esto es para la AUC
  predicciones[[i]]<-predicciones_prob_boot
  etiquetas[[i]]<-prueba$LITO_COLANGIO

  #ESTO ES PARA EL CALCULO DE AUC
  pred<-prediction(predicciones_prob_boot,prueba$LITO_COLANGIO)
  metricas$AUC[i]<-performance(pred,"auc")@y.values

  #NOTA: HACEMOS MATRIZ DE CONFUSION##
  matriz_confusion<-confusionMatrix(as.factor(predicciones_clase),
reference=prueba$LITO_COLANGIO,
positive="1")

  #NOTA:ahora se almacenan las metricas una por una
  metricas$SENSITIVITY[i] <- matriz_confusion$byClass[1]
  metricas$SPECIFICITY[i] <- matriz_confusion$byClass[2]
  metricas$POS.PRED.VALUE[i] <- matriz_confusion$byClass[3]
  metricas$NEG.PRED.VALUE[i] <- matriz_confusion$byClass[4]
  metricas$PRECISION[i] <- matriz_confusion$byClass[5]
  metricas$ACCURACY[i] <- matriz_confusion$overall[1]
  metricas$RVP[i] <- metricas$SENSITIVITY[i] / (1-metricas$SPECIFICITY[i])
  metricas$RVN[i] <- (1-metricas$SENSITIVITY[i]) / metricas$SPECIFICITY[i]
  #CALCULO DE ODDS RATIO
  metricas$ODDS.RATIO[i] <- metricas$RVP[i] / metricas$RVN[i]
}
metricas

mean(metricas$SENSITIVITY)
mean(metricas$SPECIFICITY)
mean(metricas$POS.PRED.VALUE)
mean(metricas$NEG.PRED.VALUE)
mean(metricas$ACCURACY)
mean(metricas$PRECISION)
mean(metricas$RVP)
mean(metricas$RVN)
mean(metricas$ODDS.RATIO)
mean(metricas$AUC)
promedio_AUC <- as.numeric(metricas$AUC,na.rm = TRUE)
mean(promedio_AUC)

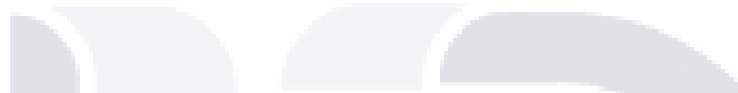
#apply(metricas,2,mean,na.rm=TRUE)
##AHORA SE CALCULARIA CURVA ROC
ROCpred <- prediction(predicciones,etiquetas)
ROCperf <- performance(ROCpred, "tpr", "fpr")

#NOTA SE GRAFICA ASI
plot(ROCperf, colorize=TRUE, print.cutoff.at=seq(0.1,by=0.1), avg="threshold", lwd=3)
plot(ROCperf, col="grey78", add=TRUE)

#CALCULO DE PREVALENCIA DE LA MUESTRA
prevalencia_tabla <- prop.table(table(DATOS$LITO_COLANGIO))
probabilidad_pre_prueba <- prevalencia_tabla["1"]
cat("Usando prop.table:", probabilidad_pre_prueba, "\n")

#CALCULO DE PREVALENCIA POST-PRUEBA DE LA MUESTRA
prevalencia_tabla <- prop.table(table(prueba$LITO_COLANGIO))
probabilidad_pre_prueba <- prevalencia_tabla["1"]
cat("Usando prop.table:", probabilidad_pre_prueba, "\n")

```



```
#12.MODELO SECUNDARIO.....

library("caret")
library(ROCR)

#SE INDICA COMO LEER CADA UNA DE LAS VARIABLES:
DATOS$EDAD<-as.numeric(DATOS$EDAD)
DATOS$SEXO<-as.factor(DATOS$SEXO)
DATOS$IMC <- as.numeric(DATOS$IMC)
DATOS$DM2<-as.factor(DATOS$DM2)
DATOS$HTA<-as.factor(DATOS$HTA)
DATOS$TAB<-as.factor(DATOS$TAB)
DATOS$ALCOHOL<-as.factor(DATOS$ALCOHOL)
DATOS$DX_PRE<-as.factor(DATOS$DX_PRE)
DATOS$RIESGO_ASGE <- as.factor(DATOS$RIESGO_ASGE)
DATOS$FA <- as.numeric(DATOS$FA)
DATOS$BT<-as.numeric(DATOS$BT)
DATOS$AST_TGO<-as.numeric(DATOS$AST_TGO)
DATOS$ALT_TGP<-as.numeric(DATOS$ALT_TGP)
DATOS$GGT <- as.numeric(DATOS$GGT)
DATOS$CD<-as.numeric(DATOS$CD)
DATOS$LITO <- as.factor(DATOS$LITO)
DATOS$PANCREATITIS<-as.factor(DATOS$PANCREATITIS)
DATOS$COLANGITIS<-as.factor(DATOS$COLANGITIS)
DATOS$LITO_COLANGIO<-as.factor(DATOS$LITO_COLANGIO)

#MODELO PREDICTIVO Y DESCARTE POR AKAIKE
modelo_de_ary<- glm(LITO_COLANGIO~EDAD+SEXO+IMC+DM2+HTA+TAB+ALCOHOL+DX_PRE+RIESGO_ASGE+FA+BT+
                    AST_TGO+ALT_TGP+GGT+CD+LITO+PANCREATITIS+COLANGITIS, data=DATOS,family="binomial")

summary(modelo_de_ary)

#SE REALIZA LA SELECCION DEL MODELO CON AKAIKE
step(modelo_de_ary, direction="backward")

#SELECCION DE MODELO CON AKAIKE MAS BAJO= 251.69
modelo.5<-glm(formula = LITO_COLANGIO ~ DM2 + HTA + DX_PRE + GGT + LITO,
              family = "binomial", data = DATOS)

summary(modelo.5)

### IDENTIFICACION DE LOS 2 GRUPOS DE LITO_COLANGIO###
indice.lc.ausente<-which(DATOS$LITO_COLANGIO==0)
indice.lc.presente<-which(DATOS$LITO_COLANGIO==1)

##CORROBORAR LAS SEPARACIONES##
table(DATOS$LITO_COLANGIO[indice.lc.ausente])
table(DATOS$LITO_COLANGIO[indice.lc.presente])

##INDICO BOOTSTRAP ( 20 R)
n_repeticiones <- 20

#DEFINIR LA TABLA BASE##
metrics <- data.frame(BOOTSTRAP=seq(1,n_repeticiones), SENSITIVITY=0, SPECIFICITY=0,
                      POS.PRED.VALUE=0, NEG.PRED.VALUE=0, ACCURACY=0,
                      PRECISION=0, RVP=0, RVN=0, ODDS.RATIO=0, AUC=0)
```

TESIS TESIS TESIS TESIS TESIS



TESIS TESIS TESIS TESIS TESIS

```

## SE INICIA EL BOOTSTRAP##
predicciones<-list()
etiquetas<-list()
set.seed(14042025)

for(i in 1:n_repeticiones) {
  ##DEFINIMOS GRUPO DE ENTRENAMIENTO##
  posiciones<-c(sample(x=indice.lc.ausente, size=length(indice.lc.ausente), replace=TRUE),
                sample(x=indice.lc.presente, size=length(indice.lc.presente), replace=TRUE))
  ###GRUPO DE ENTRENAMIENTO##
  entrenamiento <- DATOS[posiciones,]
  ##GRUPO DE PRUEBA##
  prueba <- DATOS[-posiciones,]
  modelo.8.e <- glm(formula = LITO_COLANGIO ~ DM2 + HTA + DX_PRE + GGT + LITO,
                    family = "binomial", data = DATOS)
  ##NOTA: se realizan predicciones
  predicciones_prob_boot <- predict(modelo.8.e, newdata = prueba, type = "response")
  predicciones_clase <- ifelse(predicciones_prob_boot > 0.60, 1, 0)

  #esto es para la AUC
  predicciones[[i]]<-predicciones_prob_boot
  etiquetas[[i]]<-prueba$LITO_COLANGIO

  #ESTO ES PARA EL CALCULO DE AUC
  pred<-prediction(predicciones_prob_boot,prueba$LITO_COLANGIO)
  metricas$AUC[i]<-performance(pred,"auc")@y.values

  #NOTA: HACEMOS MATRIZ DE CONFUSION##
  matriz_confusion<-confusionMatrix(as.factor(predicciones_clase),
                                    reference=prueba$LITO_COLANGIO,
                                    positive="1")

  #NOTA:ahora se almacenan las metricas una por una
  metricas$SENSITIVITY[i] <- matriz_confusion$byClass[1]
  metricas$SPECIFICITY[i] <- matriz_confusion$byClass[2]
  metricas$POS.PRED.VALUE[i] <- matriz_confusion$byClass[3]
  metricas$NEG.PRED.VALUE[i] <- matriz_confusion$byClass[4]
  metricas$PRECISION[i] <- matriz_confusion$byClass[5]
  metricas$ACCURACY[i] <- matriz_confusion$overall[1]
  metricas$RVP[i] <- metricas$SENSITIVITY[i] / (1-metricas$SPECIFICITY[i])
  metricas$RVN[i] <- (1-metricas$SENSITIVITY[i]) / metricas$SPECIFICITY[i]
  #CALCULO DE ODDS RATIO
  metricas$ODDS.RATIO[i] <- metricas$RVP[i] / metricas$RVN[i]
}
metricas

mean(metricas$SENSITIVITY)
mean(metricas$SPECIFICITY)
mean(metricas$POS.PRED.VALUE)
mean(metricas$NEG.PRED.VALUE)
mean(metricas$ACCURACY)
mean(metricas$PRECISION)
mean(metricas$RVP)
mean(metricas$RVN)
mean(metricas$ODDS.RATIO)
mean(metricas$AUC)
promedio_AUC <- as.numeric(metricas$AUC,na.rm = TRUE)
mean(promedio_AUC)

#apply(metricas,2,mean,na.rm=TRUE)
##AHORA SE CALCULARIA CURVA ROC
ROCpred <- prediction(predicciones,etiquetas)
ROCperf <- performance(ROCpred, "tpr", "fpr")

plot(ROCperf, colorize=TRUE, print.cutoff.at=seq(0.1,by=0.1), avg="threshold", lwd=3)
plot(ROCperf, col="grey78", add=TRUE)

```

```

#13.MODELO TERCARIO.....

library("caret")
library(ROCR)

#SE INDICA COMO LEER CADA UNA DE LAS VARIABLES:
DATOS$EDAD<-as.numeric(DATOS$EDAD)
DATOS$SEXO<-as.factor(DATOS$SEXO)
DATOS$IMC <- as.numeric(DATOS$IMC)
DATOS$DM2<-as.factor(DATOS$DM2)
DATOS$HTA<-as.factor(DATOS$HTA)
DATOS$TAB<-as.factor(DATOS$TAB)
DATOS$ALCOHOL<-as.factor(DATOS$ALCOHOL)
DATOS$DX_PRE<-as.factor(DATOS$DX_PRE)
DATOS$RIESGO_ASGE <- as.factor(DATOS$RIESGO_ASGE)
DATOS$FA <- as.numeric(DATOS$FA)
DATOS$BT<-as.numeric(DATOS$BT)
DATOS$AST_TGO<-as.numeric(DATOS$AST_TGO)
DATOS$ALT_TGP<-as.numeric(DATOS$ALT_TGP)
DATOS$GGT <- as.numeric(DATOS$GGT)
DATOS$CD<-as.numeric(DATOS$CD)
DATOS$LITO <- as.factor(DATOS$LITO)
DATOS$PANCREATITIS<-as.factor(DATOS$PANCREATITIS)
DATOS$COLANGITIS<-as.factor(DATOS$COLANGITIS)
DATOS$LITO_COLANGIO<-as.factor(DATOS$LITO_COLANGIO)

#MODELO PREDICTIVO Y DESCARTE POR AKAIKE
modelo.18<- glm(LITO_COLANGIO~EDAD+SEXO+IMC+DM2+HTA+TAB+ALCOHOL+DX_PRE+RIESGO_ASGE+FA+BT+
                AST_TGO+ALT_TGP+GGT+CD+LITO+PANCREATITIS+COLANGITIS, data=DATOS,family="binomial")

summary(modelo.18)

### IDENTIFICACION DE LOS 2 GRUPOS DE LITO_COLANGIO###
indice.lc.ausente<-which(DATOS$LITO_COLANGIO==0)
indice.lc.presente<-which(DATOS$LITO_COLANGIO==1)

##CORROBORAR LAS SEPARACIONES##
table(DATOS$LITO_COLANGIO[indice.lc.ausente])
table(DATOS$LITO_COLANGIO[indice.lc.presente])

##INDICO BOOTSTRAP ( 20 R)
n_repeticiones <- 20

#DEFINIR LA TABLA BASE##
metricas <- data.frame(BOOTSTRAP=seq(1,n_repeticiones), SENSITIVITY=0, SPECIFICITY=0,
                        POS.PRED.VALUE=0, NEG.PRED.VALUE=0, ACCURACY=0,
                        PRECISION=0, RVP=0, RVN=0, ODDS.RATIO=0, AUC=0)
  
```



```
## SE INICIA EL BOOTSTRAP##
predicciones<-list()
etiquetas<-list()
set.seed(14042025)

for(i in 1:n_repeticiones) {
  ##DEFINIMOS GRUPO DE ENTRENAMIENTO##
  posiciones<-c(sample(x=indice.lc.ausente, size=length(indice.lc.ausente), replace=TRUE),
                sample(x=indice.lc.presente, size=length(indice.lc.presente), replace=TRUE))
  ###GRUPO DE ENTRENAMIENTO##
  entrenamiento <- DATOS[posiciones,]
  ##GRUPO DE PRUEBA##
  prueba <- DATOS[-posiciones,]
  modelo.18.e <- glm(formula = LITO_COLANGIO~EDAD+SEXO+IMC+DM2+HTA+TAB+ALCOHOL+DX_PRE+RIESGO_ASGE+FA+BT+
                    AST_TGO+ALT_TGP+GGT+CD+LITO+PANCREATITIS+COLANGITIS, data = entrenamiento,
                    family = "binomial")
  ##NOTA: se realizan predicciones
  predicciones_prob_boot <- predict(modelo.18.e, newdata = prueba, type = "response")
  predicciones_clase <- ifelse(predicciones_prob_boot > 0.60, 1, 0)

  #esto es para la AUC
  predicciones[[i]]<-predicciones_prob_boot
  etiquetas[[i]]<-prueba$LITO_COLANGIO

  #ESTO ES PARA EL CALCULO DE AUC
  pred<-prediction(predicciones_prob_boot,prueba$LITO_COLANGIO)
  metricas$AUC[i]<-performance(pred,"auc")@y.values

  #NOTA: HACEMOS MATRIZ DE CONFUSION##
  matriz_confusion<-confusionMatrix(as.factor(predicciones_clase),
                                    reference=prueba$LITO_COLANGIO,
                                    positive="1")

  #NOTA:ahora se almacenan las metricas una por una
  metricas$SENSITIVITY[i] <- matriz_confusion$byClass[1]
  metricas$SPECIFICITY[i] <- matriz_confusion$byClass[2]
  metricas$POS.PRED.VALUE[i] <- matriz_confusion$byClass[3]
  metricas$NEG.PRED.VALUE[i] <- matriz_confusion$byClass[4]
  metricas$PRECISION[i] <- matriz_confusion$byClass[5]
  metricas$ACCURACY[i] <- matriz_confusion$overall[1]
  metricas$RVP[i] <- metricas$SENSITIVITY[i] / (1-metricas$SPECIFICITY[i])
  metricas$RVN[i] <- (1-metricas$SENSITIVITY[i]) / metricas$SPECIFICITY[i]
  #CALCULO DE ODDS RATIO
  metricas$ODDS.RATIO[i] <- metricas$RVP[i] / metricas$RVN[i]
}
metricas

mean(metricas$SENSITIVITY)
mean(metricas$SPECIFICITY)
mean(metricas$POS.PRED.VALUE)
mean(metricas$NEG.PRED.VALUE)
mean(metricas$ACCURACY)
mean(metricas$PRECISION)
mean(metricas$RVP)
mean(metricas$RVN)
mean(metricas$ODDS.RATIO)
mean(metricas$AUC)
promedio_AUC <- as.numeric(metricas$AUC,na.rm = TRUE)
mean(promedio_AUC)

#apply(metricas,2,mean,na.rm=TRUE)
##AHORA SE CALCULARIA CURVA ROC
ROCpred <- prediction(predicciones,etiquetas)
ROCperf <- performance(ROCpred, "tpr", "fpr")

plot(ROCperf, colorize=TRUE, print.cutoff.at=seq(0.1,by=0.1), avg="threshold", lwd=3)
plot(ROCperf, col="grey78", add=TRUE)
```

```
# Vamos a seleccionar un 70% de los datos para
# ajustar un modelo de regresión logística
# Las muestras de entrenamientos serán del mismo
# tamaño para que evitar posibles problemas de
# desbalance entre las clases.
# El 70% de 194 pacientes es 136 pacientes
# de los cuales serán seleccionados 68 de
# cada uno de los dos grupos
set.seed(14042025)
indice.lc.ausente <- which(DATOS$LITO_COLANGIO==0)
indice.lc.presente <- which(DATOS$LITO_COLANGIO==1)
posiciones <- c(sample(x=indice.lc.ausente, size=68),
                sample(x=indice.lc.presente, size=68))
entrenamiento <- DATOS[posiciones, ]
prueba <- DATOS[-posiciones, ]

###SE AJUSTA EL MODELO A LOS DATOS DE ENTRENAMIENTO###
modelo.0.e <- glm(LITO_COLANGIO ~ EDAD + SEXO + IMC + DM2 + HTA + TAB +
                  ALCOHOL + DX_PRE + RIESGO_ASGE + FA +
                  BT + ALT_TGP + AST_TGO + GGT + CD + LITO +
                  PANCREATITIS + COLANGITIS, data=entrenamiento, family="binomial")
summary(modelo.0.e)

modelo.1.e <- glm(LITO_COLANGIO ~ DM2 + HTA + DX_PRE + GGT + LITO,
                  data=entrenamiento, family="binomial")
summary(modelo.1.e)

### SE EVALUA EL MODELO CON LOS DATOS DE PRUEBA, EL MODELO EVALUA LA PROBABILIDAD DE QUE UN PACINETE TENGA LITO###
Se evalua el modelo con los datos de prueba, el
# modelo evalua la probabilidad de que un paciente
# tenga Lito
prediccion.prob.0 <- predict.glm(modelo.0.e, newdata=prueba,
                                type="response")

prediccion.prob.1 <- predict.glm(modelo.1.e, newdata=prueba,
                                type="response")

###SE REALIZAN PREDICCIONES EN BASE A UN VALOR UMBRAL ###

umbral <- seq(0.1, 0.9, by=0.01)
metricas <- matrix(0, nrow=length(umbral), ncol=10)

for(i in 1:length(umbral)) {
  prediccion <- ifelse(prediccion.prob.0 < umbral[i], 0, 1)
  resultado <- confusionMatrix(as.factor(prediccion),
                              reference=as.factor(prueba$LITO_COLANGIO),
                              positive='1')

  metricas[i, 1] <- resultado$overall[1]
  metricas[i, 2] <- resultado$byClass[1]
  metricas[i, 3] <- resultado$byClass[2]
  metricas[i, 4] <- resultado$byClass[3]
  metricas[i, 5] <- resultado$byClass[4]
  prediccion <- ifelse(prediccion.prob.1 < umbral[i], 0, 1)
  resultado <- confusionMatrix(as.factor(prediccion),
                              reference=as.factor(prueba$LITO_COLANGIO),
                              positive='1')

  metricas[i, 6] <- resultado$overall[1]
  metricas[i, 7] <- resultado$byClass[1]
  metricas[i, 8] <- resultado$byClass[2]
  metricas[i, 9] <- resultado$byClass[3]
  metricas[i, 10] <- resultado$byClass[4]
}
```



```

plot(umbral, metricas[,1], type="l", ylim=c(0,1),
     ylab = 'Sensibilidad y especificidad')
points(umbral, metricas[,2], col="green", type="l")
points(umbral, metricas[,3], col="red", type="l")
points(umbral, metricas[,6], lty=2, type="l")
points(umbral, metricas[,7], col="green", type="l", lty=2)
points(umbral, metricas[,8], col="red", type="l", lty=2)
abline(v=0.60)
abline(v=0.55)
abline(h=0.5)

plot(umbral, metricas[,4], type="l", col="blue", ylim=c(0,1),
     ylab = 'Valor predictivo positivo y valor predictivo negativo')
points(umbral, metricas[,5], col="red", type="l")
points(umbral, metricas[,9], col="blue", type="l", lty=2)
points(umbral, metricas[,10], col="red", type="l", lty=2)
abline(v=0.21)

#umbral 0.60
umbral_especifico <- 0.60

#POSICION SEGUN INDICE
indice <- which.min(abs(umbral - umbral_especifico))

#EXTRACCION DE VALORES
sensibilidad_en_060 <- metricas[indice, 7]
especificidad_en_060 <- metricas[indice, 8]

#RESULTADOS
cat("Para un umbral de:", umbral[indice], "\n")
cat("  - Sensibilidad:", round(sensibilidad_en_060, 4), "\n")
cat("  - Especificidad:", round(especificidad_en_060, 4), "\n")

#umbral 0.21
umbral_especifico <- 0.21

#POSICION SEGUN INDICE
indice <- which.min(abs(umbral - umbral_especifico))

#EXTRACCION DE VALORES
VPP_en_060 <- metricas[indice, 9]
VPN_en_060 <- metricas[indice, 10]

#RESULTADOS
cat("Para un umbral de:", umbral[indice], "\n")
cat("  - VPP:", round(VPP_en_060, 4), "\n")
cat("  - VPN:", round(VPN_en_060, 4), "\n")

```