

Resumen de trabajo de tesis

Propuesta de predictor de coledocolitiasis para pacientes con pancreatitis aguda de origen biliar

MARÍA JOSÉ VÁSQUEZ CASTRO*

GRADUADA DE MAGÍSTER EN BIOESTADÍSTICA, ESCUELA DE SALUD PÚBLICA
FACULTAD DE MEDICINA, UNIVERSIDAD DE CHILE

A nivel nacional la prevalencia de los casos de pancreatitis aguda (PA) es de $42,6 \times 100.000$ habitantes, la cual es alta y muy característica de la población latinoamericana. La génesis de este cuadro está ligada estrechamente con la coledocolitiasis, siendo éste el diagnóstico para cuando existe obstrucción del colédoco a causa de la litiasis biliar. Esto genera que el jugo pancreático no pueda ser liberado hacia el duodeno a causa de la obstrucción y, por lo tanto, se comienzan a acumular las enzimas dentro del órgano, dando inicio al cuadro inflamatorio agudo. Con el fin de prevenir nuevos eventos de PA se extrae la vesícula biliar mediante colecistectomía (CL), que es un procedimiento endoscópico. En ocasiones, la litiasis biliar no sigue su camino natural hacia el duodeno y queda obstruyendo permanentemente el colédoco, dando origen al cuadro de coledocolitiasis. La resolución requiere personal endoscopista durante la CL, pero no se tiene certeza sobre la presencia de litiasis en el colédoco previo a la cirugía. Es por esto que se vuelve interesante la construcción de un predictor de coledocolitiasis con el fin de mejorar la planificación quirúrgica y optimizar recursos [1,2].

La construcción de *scores* en el ámbito médico se ha visto influenciada por técnicas de *Machine Learning* (ML), tanto por métodos de clasificación como de regresión, y paramétricos como no paramétricos. Para el caso de esta investigación se hace útil trabajar con modelos de clasificación supervisados, en los cuales se entregan los valores de la variable respuesta para la construcción del *score*. Dentro de estos métodos se encuentra la regresión logística, los *K* vecinos más cercanos (KNN) y los árboles de decisión, los cuales ya han sido explorados como métodos de clasificación en investigaciones previas, esto para determinar heridas malignas de piel, pronóstico de Fibrosis Quística (FQ) y eventos de hipotensión en estadias en UCI [3–5].

A partir de una base de datos de 52 sujetos que acudieron por PA al Hospital del Salvador en los años 2016 y 2017, que contiene tanto las características de los individuos como los exámenes de laboratorio e imagenológicos, se construyeron distintos *scores* con las técnicas de ML nombradas anteriormente. La variable respuesta correspondió a la necesidad de colangiopan-

creatografía retrógrada endoscópica (ERCP por sus siglas en inglés), técnica que extrae litiasis desde el colédoco por medio de endoscopia, lo que indirectamente indica que el paciente presentaba coledocolitiasis. Para seleccionar el modelo que entrega la mejor precisión en la predicción de coledocolitiasis, se compararon entre sí mediante precisión, sensibilidad, especificidad, valor predictivo positivo y negativo, y porcentaje de falsos negativos. Los valores de estos indicadores fueron calculados mediante la técnica de *K-fold Cross Validation*, en donde se compara *K* veces la base de entrenamiento con la base de prueba. El análisis comparativo arrojó que la regresión logística es el modelo que entrega la mejor predicción, presentando el mayor valor de precisión, especificidad y valor predictivo positivo [6].

Bajo el contexto de la práctica clínica, el método de árboles de decisión entrega una herramienta útil y práctica para poder predecir la presencia de coledocolitiasis, ya que el algoritmo construido permite una interpretación rápida y sencilla de las variables predictoras. Sin embargo, en esta investigación este método de clasificación fue muy sensible al tamaño muestral y el método se sobreajustó, entregando valores idénticos en todos los tipos de árboles construidos. En relación al modelo elegido, éste es más sencillo de utilizar que el método de KNN, y además incluye menos costos de implementación.

El modelo de regresión logística contiene variables que fueron seleccionadas mediante el método *Backward Stepwise*, siendo algunas de éstas categorizadas por medio de criterio experto o por curva ROC y, a partir de aquello, se creó el siguiente predictor de la probabilidad de requerir ERCP:

$$P(ERCP) = \frac{e^{-4,3+3,7CVB.TC-1,5BT+1,9BD+2,8AST}}{1 + e^{-4,3+3,7CVB.TC-1,5BT+1,9BD+2,8AST}},$$

donde

CVB.TC: calibre de la vía biliar, en milímetros, medido en tomografía computarizada y trabajada de manera dicotómica, asumiendo un valor igual a 0 si es inferior o igual a 7 mm, o un valor igual a 1 en caso contrario (> 7 mm);

BT: bilirrubina total, medida en mg/dL;

*Escuela de Kinesiología, Facultad de Medicina, Universidad de Valparaíso, maria.vasquez@uv.cl.

BD: bilirrubina directa, medida en mg/dL;
 AST: aspartato aminotransferasa, medido en U/L, dicotomizado de modo que asuma un valor igual a 0 si es inferior a 270 U/L, o un valor igual a 1 en caso contrario (≥ 270 U/L).

Si el valor de la probabilidad de ERCP es superior a 0,45, se predice que el paciente tiene coledocolitiasis, por lo que se sugiere planificar la cirugía contando con las condiciones necesarias para realizar un ERCP.

A causa de trabajar con un tamaño muestral pequeño, se decidió poner a prueba el modelo en una base de mayor tamaño. Esto se logró por medio del método *Bootstrap*, correspondiente a una técnica de remuestreo con reposición que permitió crear 3 bases de datos con tamaños de 102, 510 y 1020 observaciones, respectivamente. El modelo fue puesto a prueba y entregó una precisión mayor al 90 % en las 3 bases simuladas, por lo que podría ser una buena herramienta a utilizar en pacientes con PA sometidos a CL (ver Tabla).

Tabla: Odds Ratio estimado para la necesidad de ERCP con distintos tamaños muestrales obtenidos mediante remuestreo *Bootstrap*.

Tamaño muestral	Variables predictoras				AUC
	CVB.TC	BT	BD	AST	
102	47,8***	0,4	4,2*	32,9***	0,908
510	51,0***	0,2***	8,9***	20,1***	0,935
1020	32,8***	0,2***	7,2***	14,7***	0,922

p-value: *: $< 0,1$; **: $< 0,05$; ***: $< 0,01$.

CVB.TC: Calibre de la vía biliar medida en tomografía computarizada;

BT: Bilirrubina total; BD: Bilirrubina directa; AST: Aspartato aminotransferasa.

AUC: Área bajo la curva (*area under curve*).

Finalmente, el predictor propuesto contiene variables que previamente se ha descrito que se relacionan con la presencia de coledocolitiasis, tanto en pacientes pediátricos como en pacientes adultos, lo que respalda las variables seleccionadas para la construcción. Junto con lo anterior, el modelo seleccionado también ha sido ampliamente utilizado para la construcción de predictores, obteniendo buenos resultados en su uso, lo que coincide con lo plasmado en esta tesis de máster.

Información adicional

Directora: Prof. Andrea Canals C. Programa de Bioestadística, Escuela de Salud Pública, Universidad de Chile.

Codirector: Prof. Francisco Riquelme M. Servicio de Cirugía, Hospital del Salvador.

Fecha de la graduación: 7 de agosto de 2024.

Referencias

- [1] Berger Z, Mancilla C, Tobar E, *et al.* (2020). Acute pancreatitis in Chile: A multicenter study on epidemiology, etiology and clinical outcome. Retrospective analysis of clinical files. *Pancreatology: official journal of the International Association of Pancreatology (IAP)*, 20(4), 637-643. [10.1016/j.pan.2020.04.016](https://doi.org/10.1016/j.pan.2020.04.016)
- [2] Rosing DK, de Virgilio C, Yaghoobian A, *et al.* (2007). Early cholecystectomy for mild to moderate gallstone pancreatitis shortens hospital stay. *Journal of the American College of Surgeons*, 205(6), 762-766. [10.1016/j.jamcollsurg.2007.06.291](https://doi.org/10.1016/j.jamcollsurg.2007.06.291)
- [3] Hatem MQ. (2022). Skin lesion classification system using a K-nearest neighbor algorithm. *Visual computing for industry, biomedicine, and art*, 5(1), 7. [10.1186/s42492-022-00103-6](https://doi.org/10.1186/s42492-022-00103-6)
- [4] McCarthy C, Dimitrov BD, Meurling IJ, *et al.* (2013). The CF-ABLE score: a novel clinical prediction rule for prognosis in patients with cystic fibrosis. *Chest*, 143(5), 1358-1364. [10.1378/chest.12-2022](https://doi.org/10.1378/chest.12-2022)
- [5] Moghadam MC, Abad EMK, Bagherzadeh N, *et al.* (2020). A machine-learning approach to predicting hypotensive events in ICU settings. *Computers in Biology and Medicine*, 118. [10.1016/j.combiomed.2020.103626](https://doi.org/10.1016/j.combiomed.2020.103626)
- [6] Vabalas A, Gowen E, Poliakoff E, *et al.* (2019). Machine learning algorithm validation with a limited sample size. *PLoS One*, 14(11):e0224365. [10.1371/journal.pone.0224365](https://doi.org/10.1371/journal.pone.0224365)