

Sr. Director

He leído el artículo de Gasque y colaboradores¹ que presenta, en mi opinión, defectos metodológicos que atentan contra su reproducibilidad y consistencia interna, por lo que merecen ser discutidos:

1. No fue calculado el poder estadístico de la muestra. La incidencia de la apendicitis aguda no ofrece motivos para resignarse a un tamaño muestral exiguo y privar a los lectores de un valioso elemento para interpretar los resultados.
2. Se compararon las proporciones de diagnósticos positivos entre hombres y mujeres y en diferentes intervalos de edad pero sin inferencia estadística, por lo que el lector no sabe si los resultados pueden generalizarse.
3. La mención a la falta de significación estadística ($p = 0,11$) en los promedios de duración del cuadro clínico demuestra que, para esa variable, se utilizó una prueba de contrastación de hipótesis pero no se menciona cuál, ni el nivel de significación establecido, ni el programa informático con el que se realizaron los cálculos.
4. Los autores consignaron el promedio global de duración del cuadro clínico con su desvío estándar y los de los grupos con apendicitis y sin ella, apenas una línea por debajo, sin desvío estándar (?)
5. Los valores de cribaje fueron comunicados sin intervalos de confianza. Debido a lo reducido de la muestra, en especial del grupo sin apendicitis aguda, el correspondiente a la especificidad es de una amplitud tal (IC 95% = 0,2 -0,78) que viste de incertidumbre al 0,4 (40%) calculado por los autores.
6. La muestra es poco representativa de la población para la que está destinada la escala RIPASA, ya que incluye más de 85% de diagnósticos positivos, proporción que magnifica la incidencia observada en pacientes que consultan a servicios de emergencias con "síndrome de fosa ilíaca derecha" (Chong comunicó 52,60% en su estudio prospectivo)². Esto distorsiona los valores predictivos de la prueba, esenciales para estimar su rendimiento en el escenario asistencial.
7. Newman y col.³, entre muchos otros, proscriben la estimación de los valores predictivos en muestras no consecutivas.
8. La evaluación de registros de pacientes con diagnóstico final conocido y por un solo observador implica un potencial sesgo de interpretación. Para menguar el riesgo se aconseja emplear al menos dos evaluadores independientes no avisados de los objetivos de la investigación (enmascaramiento o ceguera) y calcular la concordancia interobservador⁴.
9. La conclusión es propia de una revisión bibliográfica y no del análisis de datos primarios, ya que se funda en estudios de otros autores.

La valoración del rendimiento de una escala para el diagnóstico de apendicitis aguda ofrece facilidades para la investigación que, lamentablemente, no fueron aprovechadas:

1. Existen datos de estudios anteriores para calcular, con buena aproximación, la cantidad de pacientes que permita obtener estimaciones por intervalo apropiadas.
2. La patología es frecuente, lo que permitiría alcanzar el tamaño muestral requerido en un tiempo de reclutamiento breve.
3. No es necesario un seguimiento prolongado, lo que garantiza pocos abandonos y facilita la realización de estudios prospectivos.

Daniel E. Tripoloni

dtripoloni@hotmail.com

División "A" de Cirugía General. Hospital José María Ramos Mejía. Ciudad Autónoma de Buenos Aires. Argentina

■ ENGLISH VERSION

Dear Sir,

I have read the article by Gasque et al.¹ which, in my opinion, has methodological flaws that affect its reproducibility and internal consistency which merit discussion:

1. The statistical power of the sample was not calculated. The incidence of acute appendicitis does not provide a reason for a small sample size and deprives readers of a valuable element for the interpretation of the results.
2. The authors compared the proportions of positive diagnoses between men and women and at different age intervals but without statistical inference, so the reader does not know if the results can be generalized.
3. The lack of statistical significance ($p = 0.11$) in the mean duration of symptoms shows that a hypothesis test was used for this variable, but the test used, the level of significance established and the software used to perform the calculations are not mentioned.
4. The authors reported the mean overall duration of symptoms with its standard deviation, while the standard deviation of duration of symptoms in the groups with and without appendicitis, which was just one point lower, is not mentioned.
5. The screening values were reported without the corresponding confidence intervals. Due to the small sample size, especially in the group without acute appendicitis, the confidence interval of specificity

is so wide (95% CI = 0.2 -0.78) that the authors calculated an uncertainty of 0.4 (40%).

6. The sample is poorly representative of the target population of the RIPASA score, since it includes more than 85% of positive diagnoses, a proportion that magnifies the incidence observed in patients visiting the emergency departments with "right iliac fossa pain" (Chong reported 52.60% in his prospective study)², producing a distortion of the predictive values of the test, which are essential to estimate its performance in the healthcare scenario.
7. Newman et al.³, among many other authors, ban the estimation of predictive values in non-consecutive samples.
8. The evaluation of patient records with a known final diagnosis by a single observer involves potential interpretation bias. To reduce the risk, a minimum of two independent blinded assessors are recommended with estimation of interobserver agreement⁴.
9. The conclusion is typical of a review of the literature rather than an analysis of primary data, as it is based on studies by other authors.

The assessment of the performance of a score for the diagnosis of acute appendicitis offers opportunities for research that, unfortunately, were not considered:

1. There are data from previous studies to calculate, with reasonable approximation, the number of patients to obtain appropriate estimates per interval.
2. The disease is common, which would allow to reach the required sample size in a short recruitment period.
3. There is no need for prolonged follow-up, which ensures few dropouts and facilitates prospective studies.

Daniel E. Tripoloni
dtripoloni@hotmail.com

División "A" de Cirugía General. Hospital José María Ramos Mejía. Ciudad Autónoma de Buenos Aires. Argentina

Referencias bibliográficas | Reference

1. Gasque RA, Moreno WA, Vigilante GE. Correlación diagnóstica de la escala RIPASA en pacientes intervenidos quirúrgicamente por apendicitis aguda. Rev Argent Cirug. 2020; 112:178-84.
2. Chong CF, Thien A, Mackie AJ, Tin AS, Tripathi S, Ahmad MA, et al. Comparison of RIPASA and Alvarado scores for the diagnosis of acute appendicitis. Singapore Med J. 2011;52:340-5.
3. Newman TB, Browner WS, Cummings SR, Hulley SB. Diseño de estudios de pruebas médicas. En: Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Diseño de investigaciones clínicas. 4ª ed. Barcelona: Wolters Kluwer; 2014. Pp.171-91.
4. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960; 20:37-46.

Respuesta de los autores

Sr. Director:

Agradecemos al Dr. Tripoloni por haberse tomado el tiempo de leer nuestro artículo "Correlación diagnóstica de la escala RIPASA en pacientes intervenidos quirúrgicamente por apendicitis aguda" y habernos enviado sus valiosas apreciaciones. En respuesta a sus correcciones:

- Si bien es cierto que no fue calculado el tamaño muestral, decidimos incluir la mayor cantidad de pacientes en la muestra debido a que nuestro centro atiende principalmente adultos mayores (con un promedio de 63 años); por ende, no contamos con un número elevado de casos de apendicitis aguda.
- Para todas las pruebas estadísticas se utilizó una $p \leq 0,05$ (IC 95%) como estadísticamente significativa y se calcularon con las plataformas MedCalc v. 19.1.3[®] y Microsoft Excel 2016[®]. Utilizamos la prueba "t" de Student para contrastar la duración del cuadro clínico entre los pacientes con apendicitis aguda y sin ella en su estudio histopatológico.
- La desviación estándar de la duración del cuadro clínico para pacientes con apendicitis aguda fue de $8,81 \pm 3,41$ horas y para aquellos sin apendicitis aguda fue de $7,95 \pm 4,32$ horas.

- Una de las principales falencias de este estudio es su naturaleza retrospectiva y el hecho de que nos hayamos basado en las muestras de anatomía patológica para el diagnóstico de certeza. Igualmente, este trabajo fue presentado en el 89° Congreso Argentino de Cirugía y una de las críticas del jurado fue que la escala no fue aplicada de manera regular a todos los pacientes que se presentaron por guardia con "síndrome de fosa ilíaca derecha" como usted remarca en el punto 6; por ende, queda trunca su potencial función discriminatoria.
 - Las historias clínicas electrónicas fueron analizadas por el autor principal del trabajo. Consideramos que el principal sesgo fue la subjetividad de cada cirujano tratante a la hora de la evaluación inicial del paciente. Este podría reducirse –como dijo usted– utilizando más de un evaluador y de manera prospectiva.
 - La conclusión responde a los resultados y objetivos del estudio y no así a la discusión como usted propone.
- Nuevamente, reiteramos la necesidad de realizar este estudio de manera prospectiva y con un tamaño muestral adecuado para la patología a fin de validar su utilidad en nuestro medio.

Rodrigo A. Gasque
Walter A. Moreno
Gabriel E. Vigilante