

Sobre la calibración de los sistemas de predicción de muerte en los pacientes críticos

Sr. Director:

He leído con interés los dos estudios publicados recientemente en *MEDICINA INTENSIVA* por Domínguez et al sobre la efectividad y la eficiencia en las unidades de cuidados intensivos (UCI)¹ y sobre la reproducibilidad de las distintas puntuaciones de ajuste de riesgo (APACHE II, APACHE III y SAPS II)². Ambos casos ponen de manifiesto el alto nivel de la revista que usted dirige y de los investigadores que en ella publican. Quisiera, sin embargo, puntualizar algunas cuestiones estadísticas que, en mi opinión, no han sido correctamente expresadas por los autores.

En el primer estudio¹, se dice que la calibración del sistema APACHE III adaptado para España es buena

para predecir la mortalidad hospitalaria, pero se cita una *p* del estadístico de Hosmer y Lemeshow cercana a la significación estadística ($p < 0,1$); dado que cuanto menor sea el valor de *p*, peor es la calibración, una $p < 0,1$ no indica una buena calibración, sino más bien lo contrario³. Los autores deberían haber expresado el valor exacto de la *p*; sin embargo, éste se puede calcular con los datos que se aportan en el estudio, ya que sí se da el valor de la χ^2 , que es de 14,95, y asumiendo que el número de grados de libertad para el estadístico de Hosmer y Lemeshow es habitualmente de 8, el valor de la *p* sería de 0,06, lo que la coloca al borde de la significación estadística, pero sin alcanzarla. En cualquier caso, aunque no se pueda afirmar formalmente que el APACHE III adaptado para España no tiene buena calibración, tampoco se puede decir que sí la tenga, como hacen los autores.

En su estudio sobre la reproducibilidad de las puntuaciones APACHE II, APACHE III y SAPS II², los autores expresan, en la tabla 2, los resultados numéricos de la prueba de Hosmer-Lemeshow para evaluar la calibración de los citados sistemas (el grado en que predicen correctamente la mortalidad real), pero no dan ninguna explicación al respecto, por lo que puede pasar inadvertida, para la mayoría de los lectores, su conclusión práctica: que ninguno de los tres sistemas de puntuación tuvo una buena calibración en la muestra estudiada. Éste es un dato preocupante, pues cuestiona la utilidad de los sistemas de puntuación empleados habitualmente para la predicción pronóstica de nuestros pacientes.

Por último, en la figura 1 del estudio sobre la reproducibilidad² parece apreciarse que el grado de

correlación en el riesgo de muerte estimado por dos grupos distintos de intensivistas presenta una mayor dispersión cuando el riesgo es mayor, mientras que el acuerdo parece mejorar cuando el riesgo de muerte es menor. Los autores no hacen una prueba formal de esta apreciación, y no informan a los lectores sobre ella, a pesar de tratarse de un hallazgo potencialmente importante.

En conclusión, los autores deberían hacer un mayor esfuerzo en la redacción de los manuscritos más técnicos, en el sentido de hacerlos más comprensibles para el grueso de los lectores, destacando más sus implicaciones prácticas. De esta manera, su publicación ganaría en interés e impacto, como se merecen el esfuerzo invertido y la calidad de la investigación realizada.

E. PALENCIA HERREJÓN
*Servicio de Medicina Intensiva.
Hospital Infanta Leonor. Madrid. España.*

BIBLIOGRAFÍA

1. Domínguez L, Enríquez P, Álvarez P, De Frutos M, Sagredo V, Domínguez A, et al. Mortalidad y estancia hospitalaria ajustada por gravedad como indicadores de efectividad y eficiencia de la atención de pacientes en Unidades de Cuidados Intensivos. *Med Intensiva*. 2008;32:8-14.
2. Domínguez L, Enríquez P, Álvarez P, De Frutos M, Sagredo V, López-Messa J, et al. Evaluación de la reproducibilidad de la recogida de datos para el APACHE II, APACHE III adaptado para España y SAPS II en 9 Unidades de Cuidados Intensivos en España. *Med Intensiva*. 2008;32:15-22.
3. Hosmer DW Jr, Lemeshow S. *Applied logistic regression*. 2.a ed. New York: Wiley; 2000. p. 147-56.