

Sobre la reproducibilidad y la efectividad del APACHE II, el APACHE III adaptado para España y el SAPS II en 9 unidades de cuidados intensivos en España

Sr. Director:

Agradecemos los interesantes comentarios del Dr. Eduardo Palencia en relación con nuestros artículos sobre la evaluación de la efectividad y la eficiencia¹ y de la reproducibilidad² de distintos modelos pronósticos utilizados en 9 unidades de cuidados intensivos en España. Sin embargo, quisiéramos hacer las siguientes observaciones:

1. En relación con el primer estudio¹, el Dr. Palencia, tras el cálculo de la *p* del test C de bondad de ajuste Hosmer-Lemeshow, afirma que no se puede decir que la calibración sea buena al ser ésta de 0,06. Al ser el

test de Hosmer-Lemeshow una prueba que compara la distribución de dos curvas de riesgo, la que predice nuestra muestra y la que se observa, podemos decir, tal y como describen Hosmer y Lemeshow³, que ambas distribuciones no son diferentes y, por lo tanto, la calibración es buena. En cualquier caso, el valor de la *p* no se debe interpretar como si hubiera una relación del tipo «a más *p* mejor calibración, a menos *p* peor calibración». Por otro lado, si el fin de ese artículo hubiera sido el estudio de la calibración del modelo, ésta habría quedado perfectamente descrita con la observación de la curva de calibración.

2. En cuanto al estudio sobre la reproducibilidad², efectivamente los modelos APACHE II y SAPS II no calibran correctamente en nuestra muestra. Sí calibran correctamente el APACHE III (*p* < 0,1). Sin embargo, puesto que el estudio pretende hacer hincapié en uno de los numerosos problemas que pueden justificar que los modelos no calibren correctamente, como es la reproducibilidad en la recogida de los datos, el resto de la metodología del estudio está centrado en este punto.

Por último, efectivamente la dispersión de los datos es mayor en los deciles altos de riesgo. Probablemente, el factor determinante es el menor número de pacientes en estos deciles de riesgo. Por ello, consideramos de poca utilidad estudiar la correlación por niveles de riesgo, ya que en todo caso se habrían producido estimadores muy inestables y, por ello, de poca utilidad.

En definitiva, efectivamente la calibración y el estudio de los modelos pronósticos son un tema complejo, que requiere un profundo conocimiento de la metodología a emplear y que cuando se hace correctamente deja muchas preguntas por responder y genera muchas dudas sobre su utilidad.

L. DOMÍNGUEZ, P. ENRÍQUEZ Y J. BLANCO

Unidad de Cuidados Intensivos.

Hospital Universitario Río Hortega. Valladolid. 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, Lemeshow S. *Applied logistic regression*. New York: John Wiley & sons; 1989. p. 135-86.