

Cartas al director

Validación de los Mortality Probability Models II comparados con las predicciones de mortalidad hospitalaria de APACHE II y SAPS II

Sr. Director:

Hemos leído con atención el reciente artículo de Serrano Henández et al¹ en el que se comparan varios modelos predictivos de mortalidad en pacientes críticos. En primer lugar hemos de reconocer el interesante enfoque del problema y la utilización de herramientas predictivas que, a pesar de su amplio reconocimiento y su vigencia, no son usadas en todo su alcance (especialmente el APACHE II) de manera habitual por los intensivistas de nuestro país.

Sin embargo, queremos realizar alguna puntualización sobre el acercamiento metodológico al problema que, a nuestro juicio, requiere alguna aclaración. La finalidad reconocida del estudio es "comparar la certeza de las predicciones de mortalidad hospitalaria de (los modelos estudiados)", y para ello se analiza la calibración y el poder discriminativo en el punto de corte del 50% de los distintos modelos.

Para comparar la calibración se utiliza como herramienta estadística la prueba de bondad del ajuste de Lemeshow-Hosmer, y en el apartado de Métodos se reconoce que el valor del estadístico χ^2 depende del tamaño de la muestra². Pero después se afirma que el coeficiente de significación p no depende de dicho tamaño, lo cual es cuando menos cuestionable, pues el grado de significación p proviene directamente del valor de χ^2 y por tanto, ante una misma distribución, puede variar con el tamaño de la muestra. Además, posteriormente se atribuye superior calibración al modelo predictivo MPM II-0 respecto a los obtenidos a las 24 horas, en base a que su valor de p es menor que los otros (todos ellos no significativos). Extraer este tipo de conclusiones de esta prueba de significación es arriesgado tanto por la definición en sí del contraste de hipótesis como porque la finalidad de la prueba de Lemeshow-Hosmer está en definir si existen diferencias significativas en la calibración de un modelo al compararlo con una distribución teórica, y no la comparación de calibraciones entre diferentes modelos³. De hecho, todos los modelos analizados muestran valores de p no significativos, por lo que resulta difícil afirmar la superioridad de unos frente a otros.

En cuanto al análisis de la discriminación, la utilización de curvas ROC y el cálculo del área bajo la curva constituyen una magnífica aproximación para la comparación de modelos predictivos, pero echamos de menos la utilización de los errores estándar y la construcción de los intervalos de confianza al 95% de dichas áreas, así como la aplicación de alguna prueba estadística de comparación como la propuesta por Hanley⁴ en su artículo original. Ello ofrecería otro tipo de argumentos para establecer el mejor poder discriminativo de alguno de los modelos predictivos.

Por último, sería de interés conocer los índices de mortalidad estandarizado y sus intervalos de confianza al 95% para cada modelo predictivo, que constituyen otro medio validado, y de muy intuitiva interpretación para analizar la adecuación entre la mortalidad real y la esperada.

Consideramos por tanto que no se ofrecen datos suficientes para poder afirmar la superioridad de un modelo predictivo frente a otro, bien porque son de hecho superponibles o porque el tamaño de la muestra no es suficientemente grande.

BIBLIOGRAFÍA

1. Serrano Hemández N, García de Lorenzo y Mateos A, Mora Quintero UL, Fedriani Gorría J. Validación de los Mortality Probability Models II al ingreso (MPM II-0), a las 24 horas (MPM II-24) y a las 48 horas (MPM II-48) comparados con las predicciones de mortalidad hospitalaria de APACHE II y SAPS II realizadas en los días 1 y 2 de estancia en UCI. *Med Intensiva* 2000; 24: 49-60.
2. Silva Ayçaguer LC. Bondad del ajuste. En: Silva Ayçaguer LC, ed. *Excursión a la regresión logística en ciencias de la salud*. Madrid, Ediciones Díaz de Santos, 1995; 63-68.
3. Silva Ayçaguer LC ¿Qué significan las pruebas de significación? En: Silva Ayçaguer LC, ed. *Cultura estadística e investigación científica en el campo de la salud: una mirada crítica*. Madrid: Ediciones Díaz de Santos, 1997; 133-159.
4. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29-36.

J. L. GARCÍA GARMENDIA,
F. J. JIMÉNEZ JIMÉNEZ,
J. GARNACHO MONTERO y C. ORTIZ LEYBA

Servicio de Cuidados Críticos y Urgencias. UCI, Hospital General. Hospitales Universitarios Virgen del Rocío. Sevilla.