

Hospital Militar Central "Dr. Luis Díaz Soto"

## VALIDACIÓN DEL SISTEMA PREDICTIVO APACHE II EN UN GRUPO DE PACIENTES INGRESADOS EN LA UNIDAD DE CUIDADOS INTENSIVOS

*Dr. Armando Padrón Sánchez<sup>1</sup> Dr. Jorge Luis Ayala Pérez<sup>1</sup> Dr. Mario Santiago Puga Torres<sup>1</sup> My. Tammy Alonso Díaz<sup>2</sup> Cap. Tania Salazar González<sup>2</sup> y Tte. Navio Andrés Quiñones Zamora<sup>3</sup>*

### RESUMEN

Los modelos predictivos se han convertido en una herramienta eficaz siempre que prime el sentido común y la experiencia en su uso. Se realizó un estudio prospectivo longitudinal de tipo cohorte e inferencial en la Unidad de Cuidados Intensivos de adultos del Hospital Militar Central "Dr. Luis Díaz Soto". Las variables objeto de estudio son las comprendidas en el sistema de Evaluación del Estado Fisiológico Agudo y Crónico (APACHE II), calculándose su valor, razón de predicción-muerte y la razón de predicción-muerte ajustada, tanto a las 24 como a las 48 h. La calibración fue comparada utilizando los estadísticos chi cuadrado, y la discriminación comparando el área bajo la curva de probabilidad (ROC). Valores superiores a 30 puntos demostraron más del 80 % de probabilidad de muerte; la cifra de 40 puntos o más fue 100 % fatal en el estudio; el sistema APACHE II demostró mejor calibración y discriminación a las 24 h que a las 48 h ( $p < 0,05$ ).

*DeCS:* PREDICCIÓN/métodos; APACHE; CUIDADOS INTENSIVOS; ESTUDIOS PROSPECTIVOS.

Los modelos predictivos se han convertido en una herramienta eficaz, siempre que prime el sentido común y la experiencia en su uso, para el médico dedicado al cuidado del enfermo crítico, respaldo de acciones éticas y legales, han sido elementos importantes a tener en cuenta a la hora de privar de una terapéutica o proceder a al-

gún enfermo. Como sucede con todas las decisiones que tienen un impacto sobre los cuidados del paciente, el director-médico debe ponderarse del conjunto de conocimientos médicos disponibles, los deseos de los pacientes, familiares y médicos, y la probabilidad de que los cuidados intensivos beneficien al paciente o no. A veces

<sup>1</sup> Especialista de I Grado en Medicina Interna. Instructor.

<sup>2</sup> Especialista de I Grado en Medicina Interna.

<sup>3</sup> Especialista I Grado en Medicina General Integral Militar.

estas decisiones incluirán solo un juicio médico; otras veces, la elección reflejará una perspectiva ética, legal o filosófica<sup>1-5</sup>

En muchos servicios, el número de cama en las unidades de cuidados intensivos (UCI) al paciente crítico es limitado, y con suma frecuencia los pacientes que llegan con diversos grados de morbilidad se comparan para determinar cuales se pueden tratar mejor en estas salas. Todas las escalas predictivas de uso en cuidados intensivos no son más que sistemas de valores numéricos para describir la posible evolución de la enfermedad del paciente.<sup>6-8</sup> El primer modelo que sobrevivió a la crítica y vio la luz de la publicación médica fue el propuesto por *William Knaus*<sup>1</sup> en 1981, que consistió en un conjunto de modelos a los cuales se le denominó de forma genérica Acute Physiology and Chronic Health Evaluation conocido por las siglas de APACHE, el cual viene siendo diseñado y actualizado por *Knaus* y otros desde finales de la década de los 70.

## MÉTODOS

Se realizó un estudio prospectivo longitudinal de tipo cohorte e inferencial en la UCI de adultos del Hospital Militar Central "Dr. Luis Díaz Soto".

Para la selección de la muestra se utilizó los criterios siguientes:

*Criterios de inclusión.* a) Todo paciente que ingrese en el servicio de la UCI Polivalente. b) Síndrome coronario agudo con clasificación de Killip y Kimball III-IV o reanimación cardiopulmonar exitosa.

Para el cálculo de la muestra se estimó la aproximación del resultado de la media de los ingresos mensuales en la UCI (que no sean por causa coronaria aguda) y se multiplicó por el tiempo que durará la

recolecta de los datos que es aproximadamente 5 meses mediante la fórmula siguiente:  $N = (Z - N_1) \times t$  donde  $N$ , es el valor de la muestra estimada,  $Z$  el número de casos que ingresan mensualmente,  $N_1$  el número de casos con diagnóstico de síndrome coronario agudo y  $t$  el tiempo que duró la fase investigativa. Los cálculos ofrecieron un tamaño muestral de 116 pacientes. Las variables objeto de estudio son las comprendidas por el sistema APACHE II que cuenta con 12 y un número de 6 restantes que se escogieron según el criterio y experiencia del tutor y demás profesores, lo que constituyen 18 variables en el estudio. Se confeccionó una planilla individual para cada caso donde se recogieron los datos, la cual fue llenada y recolectada por el autor y el tutor. La recolección de los datos se efectuó a las 24 y 48 h de ingreso en UCI de adultos, para su comparación. A cada paciente que se escogió para el estudio se le llenó la planilla a las 24 y 48 h, calculándose el valor del APACHE II, la razón de predicción muerte y la razón de predicción-muerte ajustada por diagnóstico al ingreso, según los coeficientes de ajuste calculados por *Knaus* y otros para el APACHE II. Se confeccionó una base de datos con el paquete estadístico computadorizado SPSS y se utilizó el cálculo de la media y la desviación así como el estadígrafo  $t$  de Student, y la regresión lineal. Se empleo el 95 % de confianza para el estudio de las relaciones entre las variables. Dado que no existe una única medida estándar para describir la bondad del ajuste o capacidad predictiva de estos modelos de regresión logística,<sup>9,10</sup> la calibración fue comparada utilizando los estadísticos  $X^2$  propuestos por *Lemeshow* y *Hosmer*,<sup>9</sup> y la discriminación comparando el área bajo la curva ROC (características operativas del receptor) de cada modelo. Las probabilidades individuales de mortalidad obtenidas por los 2 modelos fueron

ordenadas según valores crecientes y agrupadas en estratos (en este estudio se adoptaron 10 estratos) con un número similar de pacientes en cada uno. El número de fallecimientos esperados en cada estrato corresponde a la suma de las probabilidades individuales de muertes en los pacientes en dicho estrato. El número de supervivientes esperados se calcula por diferencia. Para la comparación de los resultados entre las 24 y 48 h se optó por la comparación del valor más alto del coeficiente de significación  $p$ . Fueron comparadas la sensibilidad, especificidad y cociente de clasificación correcta total (CCT) de los distintos índices predictivos. Se comparó la calibración y la discriminación de la predicción del modelo APACHE II a las 24 y 48 h de ingreso en la UCI.

## RESULTADOS

Se estudiaron 116 pacientes críticos, 73 hombres y 43 mujeres. La edad promedio en la muestra estudiada fue de 56,45 años con un mínimo de 19 años y un máximo de 90 años, la desviación estándar (DE) de 17,71 años. Los antecedentes patológicos personales fueron hipertensión arterial 60,45 %, cardiopatía isquémica 26,02 %, diabetes mellitus 5,48 %, enfermedad pulmonar obstructiva crónica 4,24 %, insuficiencia renal crónica 1,16 % y otras 2,65 %. El 4,12 % de los encuestados recibió maniobras de reanimación cardiopulmonar antes de ser admitidos en la UCI. Los padecimientos crónicos resultaron la hipertensión arterial en primer lugar, que aumenta su frecuencia en muestras de enfermos críti-

cos, igual ocurre con la cardiopatía isquémica, la diabetes mellitus y demás enfermedades crónicas. Al utilizar una prueba  $t$  para el análisis univariado mostró relaciones significativas ( $p < 0,05$ ) entre la muerte y las variables: la edad, la frecuencia cardíaca, la tensión arterial media, la  $PO_2$ , la hemoglobina, la glicemia, la creatinina, el nivel de Glasgow y la puntuación de APACHE II.

La prueba de chi cuadrado mostró una alta significación ( $p < 0,001$ ) entre el fallecimiento y la condición de haber recibido reanimación cardiopulmonar antes de su admisión en UCI. La distribución de la puntuación de APACHE II en la muestra se observa en la figura 1. Para determinar la relación entre la puntuación de APACHE II con respecto a las variables: edad y mortalidad, se realizó un modelo gráfico de regresión lineal (figs. 2 y 3) donde se puede observar una dependencia con evidente proporcionalidad directa entre la edad, mortalidad y la puntuación de APACHE II. La figura 2 pilota un modelo de regresión entre la edad y la puntuación del APACHE II; se demuestra que a medida que la edad avanza la puntuación del modelo se eleva. Al distribuir la muestra usando el modelo del APACHE II y calculando la razón de predicción de muerte se observó que valores entre 30 y 35 puntos tenían el 80 % de probabilidad de fallecimiento y aumentaba hasta el 90 % cuando alcanzaba los 40 puntos, valores superiores fueron 100 % fatales. En el estudio aproximadamente el 60 % de la muestra demostró valores de APACHE inferiores a 30 puntos.

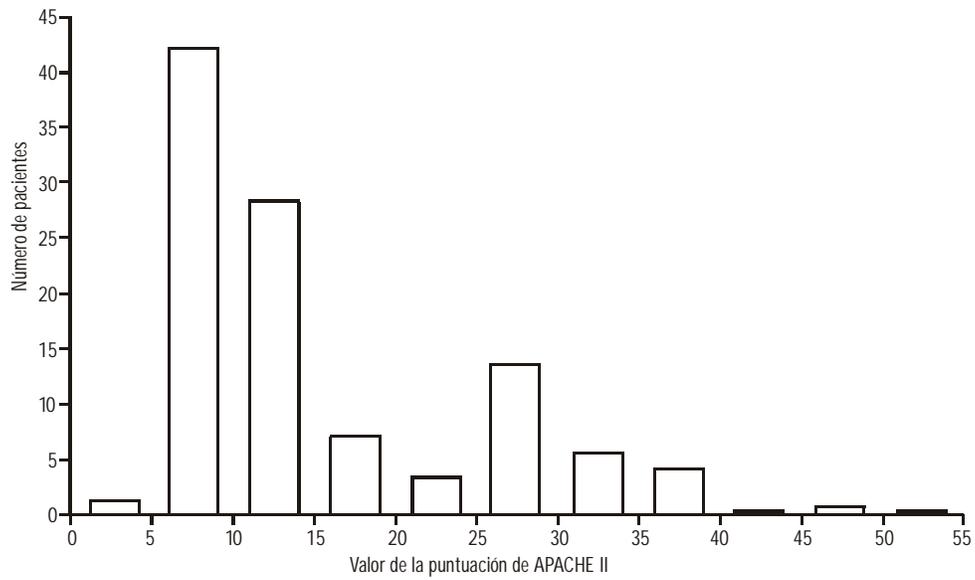


FIG. 1. Distribución de la muestra según puntuación de APACHE II.

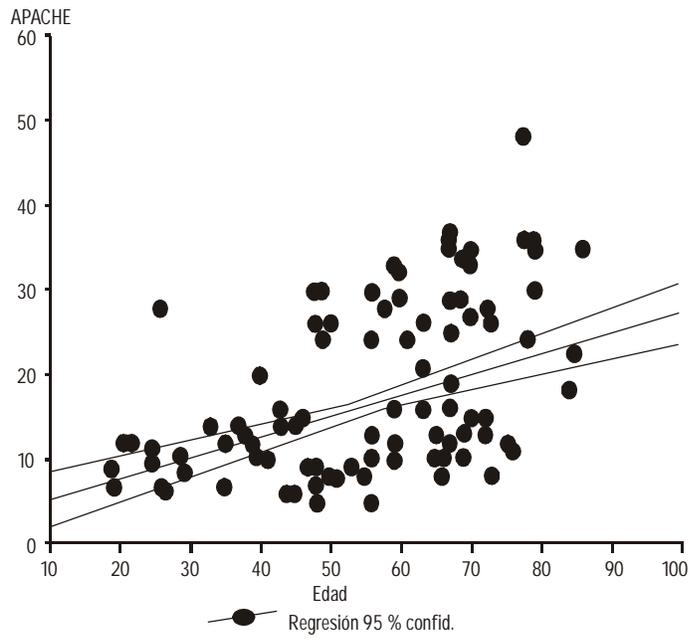


FIG. 2. Modelo de regresión (APACHE vs. edad) n= 116.

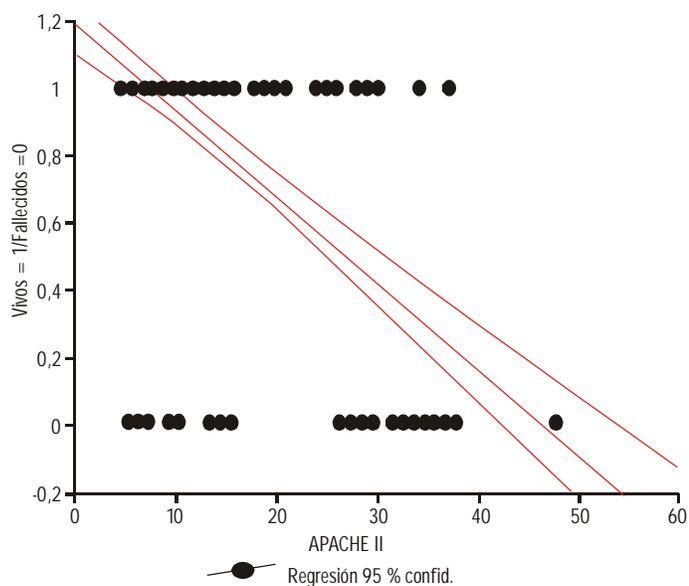


FIG. 3. Modelo de regresión (Supervivencia vs APACHE II).

TABLA 1. Comparación de predicciones realizadas en los 116 pacientes a las 24 y 48 h con el modelo APACHE II

Modelo predictivo	Calibración		Discriminación			
	$\hat{c}$	p	Sens.	Espec.	CCT	AUC
A las 24 h de ingreso						
APACHE II- 24 h	4,45	0,56	33,56	90,45	76,75	0,634
A las 48 h de ingreso						
APACHE II- 48 h	15,4	0,10	21,21	92,12	67,23	0,707

c = estadístico Chi cuadrado de Hosmer-Lemeshow, p = significación estadística, Sens = sensibilidad, Espec = especificidad, CCT = cociente de clasificación correcta total, AUC = área bajo la curva COR.

TABLA 2. Comparación cualitativa de la calibración y la discriminación del modelo APACHE II a las 24 y 48 h.

Modelo predictivo	Calibración	Discriminación
A las 24 h de ingreso		
APACHE II	+++++	++++
A las 48 h de ingreso		
APACHE II	+++	++

## DISCUSIÓN

Los primeros estudios diseñados para la validación de modelos predictivos comenzaron en la década de los 80. *Lemeshow* y

*Hosmer*<sup>10</sup> publicaron varios estudios sobre los modelos más empleados en aquel tiempo y resultó que tenían similares comportamientos en cuanto a calibración y discriminación. *Schafer* y otros,<sup>11</sup> en 1990, compararon la capacidad predictiva del modelo APACHE II con las versiones iniciales del Simplified Acute Physiology Score (SAPS I) y las del Mortality Probability Models (MPM) al ingreso, y obtuvieron resultados similares a los de Lemeshow<sup>12-16</sup> años antes. En el presente estudio se escogió el modelo APACHE II por su poder predictivo demostrado y así obtener una validación a

pesar del tamaño muestral que resultó ser pequeño en comparación con otras investigaciones.

Con respecto al precedente encontrado en la literatura de la utilización del índice predictivo APACHE II como score dinámico, Biot<sup>17-19</sup> y otros en 1988, demostraron en 128 pacientes que el uso longitudinal de una variante del APACHE II, el sistema *sickness scoring*, fue el mejor predictor de la evolución. También en ese año Shuster y otros<sup>13</sup> propusieron el uso del APACHE II corregido con el número de órganos o sistemas en fallo para el pronóstico individual de pacientes. Con posterioridad, en 1994 Douglas Wager y otros,<sup>20</sup> miembros del mismo equipo del APACHE II, utilizaron secuencialmente este modelo, y sugirieron que su uso más allá de 24 h podría proporcionar un refinamiento de sus predicciones de mortalidad.

En este estudio se encontró que las predicciones sobre los 116 pacientes a las 24 h de ingreso con el modelo APACHE II presentó mejor calibración que las realizadas a las 48 h. Al revisar la literatura se encontró que el modelo APACHE II demuestra una excelente calibración a las 24 h, aceptable discriminación. Pero su poder predictivo disminuye ostensiblemente a las 48 h, de forma similar a los modelos más estudiados en el mundo.

La tabla 1 muestra los resultados de la prueba de bondad de ajuste de *Hosmer-Lemeshow* y los valores de calibración y discriminación para las predicciones realizadas sobre los pacientes a las 24 h y a las 48 h. En relación con las 48 h sobre los 116 pacientes las predicciones de la escala APACHE II tuvieron menor calibración y muy precaria discriminación tabla 2. El resultado más significativo del estudio resultó ser la falta de significación en la discriminación a las 48 h de la escala APACHE II, pese al carácter de *score* dinámico que algunos autores quieren ofrecerle. Es muy posible que el peso estadístico de la cate-

goría diagnóstica al ingreso, que tan excelentes resultados otorga al APACHE II en las primeras 24 h, le conduzca a un mal ajuste de sus predicciones a las 48 h. Con respecto a la comparación de la bondad de ajuste (tabla 2) entre las 24 y las 48 h se realizaron las comparaciones que se basaron en los valores del coeficiente p de significación estadística. Para comparar la discriminación del modelo se eligió el área bajo la curva ROC como primer criterio, por encima de sensibilidad, especificidad y tanto por ciento de pacientes bien clasificados en el punto corte del 50 %.

De estudios previos<sup>6-9</sup> se conoce que usando un punto de corte del 50 % virtualmente todos los modelos predictivos propuestos tendrán, al menos, un cociente de clasificación falsa del 10 al 15. Esto ha sido interpretado por algunos autores<sup>16-19</sup> como que todos los sistemas predictivos son aproximadamente equivalentes y ninguno es suficientemente bueno para la predicción individual. Esta afirmación eminentemente matemática aunque se cumple, al menos en teoría, no se puede ver de forma absoluta, pues el criterio de la predicción individual de un modelo está basado en el refinamiento que alcance en su evolución longitudinal, en el tamaño de la muestra utilizada y en la posibilidad de que las variables se puedan ajustar al medio donde se desarrolle. Otros autores comparten nuestros criterios. *Knausy* otros,<sup>14</sup> en 1986, sobre una muestra de 210 pacientes, clasificaron correctamente con el APACHE II al 84,3 % con una especificidad del 95,5 % considerando también un punto de corte del 50 %. En otro estudio, *Castella* y otros,<sup>11</sup> en 1995, obtuvieron una clasificación correctamente con el APACHE II el 68,3 % de 593 pacientes polivalentes con una especificidad del 74,9 % en el mismo punto corte. Finalmente, *Wagner* y otros,<sup>20</sup> en 1994, clasificaron correctamente con el APACHE II el 83,4 % de 332 pacientes, con el 95,4 % de especificidad.

## SUMMARY

Predictive models have become an efficient tool as long as common sense and the experience in its use prevail. A prospective longitudinal cohort inference study was conducted at the Intensive Care Unit of "Dr. Luis Díaz Soto" Military Central Hospital. The analyzed variables were those included in the system of Evaluation of the Acute and Chronic Physiological State (APACHE II). Their value, the prediction-death ratio, and the adjusted prediction-death ratio were calculated at 24 and 48 hours. The calibration was compared by using the Chi square test, and the discrimination by comparing the area under the odds curve (ROC). Values over 30 points demonstrated more than 80 % of death probability. The figure of 40 points or more proved to be 100 % fatal in the study. The APACHE II system showed better calibration and discrimination at 24 hours than at 48 hours ( $p < 0.05$ ).

*Subject headings:* FORECASTING / methods; APACHE; INTENSIVE CARE; PROSPECTIVE STUDIES.

## REFERENCIAS BIBLIOGRÁFICAS

1. Bongard FS, Sue DY. Diagnóstico y tratamiento en Cuidados Intensivos. México, D.F: Editorial Manual Moderno SA. 1996; t 1:35-38.
2. Cowen JS, Kelley MA. Predicting intensive care unit outcome: Errors and bias in using predictive scoring systems. *Crit Care Clin* 1999;10:53-8.
3. Escarce JJ, Kelley MA. Admission source to the medical intensive care unit predicts hospital death independent of APACHE II score. *JAMA* 1999;264:2389-92.
4. Knaus WA, Wagner DP, Drape EA. The APACHE III prognostic: system: Risk prediction of hospital mortality for critically hospitalized adults. *Chest* 1991;100:1619-21.
5. Zimmerman JE, Wagner DP, Draper EA. Evaluation of acute physiology and chronic health evaluation III predictions of hospital mortality in an independent database. *Crit Care Med* 1998;26:1317-9.
6. Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European / North American multicenter study. *JAMA* 1993;270:2957-9.
7. Auriant I, Vinatier I, Thaler F, Drapper B. Simplified acute physiology score II for measuring severity of illness in intermediate care units. *Crit Care Med* 1998;26:1368-70.
8. Lemeshow S, Teres D, Klar J. Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 1998;270:2478.
9. Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health* 1991;81:1630-5.
10. Lemeshow S, Teres D, Avrunin JS, Pastides H. A comparison of methods to predict mortality of intensive care unit patients. *Crit Care Med* 1987;15:715-22.
11. Castella X, Artigas A, Bion J, Kari A. The European / North American Severity Study Group. A comparison of severity of illness scoring systems for intensive care unit patients: Results of a multicenter, multinational study. *Crit Care Med* 1995; 23: 1327-9.
12. Lemeshow S, LeGall JR. Modeling the severity of illness of ICU patients. A systems update. *JAMA* 1994;272:1049-54.
13. Shuster DP, Kollef MH. Predicting intensive care unit outcome with scoring systems: Underlying concepts and principles. *Crit Care Clin* 1994;10:1-4.
14. Knaus WA. Variations in mortality and length of stay in intensive care units. *Ann Intern Med* 1993;188:753-9.
15. Zimmerman JE, Wagner DP, Knaus WA. The use of risk predictions to identify candidates for intermediate care units: Implications for intensive care utilization and cost. *Chest* 1995;108:490-8
16. Li T, Phillips M. On-site physician staffing in a community hospital intensive care unit. *Crit Care Med* 1984; 252:2023-5.
17. Brown JJ, Sullivan GS. Effect on ICU mortality of a full-time critical care specialist. *Chest* 1999; 9: 127-9.
18. Carson SS, Stocking C, Podsadecki T. Effects of organizational change in the medical intensive care unit of a teaching hospital: A comparison of 'open' and 'closed' formats. *JAMA* 2000; 276: 322-6.
19. Zimmerman JE, Shortell S, Knaus WA. Value and cost of teaching hospitals: A prospective, multicenter, inception cohort study. *Crit Care Med* 2001;21:1432-3.
20. Wagner DP, Knaus WA, Harrell FE. Daily prognostic estimates for critically ill adults in intensive care units: Results from a prospective, multicenter, inception cohort analysis. *Crit Care Med* 1999;22:1359-61.

Recibido: 2 de diciembre de 2002. Aprobado: 31 de enero de 2003.

Dr. *Armando Padrón Sánchez*. Hospital Militar Central "Dr. Luis Díaz Soto". Avenida Monumental, Habana del Este, CP 11700, Ciudad de La Habana, Cuba.