

# Valor predictivo de los sistemas de calificación de gravedad: comparación de cuatro modelos en tres unidades de terapia intensiva mexicanas incluidas en la base de datos multicéntrica de terapia Intensiva

**Dr. Ulises W. Cerón Díaz,\* Dr. Juan Esponda Prado,\*\* Dr. Mauricio Borboya Paya,\*\*\* Dr. Jean Paul-Vázquez Mathieu\*\*\*\***

## RESUMEN

**Objetivo:** Comparar el rendimiento de cuatro sistemas de predicción de mortalidad: APACHE II, SAPS II, MPM II-0 y MPM II-24.

**Diseño:** Estudio prospectivo de una cohorte.

**Lugar:** Tres unidades de terapia intensiva mexicanas.

**Enfermos:** 771 enfermos adultos.

**Intervenciones:** ninguna.

**Mediciones y resultados principales:** Área por debajo de la curva ROC (capacidad discriminativa): APACHE II: 0.87, SAPS II: 0.88, MPM II-0: 0.82, MPM II-24: 0.88. Hubo diferencia estadísticamente significante entre el área de MPM II-0 y los demás sistemas. Prueba de bondad de ajuste (calibración): APACHE II:  $p > 0.2$ , SAPS II:  $p > 0.9$ , MPM II-0:  $p > 0.02$ , MPM II-24:  $p > 0.7$ .

**Conclusión:** Los cuatro modelos tienen una buena calibración y capacidad discriminativa. MPM II-0 tuvo los resultados menos ventajosos.

**Palabras clave:** Mortalidad, sistemas de predicción, pacientes críticos.

## SUMMARY

**Objective:** To compare the performance of four mortality prediction models: APACHE II, SAPS II, MPM II-0 and MPM II-24.

**Design:** A prospective cohort study.

**Setting** Three mexican intensive care units.

**Patients:** 771 adult patients.

**Interventions:** None.

**Measurements and main results:** Area under the ROC curve (discrimination capacity): APACHE II: 0.87 SAPS II: 0.88, MPM II-0: 0.82, MPM II-24: 0.88. There was a statistic difference between MPM II-0 area and the area of the other models. Goodness of fit (calibration): APACHE II:  $p > 0.2$ . SAPS II:  $p > 0.9$ , MPM II-0:  $p > 0.02$ , MPM II-24:  $p > 0.7$ .

**Conclusion:** The four models have good calibration and discrimination capacity, MPM II-0 had the least advantage of the four.

**Key words:** Mortality, prediction models, ICU patients.

\* Coordinador de BDM-TI. Responsable del Comité de Informática y Comunicaciones Electrónicas (CICE) de la Asociación Mexicana de Medicina Crítica y Terapia Intensiva. Médico adscrito de la UTI del Hospital Español de México.

\*\* Jefe de la UTI del Hospital Angeles del Pedregal.

\*\*\* Jefe de la UTI del Hospital Santelena.

\*\*\*\* Miembro del CICE de la Asociación Mexicana de Medicina Crítica y Terapia Intensiva. Médico adscrito de la UTI del Hospital Español de México.

Grupo multicéntrico BDM-TI, Comité de Informática y Comunicaciones Electrónicas de la Asociación Mexicana de Medicina Crítica y Terapia Intensiva.

Uno de los grandes avances en la Medicina Crítica ha sido la constitución de escalas generales de calificación de gravedad y modelos predictivos de mortalidad, que nos permiten hacer una mejor descripción de grupos de enfermos, nos ayudan a tomar decisiones y nos permiten evaluar el rendimiento de las unidades de terapia intensiva. El primer método, Acute Physiology and Chronic Health Evaluation (APACHE), fue publicado en 1981; sus creadores usaron un abordaje teórico para construir la escala de calificación del grado de alteración fisiológica.<sup>1</sup> En 1984 se publicó la primera versión de Simplified Acute Physiology Score

(SAPS);<sup>2</sup> y en 1985<sup>3</sup> APACHE II, que reduce el número de variables de 34 a 14 e incluye los pesos que dan la edad y el estado de salud crónico; en el mismo año se describió un modelo derivado estadísticamente, que se llamó Mortality Prediction Model (MPM), el cual ofrece tres ecuaciones para predecir la mortalidad con los datos de ingreso, de 24 y 48 horas.<sup>4</sup> En 1991 se publicó APACHE III, sin embargo no se hacen públicos la totalidad de los elementos necesarios para calcular la probabilidad de morir.<sup>5</sup> En 1993 se dieron a conocer los resultados de dos métodos, SAPS II<sup>6</sup> y MPM II;<sup>7</sup> este último ofrece modelos matemáticos para predecir la mortalidad con los datos de ingreso y de 24 horas (MPM II-0 y MPM II-24).

El APACHE II y SAPS II, además de permitir una calificación de la gravedad de la enfermedad crítica, también ofrecen modelos matemáticos para calcular la probabilidad de morir, de tal manera que se puede comparar el rendimiento de los cuatro modelos de predicción. Hay pocos trabajos multicéntricos publicados en la literatura mundial que hagan esta comparación. En México el Dr. Sánchez-Velázquez comparó la capacidad discriminativa de TISS, APACHE II, APACHE III, SAPS I, SAPS II, MPM II-0 y MPM II-24<sup>8</sup> en un grupo de 272 ingresos en un hospital privado de la ciudad de México; concluyó que el sistema con mayor capacidad discriminativa menor costo es APACHE II y desde junio de 1998 se inició la captura prospectiva en nuestro hospital en una base de datos general multicéntrica, que incluye los datos y herramientas para el cálculo de índices generales de gravedad y estimaciones de la probabilidad de morir.

El propósito de este trabajo es comparar el rendimiento de cuatro modelos matemáticos para predecir la mortalidad hospitalaria (APACHE II, SAPS II, MPM II-0 y MPM II-24), con los registros incluidos en la Base de Datos Multicéntrica de Terapia Intensiva (BDM-TI).<sup>9</sup>

## MATERIAL Y MÉTODOS

BDM-TI es el nombre de un proyecto que nace a partir de la experiencia lograda en la Unidad de Terapia Intensiva (UTI) del Hospital Español de México con una base de datos local denominada BASUTI.<sup>9</sup> Este proyecto ha sido incorporado a las actividades del Comité de Informática y Comunicaciones Electrónicas de la Asociación Mexicana de Medicina Crítica y Terapia Intensiva (AMMCTI). La planeación y desarrollo se iniciaron en 1997; en ju-

nio de 1998 se comenzó la recolección de datos en la primera UTI, en julio de 1998 se incorporó la segunda, y en septiembre de 1998 la tercera. BDM-TI es un sistema que ofrece a las UTIs una herramienta para capturar datos relacionados con aspectos demográficos, complicaciones, procedimientos, gravedad de padecimientos específicos, fallas orgánicas, evolución, cálculo de índices generales de gravedad y estimaciones de la probabilidad de morir. Esto se realiza a través de un software que contiene herramientas que permiten personalizar las listas de opciones, automatiza los cálculos y genera reportes. Los datos son centralizados a través de una herramienta de exportación. Los datos de la central son analizados con el propósito de generar información comparativa respecto a la media multicéntrica y a estándares internacionales.

Los datos se capturaron en forma prospectiva, siguiendo criterios uniformes y de acuerdo a los métodos de las publicaciones originales. El software contiene una herramienta que permite la captura de todos los datos necesarios para calcular APACHE II y SAPS II, así como para calcular las probabilidades de morir según los modelos APACHE II, SAPS II, MPM II-0 y MPM II-24. Los datos se obtuvieron por personal médico y posteriormente se capturaron en una computadora. El control de calidad de los datos incluyó supervisión por personal capacitado y herramientas del propio software. Los enfermos fueron seguidos en su evolución hospitalaria hasta su alta.

Las unidades que aportaron datos, se encuentran localizadas en la zona metropolitana de la ciudad de México y cuentan con características comunes tales como: que son privadas y disponen de atención médica especializada las 24 horas del día; las demás características se presentan en el cuadro 1.

Cada unidad aportó un número diferente de enfermos, lo cual dependió del volumen de ingresos y de los datos exportados hasta el momento de corte en la base de datos central. Los datos centralizados fueron filtrados de acuerdo a los siguientes criterios de exclusión: a) reingresos, b) menores de 18 años, c) operados de corazón, d) enfermos coronarios, e) trasladados a otro hospital, f) registros incompletos. La razón de excluirlos es el hecho de que no son incluidos en las publicaciones originales de los modelos matemáticos.

El recuento y análisis de los datos se hicieron a través de programas creados en Visual FoxPro<sup>®</sup>, para realizar las siguientes tareas: 1) recuento de los datos demográficos, 2) obtención de la distribución

Cuadro I.

	Unidad 1	Unidad 2	Unidad 3
Número de camas	12	14	5
Personal médico:			
Jefe especialista	1	1	1
Adscritos especialistas	6	6	5
Residentes de la especialidad	9	4	0
Residentes de otras especialidades	8	2	0
Personal de enfermería:			
Jefa	0	1	1
Jefas de turno	4	6	4
Enfermeras fijas	45	16	
Enfermeras eventuales	0	6	
Relación enfermo/enfermera promedio	1:2	1:1	1:1
Equipo *			
Ventiladores	11	6	5
Pulsoxímetros		14	5
Gasto cardíaco	6	4	5
Presión arterial no invasiva		7	5
Presión arterial invasiva	12	14	5

\* Número de camas que pueden contar con el equipo en forma simultánea.

de los enfermos de acuerdo a su riesgo predicho de morir, 3) cálculo de los estadísticos "C" y "H" de la prueba de bondad de ajuste de Lemeshow-Hosmer, 4) para hacer un análisis de prueba diagnóstica en diferentes puntos de corte de cada escala y poder construir las curvas ROC. El rendimiento de las escalas se analizó en dos aspectos: a) capacidad discriminativa y b) calibración. Para evaluar la capacidad discriminativa, se construyeron curvas ROC, se calculó el área por debajo de la curva<sup>10</sup> y se estableció si existía diferencia estadística entre los cuatro modelos.<sup>11</sup> Para ello se creó un programa en Excelâ . Para evaluar la calibración se hizo una prueba de bondad de ajuste siguiendo las recomendaciones de Lemeshow y Hosmer<sup>12</sup> y se elaboraron curvas de calibración. Para la construcción de las curvas ROC se hizo un análisis en cada uno de 10 puntos de corte en cada modelo: los puntos de corte se calcularon de acuerdo a la fórmula de Sturges,  $k = 1 + 3.322 (\log_{10} n)$ , donde "k" es el número de intervalos de clase y "n" es el número de valores en el conjunto de datos en observación. Para el análisis de la calibración se calcularon dos estadísticos; el estadístico "C" requiere que los intervalos de clase incluyan cantidades iguales de registros, para lo cual se programó la computadora para que ordenara en forma ascendente las probabilidades de morir y luego dividiera al grupo en diez partes iguales (deciles); el es-

tadístico "H" requiere que los rangos de los intervalos de clase sean iguales, por lo que usamos la fórmula de Sturges, con lo que cada intervalo de clase tuvo un número diferente de registros. La prueba de bondad de ajuste nos permite evaluar la discrepancia entre el número de muertes observadas y muertes esperadas, así como sobrevivientes observados y sobrevivientes esperados; mientras menor es la discrepancia, menor es el valor de los estadísticos "C" y "H" y mayor es la "p". Las curvas de calibración se construyeron comparando la mortalidad predicha y la mortalidad actual de cada intervalo de clase; y se calculó el coeficiente de correlación de Pearson y  $r^2$ .

Todos los valores son expresados en frecuencias, proporciones, media, desviación estándar y error estándar según sea el caso.

## RESULTADOS

De un total de 1,173 ingresos incluidos en la base de datos central, se analizaron 771. Las causas por la que se excluyeron los 402 ingresos restantes, se presentan en el cuadro II. Es de notar que 24 ingresos excluidos tienen dos o más razones para ello; este es el motivo por el que no coinciden el total de la tabla y el número de ingresos excluidos.

Los datos demográficos de la totalidad de los enfermos y de las tres unidades participantes se presentan en el cuadro III. Es importante notar que la unidad 3 se diferencia de las demás por tener una baja proporción de enfermos que ingresan a cuidados especiales, la gran mayoría lo hacen por presentar falla de uno o más sistemas orgánicos vitales, por la baja proporción de enfermos sometidos a cirugía en general y a cirugía de urgencia en particular.

El grupo total se caracteriza porque la proporción de enfermos quirúrgicos es alta (45.5%). El

Cuadro II.

Registros incompletos	32
Reingresos	16
Edad < 18 años	74
Enfermos coronarios	126
Operados de corazón	118
Alta a otro hospital	60

Número de enfermos excluidos por cada uno de los criterios. La razón por la que no coincide el total de este cuadro con el número de registros excluidos es porque 24 enfermos tuvieron más de un criterio.

Cuadro III.

	Unidad 1	Unidad 2	Unidad 3	Total
n=	356	276	139	771
Sexo (M/F)	177(49)/179(51)	169(61)/107(39)	76(55)/63(45)	422(55)/349(45)
Edad (X ± DE)	60 ± 19	55 ± 19	54 ± 21	57 ± 20
FDI (A/B/C)	211(59)/28(8) /117(33)	52(19)/111(40) /113(41)	89(64)/45(32) /5(4)	352(46)/184(24) /235(30)
Condición antes del ingreso				
Desconocida	0	1(0.4)	12(8.6)	13(1.7)
Encamado	1(0.3)	53(19.2)	18(12.9)	80(10.4)
< 50% con	15(4.21)	3(1.08)	1(0.7)	19(2.33)
< 50% sin	5(1.4)	4(1.44)	1(0.7)	10(1.3)
Sintomático	209(58.7)	144(52.2)	73(52.5)	426(55.3)
Normal	125(35)	66(23.9)	32(23)	223(28.9)
Condición de ingreso				
Estable	187(51.9)	204(73.9)	74(53.2)	465(59.14)
Cítico/inestable	159(44.6)	68(24.6)	59(42.4)	286(37.1)
Moribundo	10(2.8)	4(1.44)	6(4.31)	20(2.59)
Cirugía de urgencia	62(17.4)	26(9.4)	3(2.16)	91(11.8)
Categorías diagnósticas*				
Posoperados	169(47.5)	129(46.7)	53(38.1)	351(45.5)
No. Qx Falta respiratoria	73(20.5)	25(9.05)	11(7.9)	109(14.13)
No. Qx Falla cardiovascular	37(10.4)	14(5.07)	4(2.9)	55(7.13)
No. Qx Trauma	13(3.65)	11(4.98)	8(5.7)	32(4.15)
No. Qx Neurológico	12(3.37)	10(3.6)	9(6.5)	31(4)
No. Qx (otros)	10(2.8)	8(2.9)	11(7.9)	29(3.76)
No. Qx Sistemas orgánicos	41(11.5)	53(19.2)	41(29.5)	135(17.5)

FDI= factor determinante de ingreso; A= falla de uno o más sistemas orgánicos mayores; B= riesgo de establecer una falla en un sistema orgánico mayor; C= cuidados especiales; 50% con = deambula menos del 50% del tiempo y necesita asistencia; < 50% sin = deambula menos del 50% del tiempo pero no necesita asistencia; No Qx= no quirúrgico; (otros) = cetoacidosis, sobredosis, sangrado de tubo digestivo; sistemas orgánicos = enfermos cuya categoría diagnóstica no pertenece a los grupos anteriores y que son asignados de acuerdo al órgano o sistema que se afecta principalmente.

\* De acuerdo a la clasificación usada en APACHE II.

17% de los enfermos son sometidos a cirugía de urgencia. La condición previa a la enfermedad que originó su ingreso al hospital fue “sintomático por algún padecimiento crónico” en el 55.3% y le sigue la condición “normal” con el 28.9%. La mayor parte de los enfermos ingresaron estables (59%) y sólo el 2.6% ingresó en estado moribundo. De las categorías diagnósticas al ingreso la más frecuente fue el posoperatorio (45.5%) y le siguen las categorías no quirúrgicas por falla o insuficiencia respiratoria y por falla o insuficiencia cardiovascular con un 14 y 7% respectivamente; en el 17% de los casos no se pudo clasificar con precisión, por lo que se agregaron a categorías no específicas, como lo recomienda el sistema APACHE II.

La gravedad de los enfermos se expresa en el cuadro IV. El promedio de APACHE II fue 16.25 y de SAPS II de 34.53 puntos; es sobresaliente la mayor gravedad de los enfermos de la Unidad 3. La morta-

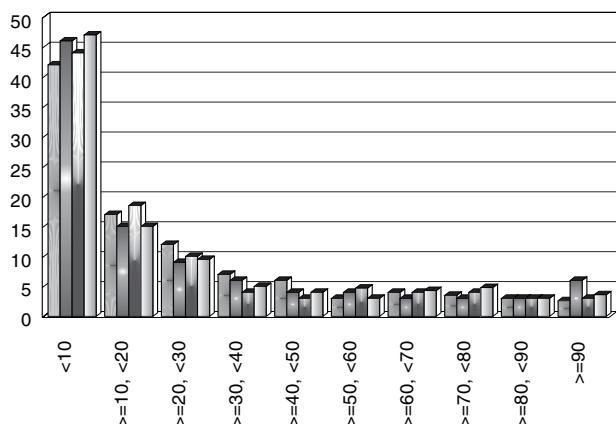
lidad hospitalaria del grupo total (23.9%) es muy parecida a la probabilidad de morir promedio obtenida con cada uno de los cuatro modelos matemáticos.

La distribución de los enfermos de acuerdo a su riesgo de morir en el hospital, se presenta en la figura 1.

Se observa que alrededor de la mitad de los enfermos tienen una probabilidad menor del 10% y que la distribución es semejante para todos los modelos matemáticos.

La probabilidad de morir promedio para cada modelo en la totalidad de los ingresos y en los grupos de vivos y muertos, se presenta en el cuadro V.

El estadístico “H” de la prueba de bondad de ajuste de Lemeshow y Hosmer fue de 8.41, 15.27, 22.02 y 16.39 para APACHE II, SAPS II, MPM II-0 y MPM II-24 respectivamente; la significancia estadística fue  $p > 0.3$ ,  $p > 0.05$ ,  $p < 0.01$  y  $p < 0.05$  respectivamente. Los resultados para el estadístico

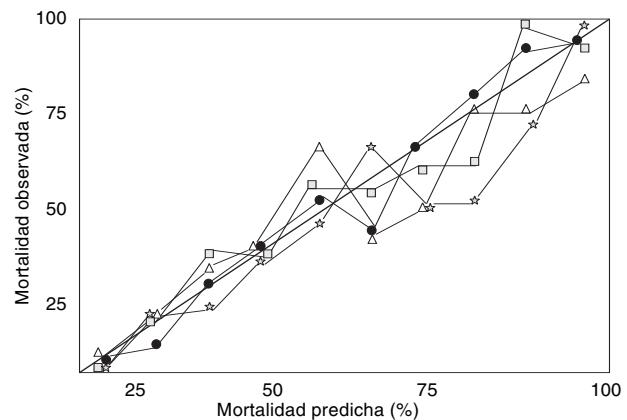


**Figura 1.** Distribución de los enfermos de acuerdo a la probabilidad de morir estimada por cada modelo matemático. En el eje de las «X» se encuentran los intervalos de clase de la probabilidad de morir y en el eje de las «Y» el porcentaje de enfermos para cada uno de los modelos. En barras APACHE II, barras SAPS II, barras MPM II-0, barras MPM II-24.

“C” de la prueba de bondad de ajuste de Lemeshow y Hosmer, fue notablemente mejor para las cuatro escalas en relación al análisis anterior (*Cuadro VI*); es de resaltar que la escala menos favorecida fue MPM II-0. Las dos escalas que obtuvieron buenos resultados con los dos estadísticos fueron APACHE II y SAPS II.

MPM II-24 tuvo resultados inconsistentes y MPM II-0 obtuvo resultados menos favorables con los dos estadísticos.

Las curvas de calibración para los cuatro modelos se representan en la *figura 2*; es de notarse que en todos existe una tendencia a mantener una mejor calibración en probabilidades de morir menores del 50%, mientras que en mortalidades altas, en general tienden a infraestimar la mortalidad actual



**Figura 2.** Curvas de calibración para los cuatro modelos. La línea recta representa la calibración perfecta. Observe que los cuatro modelos tienen una excelente calibración cuando la probabilidad de morir es menor del 50%. En probabilidades altas de morir los modelos tienden a infraestimar la mortalidad y la calibración es menos satisfactoria. Triángulo= MPM II-0; Cuadrado= MPM II-24; Círculo= APACHE II, Estrella= SAPS II.

y existe una dispersión mayor de los puntos reflejando diferencias en la calibración; la escala que visualmente presenta mejor calibración es APACHE II. Los coeficientes de correlación y  $r^2$  para cada modelo fueron: APACHE II 0.98 y 0.97, SAPS II 0.93 y 0.86, MPM II-0, 0.92 y 0.83, MPM II-24, 0.96 y 0.91. La significancia estadística fue muy buena para todos los modelos ( $p < 0.001$ ).

En la *figura 3* se presentan las curvas ROC para los cuatro modelos. Se puede observar que la curva de MPM II-0 tiene un área visualmente menor que la de los otros métodos. Esto se refleja en los cálculos del área y el significado estadístico de la comparación entre los modelos, como se presenta

**Cuadro IV.**

	Unidad 1	Unidad 2	Unidad 3	Total
n=	356	276	139	771
APACHE II	$17.08 \pm 9.8$	$13.32 \pm 8.06$	$19.88 \pm 9.46$	$16.25 \pm 9.46$
SAPS II	$36.99 \pm 23.3$	$29.98 \pm 18.44$	$37.22 \pm 21.38$	$34.53 \pm 21.59$
Mortalidad hospitalaria	26.1	15.2	35.3	23.9
Prob. APACHE II	$26.39 \pm 26.19$	$17.69 \pm 20.53$	$33.00 \pm 27.92$	$24.47 \pm 25.18$
Prob. SAPS II	$27.45 \pm 30.82$	$18.63 \pm 24.05$	$28.63 \pm 29.61$	$24.51 \pm 28.69$
Prob. MPM II-0	$23.24 \pm 26.17$	$19.47 \pm 23.80$	$33.26 \pm 29.79$	$23.69 \pm 26.49$
Prob. MPM II-24	$26.85 \pm 28.39$	$17.94 \pm 22.88$	$27.89 \pm 29.97$	$23.85 \pm 27.23$

Se expresa en media  $\pm$  desviación estándar, Prob= probabilidad de morir según el modelo correspondiente.

Cuadro V.

N(%)	Probabilidad de morir		
	Todos	Vivos	Muertos
APACHE II	24.47 ± 25.18	15.59 ± 16.39	52.81 ± 27.31
SAPS II	24.51 ± 28.69	13.99 ± 17.72	58.05 ± 31.12
MPM II-0	23.69 ± 26.49	15.80 ± 19.34	48.89 ± 30.23
MPM II-24	23.85 ± 27.23	14.38 ± 18.31	54.05 ± 28.96

Probabilidad promedio de morir en el grupo total y en los subgrupos de vivos y muertos, calculados con los cuatro modelos.

en los cuadros VII y VIII, donde MPM II-0 tiene un área significativamente menor.

## DISCUSIÓN

Los modelos predictivos son usados rutinariamente en muchas unidades del mundo se han utilizado para comparar resultados entre países,<sup>13-16</sup> para asistir en la decisión de egreso de la UTI,<sup>17</sup> para evaluar el rendimiento de las unidades,<sup>18</sup> para la aleatorización en numerosos trabajos de investigación. Puesto que el tipo de enfermo y las prácticas son diferentes en otros países distintos a los que corresponden a los estudios originales, es necesario evaluar su comportamiento en otros entornos.

En este trabajo se evalúa el rendimiento de cuatro modelos matemáticos para predecir la mortalidad (APACHE II, SAPS II, MPM II-0, MPM II-24) a

través del análisis de datos capturados en forma prospectiva y recuperados de una base de datos general multicéntrica, con la participación de tres unidades de terapia intensiva mexicanas.

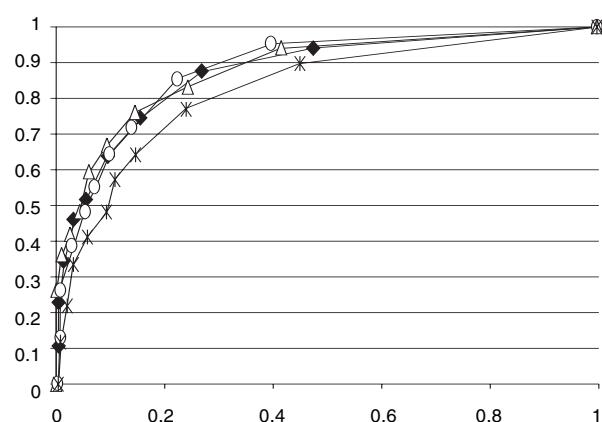
El rendimiento fue analizado en base a dos características: calibración y su capacidad discriminativa.

Aunque los coeficientes de correlación de las curvas de calibración fueron muy buenos, la mejor forma de analizar esta propiedad es a través de las pruebas de bondad de ajuste. Siguiendo las recomendaciones de Lemeshow y Hosmer, se observó superioridad de APACHE II y SAPS II respecto a MPM II-0 y MPM II-24. En relación a la capacidad discriminativa se observó inferioridad de MPM II-0 respecto a los demás modelos. Castella et al. realizaron un trabajo de investigación en 137 unidades de 12 países, incluyendo a 13.152 enfermos divididos en dos grupos, uno de desarrollo y otro de validación. Reportó que las nuevas versiones (APACHE III, SAPS II, MPM II) son superiores a las versiones anteriores (APACHE II, SAPS I, MPM I) y que APACHE II, SAPS II y MPM II tienen una buena capacidad discriminativa y buena calibración.<sup>19</sup>

Más recientemente Moreno et al. compararon en 982 enfermos de 17 unidades portuguesas, el rendimiento de dos modelos (APACHE II y SAPS II). Concluyeron que SAPS II tuvo un rendimiento superior a APACHE II y que los resultados no les permiten usarlos, al menos sin ser adaptados, para analizar la calidad de cuidado o el rendimiento entre las UCIs en este grupo de enfermos.<sup>20</sup> En 1998, Nouira et al. compararon cuatro modelos (APACHE II, SAPS II, MPM II-0 y MPM II-24) en un grupo de 1,325 enfermos ingresados a tres unidades de terapia intensiva de Túnez. Concluyeron que a pesar de una buena capacidad de discriminación, los cuatro modelos demostraron una calibración menos satisfactoria.<sup>21</sup>

En el cuadro IX se presentan los resultados de los tres trabajos comentados con anterioridad y el nuestro, así como los resultados de los trabajos originales de cada modelo. En general son notorios los mejores resultados obtenidos en nuestro estudio tanto en capacidad discriminativa como en calibración.

La razón por la que nuestros resultados reflejan una mejor calibración y capacidad discriminativa no está clara. La explicación puede estar en alguna de las diferencias que se presentan en el cuadro X, donde se observa que tanto el número de unidades incluidas como el número de ingresos son muy diferentes; nuestro estudio además de ser el que menos enfermos incluye, tiene el porcentaje más alto de exclusiones y es de mencionar que 60 enfermos no



**Figura 3.** Curvas ROC (Receiver Operating Characteristic) para los cuatro modelos. En el eje de las «X» se presenta a (1-especificidad) y en el eje de las «Y» se presenta a la sensibilidad. Observe cómo MPM II-0 tiene menos área por debajo de la curva.

**Cuadro VI. Prueba de bondad de ajuste usando la técnica de Lemeshow-Hosmer. C= estadístico «C», gL= grados de libertad.**

Sistema Probabilidad de morir	Muertos n=	Vivos Observados	Esperados	Observados	Esperados
<b>APACHE II</b>					
0-0.026	77	2	1.34	75	75.66
0.0068-0.04	77	3	2.53	74	74.47
0.04-0.06	77	3	3.78	74	73.21
0.06-0.09	77	3	5.81	74	71.18
0.09-0.14	77	3	9.09	74	67.91
0.14-0.20	77	14	13.64	63	63.36
0.21-0.29	77	19	19.07	58	57.93
0.29-0.42	77	29	27.84	48	49.16
0.43-0.65	77	38	41.76	39	35.23
0.66-1	78	70	63.8	8	14.19
Total	771	184	188.66	587	582.3
C = 10.83, gL = 8, p > 0.2					
<b>SAPS II</b>					
0-0.01	77	1	0.52	76	76.48
0.01-0.02	77	1	1.33	76	75.67
0.02-0.04	77	2	2.53	75	74.47
0.04-0.06	77	3	3.99	74	73.01
0.06-0.12	77	7	6.79	70	70.21
0.12-0.18	77	14	11.06	63	65.94
0.18-0.29	77	14	17.66	63	59.34
0.29-0.46	77	27	28.29	50	48.71
0.46-0.77	77	46	46.24	31	30.76
0.77-1	78	69	70.53	9	7.47
Total	771	184	188.94	587	582.06
C = 3.25, gL = 8, p > 0.9					
<b>MPM II-0</b>					
0-0.02	77	1	1.14	76	75.86
0.02-0.04	77	3	2.38	74	74.62
0.04-0.05	77	2	3.45	75	73.55
0.05-0.08	77	8	4.87	69	72.13
0.08-0.12	77	12	7.51	65	69.49
0.12-0.17	77	11	11.18	66	65.82
0.17-0.26	77	21	16.12	56	60.88
0.26-0.45	77	31	25.76	46	51.24
0.45-0.70	77	35	45.55	42	32.45
0.70-100	78	60	65.74	18	12.26
Total	771	184	182.7	587	588.3
C = 17.46, gL = 8, p > 0.02					
<b>MPM II-24</b>					
0-0.02	77	1	1.12	76	75.88
0.02-0.03	77	1	1.88	76	75.12
0.03-0.04	77	2	2.76	75	74.24
0.04-0.07	77	3	4.34	74	72.66
0.07-0.11	77	6	6.84	71	70.16
0.11-0.18	77	10	10.98	67	66.02
0.18-0.27	77	22	16.77	55	60.23
0.27-0.44	77	31	26.90	46	50.10
0.44-0.73	77	44	45.62	33	31.38
0.73-1	78	64	66.67	14	11.33
Total	771	184	183.88		587.12
C = 5.229, gL = 8, p < 0.07					



Cuadro VII.

	A ± EE
APACHE II	0.87 ± 0.015
SAPS II	0.88 ± 0.015
MPM II-0	0.82 ± 0.018
MPM II-24	0.88 ± 0.014

A= área por debajo de la curva ROC, EE= error estándar.

fueron incluidos porque se dieron de alta a otro hospital (dato que no es reportado en los otros trabajos). La edad de los enfermos incluidos en el estudio tunecino, es la menor debido probablemente a que incluyen a los enfermos de 16 y 17 años, lo cual no sucede en los demás estudios que se limitan a enfermos de 18 años o más. Nuestros enfermos tienen el porcentaje más alto de ingresos quirúrgicos. Los cuatro estudios coinciden en los criterios de exclusión excepto en la edad como se comentó anteriormente. Nuestros enfermos tienen la calificación de APACHE II más baja, mientras que el estudio tunecino tiene la calificación más baja de SAPS II (dato calculado del reporte original). Por otro lado, la mortalidad en este último grupo tiende a ser mayor que la mortalidad esperada, mientras que en los restantes es prácticamente igual. El presente estudio difiere de los demás en que los datos han sido obtenidos de una base de datos multicéntrica general que incluye una herramienta que automatiza el cálculo: en los demás estudios, tanto la recuperación de datos como la centralización, informatización y análisis se hicieron con propósitos del estudio. Todos los trabajos tienen sistemas de control de calidad de los da-

Cuadro VIII.

	p
APACHE II vs SAPS II	NS
APACHE II vs MPM II-0	< 0.001
APACHE II vs MPM II-24	NS
SAPS II vs VPM II-24	NS
SAPS II vs MPM II-0	< 0.001
MPM II-0 vs MPM II-24	< 0.001

Comparación del área por debajo de la curva ROC.

tos; el nuestro se basa fundamentalmente en la revisión por personal médico entrenado de la hoja de captura antes del vaciamiento a la computadora, aunque el software también dispone de algunos sistemas de control.

En conclusión este estudio contribuye a la validación internacional de los modelos de predicción de muerte utilizados a nivel mundial. Todos los modelos tienen un rendimiento aceptable, comparable y con frecuencia superior al reportado por otros estudios semejantes. APACHE II, SAPS II, MPM II-24 tienen resultados comparables; MPM II-0 es el sistema menos favorecido. Consideramos que es necesario aumentar el tamaño de la muestra así como el número de unidades participantes para obtener resultados más representativos de la realidad en México.

#### AGRADECIMIENTOS

A las autoridades de la AMMCTI por su apoyo y confianza en el proyecto. Al personal médico y de enfermería de las unidades participantes.

Cuadro IX.

Ref.	APACHE II	SAPS II	MPM II-0	MPM II-24	ADC	“C”	“H”	ADC	“C”	“H”	ADC	“C”	“H”
19	0.857	0.024	0.007	0.847	0.024	0.101	0.01	0.805	0.072	0.014	0.833	0.093	0.025
20	0.787	< 0.0001	0.0003	0.817	0.002	0.001							
21	0.82	< 0.01	< 0.05	0.84	< 0.001	< 0.05		0.85	< 0.001	< 0.05	0.88	< 0.001	< 0.05
3	0.086												
6				0.86		0.883							
7							0.837	0.623					
7											0.844	0.764	
*	0.873	> 0.2	> 0.3	0.88	> 0.9	> 0.05	0.825	> 0.02	< 0.01	0.882	> 0.7	< 0.05	

Comparación entre los resultados de la evaluación del rendimiento de cuatro modelos matemáticos para predecir la mortalidad en enfermos graves, en 6 publicaciones y nuestros resultados (\*). ADC= área debajo de la curva ROC, «C» = valor de «p» asociado al estadístico C, «H» = valor de «p» asociado al estadístico H. Recordar que mientras más alto es el valor de «p» mejor es la calibración.

Cuadro X.

	Ref. 19	Ref. 20	Ref. 21	*
No. de UTIs	137	19	3	3
No. de países	12	1	1	1
No. de ingresos	14.745	1.094		1.173
Ingresos excluidos	1593 (10.8%)	112(10.2%)		402(34.3%)
Ingresos incluidos	13.152	982	1325	771
% de varones	59.6	67	57.7	55
Edad		55 ± 19	46.6 ± 20	57 ± 0.20
Tiempo de ingreso:				
Médico (%)		68	84.2	50.5
Cirugía electiva (%)		12	6.2	37.7
Cirugía de urgencia (%)		19.6	9.6	11.8
Criterios de exclusión:				
Quemados	Si	Si	Si	Si
Cirugía cardiaca	Si	Si	Si	Si
Coronarios	Si	Si	Si	Si
Reingresos	Si	Si	Si	Si
Menores de 18 años	Si	Si		Si
Menores de 16 años			Si	
Alta a otro hospital				Si
APACHE II	19.6 ± 9.8	20.3**	16.25 ± 9.4	
SAPS II		41.4 ± 20.7	28.5**	34.53 ± 21.59
Mortalidad por APACHE II		33.5 ± 27.4	20.3	24.47 ± 25.18
Mortalidad por SAPS II		32.6 ± 29.9	19	24.51 ± 28.69
Mortalidad hospitalaria	21.8	32	24	23.9

\* Nuestros resultados. \*\* Calculados a partir de datos de la publicación original.

A nuestro maestro el Dr. Alfredo Sierra Unzueta.  
Jefe de la Unidad de Terapia Intensiva Hospital  
Español de México.

A los médicos que han colaborado directamente  
en la coordinación da la obtención y captura de los  
datos:

Dr. Sergio González Vázquez.

Médico adscrito a la UTI del Hospital Español de  
México.

Dra. Miriam Villada

Médico residente de la UTI del Hospital Angeles  
del Pedregal.

Dr. Santa López

Médico residente de la UTI del Hospital Español  
de México.

Dra. Miriam Hernández

Médico residente de la UTI del Hospital Angeles  
del Pedregal.

Dr. Alfonso García Luna.

Médico residente de la UTI del Hospital Angeles  
del Pedregal.

Al Dr. Luis David Sánchez por su colaboración  
en el análisis estadístico.

## BIBLIOGRAFÍA

1. Knaus WA, Zimmerman JE, Wagner DP et al. APACHE-Acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981; 9: 591-597.
2. Le Gall J, Loirat P, Alperovitch A et al. A simplified acute physiology score for ICU patients. *Crit Care Med* 1984; 13: 818-829.
3. Knaus WA, Draper EA, Wagner DP et al. APACHE II: A severity of disease classification system. *Crit Care Med* 1985; 13: 818-829.
4. Lemeshow S, Teres D, Pastides H et al. A method for predicting survival and mortality of ICU patients using objectively derived weights. *Crit Care Med* 1985; 13: 519-525.
5. Knaus WA, Wagner, Draper EA et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; 100: 1619-1636.
6. Le Gall J, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270: 2957-2963.
7. Lemeshow S, Teres D, Klar J et al. Mortality probability models (MPM II) based on an International cohort of intensive care unit patients. *JAMA* 1993; 270: 2478-2486.
8. Sánchez-Velázquez LD. Capacidad discriminativa y costo de los sistemas de calificación de la gravedad de la enfermedad en la unidad de terapia intensiva. (En prensa).

9. Cerón U, Sierra A, Martínez R, Vázquez JP. Base de datos para el control de calidad y utilización de recursos en la Unidad de Terapia Intensiva. *Rev Mex Med Crit y Ter Int* 1996; 10: 105-201.
10. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29-36.
11. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148: 839-843.
12. Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982; 110: 847-864.
13. Knaus WA, Le Gall GJ, Wagner DP et al. A comparison of intensive care in the USA and France. *Lancet* 1982; II: 642-646.
14. Zimmerman JE, Knaus WA, Judson JA et al. Patient selection for intensive care: A comparison of New Zealand and United States hospitals. *Crit Care Med* 1988; 16: 318-326.
15. Sirio C, Tajimi K, Tase C et al. An initial comparison of intensive care in Japan and the United States. *Crit Care Med* 1992; 20: 1207-1215.
16. Teik O, Hutchinson R, Short S et al. Verification of the acute physiology and chronic health evaluation scoring system in Hong Kong. *Intensive Care Med* 1993; 21: 698-705.
17. Zimmerman JE, Douglas P, Wagner DP et al. Improving intensive care unit discharge decisions: Supplementing physician judgment with predictions of next day risk for life support. *Crit Care Med* 1994; 22: 1373-1384.
18. Rapoport J, Teres D, Lemeshow S, Gehlbach S. A method for assessing the clinical performance and cost-effectiveness of intensive care units: a multicenter inception cohort study. *Crit Care Med* 1994; 22: 1385-1391.
19. Castella X, Artigas A, Bion J et al. A comparison of severity of illness scoring systems for intensive care unit patients: Results of a multicenter multinational study. *Crit Care Med* 1995; 23: 1327-1335.
20. Moreno R, Morais P. Outcome prediction in intensive care: results of a prospective, multicentre. Portuguese study. *Int Care Med* 1997; 23: 177-186.
21. Nouira S, Belghith M, Elatrous S et al. Predictive value of severity scoring systems: Comparison of four models in Tunisian adult intensive care units. *Crit Care Med* 1998; 26: 852-859.

**Correspondencia:**

Dr. Ulises W. Cerón Díaz  
Hospital Español de México  
Unidad de Terapia Intensiva  
Av. Ejército Nacional 163  
Col. Granada  
C.P. 11560, México D.F.  
Tel. 55-31-69-84