



## ORIGINAL

# Escalas pronósticas en la disfunción multiorgánica: estudio de cohortes



M. Sánchez-Casado<sup>a,\*</sup>, V.A. Hostigüela-Martín<sup>b</sup>, A. Raigal-Caño<sup>a</sup>, L. Labajo<sup>c</sup>, V. Gómez-Tello<sup>d</sup>, G. Alonso-Gómez<sup>a</sup> y F.M. Aguilera-Cerna<sup>a</sup>

<sup>a</sup> Unidad de Cuidados Intensivos, Complejo Hospitalario de Toledo, Toledo, España

<sup>b</sup> Unidad de Cuidados Intensivos, Hospital Fundación Jiménez Díaz, Madrid, España

<sup>c</sup> Servicio de Urgencias, Hospital Universitario La Paz, Madrid, España

<sup>d</sup> Unidad de Cuidados Intensivos, Hospital Moncloa, Madrid, España

Recibido el 6 de noviembre de 2014; aceptado el 29 de marzo de 2015

Disponible en Internet el 26 de mayo de 2015

### PALABRAS CLAVE

Escalas pronósticas;  
Mortalidad hospitalaria;  
UCI;  
Disfunción multiorgánica;  
APACHE;  
SAPS;  
MPM

### Resumen

**Objetivo:** Valorar el grado en que las escalas pronósticas son capaces de predecir la mortalidad hospitalaria.

**Diseño:** Estudio de cohortes retrospectivo de 2 años.

**Ámbito:** UCI médico-quirúrgica de un hospital de tercer nivel.

**Pacientes:** Todos los pacientes con síndrome de disfunción multiorgánica en el primer día de ingreso.

**Variables de interés:** APACHE II y IV, SAPS II y III, MPM II y muerte hospitalaria.

**Resultados:** Se incluyeron 568 pacientes. Mortalidad hospitalaria: 39,8% (226 pacientes). Discriminación (área bajo la curva; IC 95%): APACHE IV (0,805; 0,751-0,858), SAPS II (0,755; 0,697-0,814), MPM II (0,748; 0,688-0,809), SAPS III (0,737; 0,675-0,799) y APACHE II (0,699; 0,633-0,765). El MPM II es el que mejor calibra, seguido por el SAPS III. APACHE II, SAPS II y APACHE IV presentan una muy mala calibración. Razón estandarizada de mortalidad (IC 95%): APACHE IV 1,9 (1,78-2,02); APACHE II 1,1 (1,07-1,13); SAPS III 1,1 (1,06-1,14); SAPS II 1,03 (1,01-1,05); MPM 0,9 (0,86-0,94).

**Conclusiones:** APACHE IV tiene la mejor capacidad discriminativa y mala calibración. MPM II tiene una buena discriminación y la mejor calibración. En cuanto al SAPS II, mantiene la segunda mejor discriminación y una mala calibración. El APACHE II muestra unos valores de calibración y discriminación que desaconsejarían su utilización en la actualidad, y el SAPS III mantiene una adecuada calibración y una discriminación moderada. La valoración de estos resultados podría marcar el inicio de nuevos estudios a nivel regional/nacional en determinadas poblaciones de pacientes críticos.

© 2015 Elsevier España, S.L.U. y SEMICYUC. Todos los derechos reservados.

\* Autor para correspondencia.

Correo electrónico: [mmmc16@gmail.com](mailto:mmmc16@gmail.com) (M. Sánchez-Casado).

**KEYWORDS**

Predictive scoring systems;  
Hospital mortality;  
ICU;  
Multi-organ failure;  
APACHE;  
SAPS;  
MPM

**Predictive scoring systems in multiorgan failure: A cohort study****Abstract**

**Objective:** An evaluation is made of the hospital mortality predicting capacity of the main predictive scoring systems.

**Design:** A 2-year retrospective cohort study was carried out.

**Setting:** A third level ICU with surgical and medical patients.

**Patients:** All patients with multiorgan failure during the first day in the ICU.

**Main variables:** APACHE II and IV, SAPS II and III, MPM II and hospital mortality.

**Results:** A total of 568 patients were included. Mortality rate: 39.8% (226 patients). Discrimination (area under the ROC curve; 95% CI): APACHE IV (0.805; 0.751-0.858), SAPS II (0.755; 0.697-0.814), MPM II (0.748; 0.688-0.809), SAPS III (0.737; 0.675-0.799) and APACHE II (0.699; 0.633-0.765). MPM II showed the best calibration, followed by SAPS III. APACHE II, SAPS II and APACHE IV showed very poor calibration. Standard mortality ratio (95% CI): APACHE IV 1.9 (1.78-2.02); APACHE II 1.1 (1.07-1.13); SAPS III 1.1 (1.06-1.14); SAPS II 1.03 (1.01-1.05); MPM 0.9 (0.86-0.94).

**Conclusions:** APACHE IV showed the best discrimination, with poor calibration. MPM II showed good discrimination and the best calibration. SAPS II, in turn, showed the second best discrimination, with poor calibration. The APACHE II calibration and discrimination values currently disadvised its use. SAPS III showed good calibration with modest discrimination. Future studies at regional or national level and in certain critically ill populations are needed.

© 2015 Elsevier España, S.L.U. and SEMICYUC. All rights reserved.

**Introducción**

La mejora de la calidad asistencial en el paciente crítico requiere la valoración mediante metodología cuantitativa de la estructura empleada, el proceso de cuidado y la evolución del paciente<sup>1</sup>. La evolución está claramente determinada por la gravedad inicial de la enfermedad. Medir la gravedad es una constante preocupación en el enfermo crítico, y tiene múltiples beneficios<sup>2</sup>. En primer lugar, nos ayuda a realizar un pronóstico, fundamentalmente en términos de mortalidad. Además, permite conocer el perfil de nuestros pacientes, con un impacto en la gestión de los recursos materiales y de personal utilizados. La valoración pronóstica, al permitir ajustar la mortalidad real a la esperada, elabora un indicador de calidad de los cuidados que facilita la comparación con otras instituciones (*benchmarking*)<sup>1</sup>. Por último, tiene una importante función al facilitar la investigación sobre los resultados, al asegurar que los pacientes con factores de gravedad similares sean comparados. Para estandarizar la medida de gravedad se diseñaron las *escalas de gravedad*.

En la actualidad se utilizan mayoritariamente 3 grandes grupos de escalas de gravedad, que han tenido diferentes versiones<sup>3</sup>. Son la escala *Acute Physiology and Chronic Health Evaluation* (APACHE), la *Simplified Acute Physiology Score* (SAPS) y la *Mortality Prediction Models* (MPM). Estas escalas transforman su valor numérico, mediante una función *logit*, en probabilidad de mortalidad hospitalaria, agrupando los pacientes en términos de probabilidad de muerte. Estas escalas predictivas surgen del análisis de las grandes series de cohortes de pacientes con diferentes enfermedades médico-quirúrgicas<sup>4</sup>. Sus limitaciones principales son su falta de predicción individual o su utilización sobre poblaciones de pacientes no incluidas en los estudios

elaborados para su diseño o grupos de pacientes con enfermedades concretas<sup>5</sup>.

Aunque existen muchos factores relacionados con la mortalidad (edad, comorbilidad, diagnóstico de admisión en la Unidad de Cuidados Intensivos [UCI], etc.), la gravedad de las anomalías fisiológicas son su determinante fundamental<sup>6</sup>. Dentro del enfermo crítico, este tipo de anomalías se presentan en su máxima expresión en el enfermo con síndrome de disfunción multiorgánica (SDMO). Estos pacientes presentan una alta mortalidad relacionada con el número de órganos afectados<sup>7</sup>. Aunque estos enfermos forman parte de la población de pacientes de las que son derivadas las escalas de gravedad (justificando alrededor de un tercio del total de los pacientes en una UVI de tercer nivel), no hemos encontrado estudios previos basados exclusivamente en ellos<sup>4</sup>.

El *objetivo* de nuestro estudio es valorar el grado en que las escalas pronósticas más utilizadas en la actualidad son capaces de predecir de forma real la mortalidad hospitalaria en la UCI en pacientes que desarrollan un SDMO en las primeras 24 h de ingreso.

**Pacientes y métodos**

Estudio de cohortes retrospectivo, realizado en la UCI del Hospital Virgen de la Salud, Toledo. Se analizaron durante 2 años de forma consecutiva las historias clínicas de los pacientes médico-quirúrgicos con SDMO, mayores de 16 años y con un tiempo mínimo de estancia en UCI de 12 h. Se define el SDMO como la presencia de al menos 2 órganos disfuncionantes, durante las primeras 24 h de ingreso en la UCI, objetivada mediante la escala SOFA<sup>8</sup>. Se excluyeron todos los pacientes con variables perdidas necesarias para el

cálculo de las escalas de gravedad, los pacientes remitidos de otras UCI con más de 24 h de evolución, aquellos con muerte encefálica y los reingresados en la UCI durante la misma estancia hospitalaria.

Los pacientes eran seguidos durante su evolución en UCI y planta hospitalaria, hasta su fallecimiento o alta, registrándose su situación final. Se recogían variables generales de ingreso (sexo, edad, presencia de infección, situación basal –asintomático para las actividades de la vida diaria–, procedencia, diagnóstico de ingreso, necesidad de ventilación mecánica, lugar de alta de UCI, estancias en UCI y hospitalarias, *exitus* y lugar del mismo) y aquellas necesarias para realizar las escalas de gravedad de APACHE II<sup>3</sup> y IV<sup>9</sup>, SAPS II<sup>10</sup> y III<sup>11</sup> (escala europea) y MPM II<sup>12</sup>. La MPM II utilizada es al valorada a las 24 h, para de esta forma ser comparada con la SAPS y la APACHE, pues las 3 se valorarían en las primeras 24 h.

Los datos cuantitativos son expresados como media (desviación estándar), y los datos categóricos, como recuento (porcentaje). Para realizar las comparaciones entre variables categóricas utilizamos el test de *chi* cuadrado, con la aproximación de Fisher para tablas 2 × 2 cuando se consideró indicado. Para realizar comparaciones de variables cuantitativas utilizamos la prueba *t de Student*.

Se calculó el valor de todas las escalas de puntuación y las probabilidades de muerte predichas. Para evaluar la validez del modelo estudiamos la discriminación y la calibración.

La discriminación es la capacidad de distinguir los sujetos que fallecen de los que sobreviven, describiendo la precisión de una predicción dada. Por ejemplo, si el modelo predice una mortalidad del 35%, la predicción es perfecta si la mortalidad observada es del 35%. Se valora mediante la curva ROC (intervalo de confianza del 95% [IC 95%]), y el área bajo la curva puede considerarse como una medida global del poder de discriminación del modelo.

La calibración del modelo es el aspecto del ajuste que valora la concordancia entre las probabilidades observadas en la muestra y las probabilidades predichas en los modelos, describiendo cómo se ejecuta la escala pronóstica sobre un amplio rango de mortalidades predichas. Así, una buena calibración es precisa en mortalidades del 10, 50 y 80%, por ejemplo. La hemos valorado mediante la prueba de *bondad de ajuste de Hosmer-Lemeshow*, que presenta su mejor ajuste cuando la probabilidad es igual a 1 (la no significación indica un mejor ajuste)<sup>13</sup>. Para llevarla a cabo se agrupan las probabilidades de fallecimiento en 10 grupos definidos de probabilidades de mortalidad (*deciles de riesgo*), intentando categorizar un número aproximado de pacientes en cada grupo.

Complementamos la valoración de la calibración mediante la realización de la *standardized mortality ratio* (SMR, «razón estandarizada de mortalidad»); es el cociente entre la mortalidad observada y la predicha. Si presenta un valor de 1 la calibración es buena con un acuerdo perfecto, pues las mortalidades observadas serían iguales a las predichas. Si es menor de 1, ocurre una sobrestimación de la mortalidad, y si es mayor de 1, lo contrario, se subestima la mortalidad.

El análisis estadístico se realizó con el programa SPSS® v. 20 (IBM Corporation, 2011). Se ha considerado significativo todo valor de  $p > 0,05$ .

**Tabla 1** Diagnósticos clínicos más frecuentes (> 1%)

Insuficiencia cardiaca congestiva	56 (9,8)
Neumonía comunitaria	50 (8,8)
Parada cardiorrespiratoria	36 (6,4)
Peritonitis secundaria	33 (5,8)
Shock cardiogénico isquémico	20 (3,5)
Shock hemorrágico postoperatorio	19 (3,3)
EPOC agudizada	18 (3,2)
Neumonía nosocomial	18 (3,2)
Sepsis urológica obstructiva	18 (3,2)
Sepsis de origen desconocido	17 (3)
Pancreatitis aguda	16 (2,8)
Isquemia intestinal	17 (2,6)
Infección de partes blandas	14 (2,5)
Sepsis urológica no obstructiva	13 (2,3)
Shock cardiogénico no isquémico	11 (1,9)
Hepatectomía	10 (1,8)
Aneurisma abdominal roto	10 (1,8)
Tromboembolismo pulmonar	9 (1,6)
Insuficiencia respiratoria médica	9 (1,6)
Obstrucción intestinal	9 (1,6)
Insuficiencia respiratoria postoperatoria	7 (1,2)
Insuficiencia respiratoria posquirúrgica	7 (1,2)
Cirugía abdominal mayor	7 (1,2)
Hemoperitoneo	7 (1,2)
Colecistitis aguda	6 (1,1)
Aneurisma aórtico abdominal roto	6 (1,1)
Colangitis	6 (1,1)
Cirugía maxilofacial	6 (1,1)
Shock hemorrágico perigestacional	6 (1,1)

EPOC: enfermedad pulmonar obstructiva crónica.

Los datos se expresan como n (%).

## Resultados

La población de estudio durante el periodo evaluado estaba formada por 568 pacientes, por cumplir las condiciones de inclusión y no presentar ninguna de exclusión. En la [tabla 1](#) figuran los diagnósticos más frecuentes. La mortalidad global hospitalaria fue del 39,8% (226 pacientes); 165 (29,1%) fallecieron en la UCI, y 61 (10,7%), en la sala hospitalaria, tras el alta post-UCI. Las características de ingreso, evolutivas y las escalas de gravedad se muestran en la [tabla 2](#).

La calibración, evaluada para las escalas de gravedad mediante la prueba de la bondad de ajuste de Hosmer-Lemeshow, se muestra en la [tabla 3](#). En la [figura 1](#) se muestra la valoración de las distintas curvas ROC según el modelo predictivo, y en la [tabla 4](#) se muestra el área bajo la curva, con el IC 95% y su significación para cada escala. En la [tabla 5](#) se determina la SMR para cada escala pronóstica con su IC 95%. En la [figura 2](#) se muestran las diferentes curvas de calibración.

## Discusión

En nuestra serie hemos incluido pacientes con SDMO en las primeras 24 h de ingreso. Este periodo es el habitualmente tomado por las escalas pronósticas para la valoración de

**Tabla 2** Descripción de las variables de ingreso, las evolutivas y las escalas de gravedad en el total de los pacientes, los que sobreviven y los fallecidos

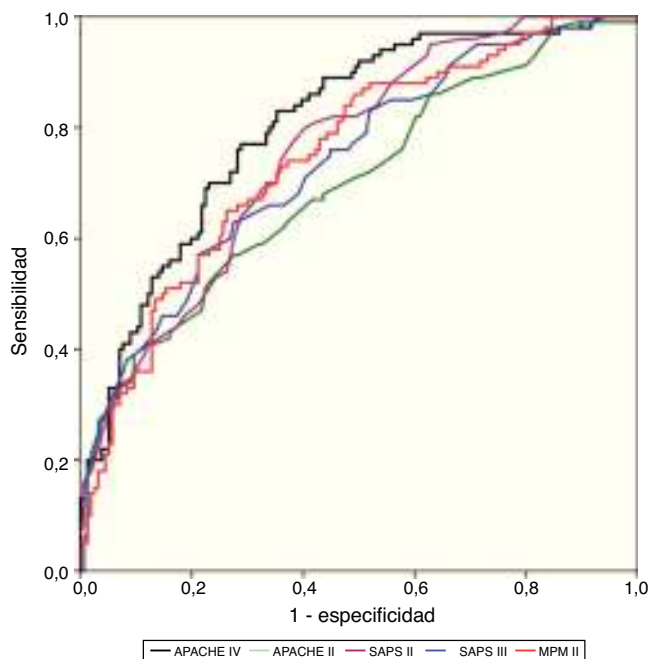
Variable	Muestra total (N = 568)	Supervivientes (n = 342)	Fallecidos (n = 226)
Sexo varón	367 (64,6%)	222 (64,9%)	145 (64,2%)
Edad (años)*	64 (16,4)	60,4 ± 17,4	69,6 ± 12,7
SFB: asintomático actividad diaria	415 (73,1%)	257 (75,1%)	157 (69,8%)
Procedencia**			
Planta	200 (35,1%)	99 (28,9%)	101 (44,7%)
Urgencias	194 (34,1%)	133 (38,9%)	60 (26,5%)
Quirófano	143 (25,2%)	90 (26,3%)	53 (23,5%)
Otro hospital	31 (5,5%)	20 (5,8%)	12 (5,3%)
Enfermo séptico	234 (41,1%)	141 (41,2%)	93 (41,2%)
Ventilación mecánica durante ingreso UCI*	448 (78,9%)	238 (69,6%)	211 (93,3%)
Estancia UCI (días)**	8,4 (12,8)	7,2 ± 8,9	10,2 ± 17
Estancia hospitalaria post-UCI (días)*	12,4 (19,9)	16,6 ± 18,4	6,2 ± 20,3
Lugar alta UCI			
Planta	383 (94,3%)	320 (93,6%)	62 (98,4%)
Otro hospital	23 (5,7%)	22 (6,4%)	1 (1,6%)
SOFA*	8 (3,4)	6,9 ± 2,8	9,7 ± 3,6
APACHE II*	18,6 (8,2)	16,3 ± 6,9	22,6 ± 8,8
APACHE IV*	60,8 (23,8)	52,3 ± 19,1	75 ± 24,2
SAPS II*	45,2 (18,2)	38,9 ± 14,5	55,3 ± 18,8
SAPS III*	63,5 (14,7)	59,1 ± 12,8	70,4 ± 14,9
MPM II*	43,2 (28,8)	33,7 ± 25,4	58,2 ± 27,3

APACHE: Acute Physiology and Chronic Health Evaluation; MPM: Mortality Prediction Model; SAPS: Simplified Acute Physiology Score; SFB: situación funcional basal; SOFA: Sequential Organ Failure Assessment; UCI: Unidad de Cuidados Intensivos.

Comparaciones entre el grupo de supervivientes y el de fallecidos:

\* p < 0,001.

\*\* p < 0,01.



**Figura 1** Valoración de la discriminación mediante la curva receiver operator characteristic -ROC- para los diferentes modelos predictivos.

los datos fisiológicos, pues coincide con el de mayor deterioro fisiológico. Aunque las predicciones que utilizan datos fisiológicos en la primera hora de la admisión han mostrado ausencia de diferencias estadísticamente significativas con los peores valores de las primeras 24 h<sup>14</sup>, en un periodo de 24 h se observan menos valores perdidos y un mayor poder explicativo<sup>15</sup>.

Nuestra serie muestra una alta mortalidad hospitalaria, falleciendo casi 4 de cada 10 enfermos, sobre todo en el periodo intra-UCI, altas estancias en UCI (más de una semana), hospitalarias post-UCI (alrededor de 2 semanas), con altos porcentajes de ventilación mecánica durante el ingreso (alrededor del 79%), y altas escalas pronósticas de disfunción (SOFA; en torno a 8). Esto es similar a lo publicado en otras series, con enfermos en estos estadios evolutivos y nivel de gravedad<sup>16</sup>.

Los diagnósticos más frecuentes son similares a los que existen en cualquier UCI polivalente (médico-quirúrgica). Los pacientes que fallecen presentan un perfil ya conocido, con más edad, procedencia de planta, mayor porcentaje de ventilación mecánica, mayor estancia en UCI, menor estancia post-UCI (posiblemente debido a su muerte), con mayor número de órganos disfuncionantes (casi un 30% más) y peor valor de las escalas pronósticas<sup>17</sup>. Para hacer un estudio más detallado de la validez de los modelos debemos estudiar la discriminación y la calibración. No debemos olvidar que una pobre calibración y discriminación puede deberse a una calidad diferente de los cuidados esperados, pero también a una mala aplicabilidad del modelo a una población dada de pacientes<sup>18</sup>.

**Tabla 3** Prueba de bondad de ajuste de Hosmer-Lemeshow en las diferentes escalas pronósticas

Deciles predictivos mortalidad (%)	N	Supervivientes observados	Supervivientes predichos	Muertes observadas	Muertes predichas	PBA H-L/gl/p
<b>APACHE IV (n = 471)</b>						
0-1,89	47	43	46,5	4	0,5	
1,9-3,65	47	44	45,7	3	1,3	
3,66-6	48	44	45,6	4	2,4	
6,04-8,97	47	36	43,5	11	3,5	
9,03-12,92	47	30	41,8	17	5,2	
12,94-17,49	47	30	39,9	17	7,1	
17,67-23,88	47	23	37,4	24	9,6	
24,04-33,95	47	21	33,7	26	13,3	130,54/gl = 8/p = 2,18624 × 10 <sup>-24</sup>
35,71-47,3	47	13	27,5	34	19,5	
48,11-97,18	47	9	16,1	38	30,9	
<b>APACHE II (n = 498)</b>						
4-9,9	52	48	48,2	4	3,8	
10-13	49	37	43,3	12	5,7	
14,6-17	49	36	41,2	13	7,8	
18,6-21	50	28	40	22	10	
23-26,2	50	34	37,5	16	12,5	
28,9-35,5	49	35	33,7	14	15,3	
36-42,4	46	31	28,1	15	17,9	
46-53	56	32	29,1	24	26,9	27,126/gl = 8/p = 0,0006721
53,3-70,3	48	21	17,8	27	30,2	
73-99,1	49	11	9,3	38	39,7	
<b>SAPS III (n = 257)</b>						
0,47-7,6	32	29	30,4	3	1,6	
8,5-12,7	24	20	21,4	4	2,6	
13,9-18,1	23	15	19,3	8	3,7	
19,7-24,7	25	17	19,5	8	5,5	
26,7-34,1	30	19	20,8	11	9,2	
36,1-40,2	21	15	13	6	8	
42,2-50,2	29	15	15,6	14	13,4	
52,2-57,9	22	14	9,8	8	12,2	10,6/gl = 8/p = 0,2254
59,7-66,4	26	9	9,9	17	16,1	
67,9-87,6	25	4	5,5	21	19,5	
<b>SAPS II (n = 258)</b>						
0-6	28	28	26,9	0	1,1	
7-11	25	21	22,7	4	2,3	
12-15	25	18	21,7	7	3,3	
17-21	25	18	20,2	7	4,8	
23-31	28	16	20,4	12	7,6	
33-40	24	14	15,3	10	8,7	
41-51	28	14	15,2	14	12,8	
53-66	25	14	10,2	11	14,8	44/gl = 8/p = 5,689 × 10 <sup>-7</sup>
68-83	26	11	6,7	15	19,3	
84-99	24	4	1,9	20	22,1	
<b>MPM II (n = 258)</b>						
1,02-6,72	26	24	25	2	1	
6,97-13,52	26	20	23,2	6	2,8	
13,79-22,55	26	22	21,2	4	4,8	
23,2-29,71	25	18	18,4	7	6,6	
30,06-38,82	26	17	17,1	9	8,9	
39,14-51,59	26	16	14,5	10	11,5	
51,6-59,88	26	15	11,3	11	14,7	
60,2-67,62	26	10	9,2	16	16,8	7,964/gl = 8/p = 0,43739
67,73-86,73	25	9	5,5	16	19,5	
87,55-99,58	26	7	1,2	19	24,8	

APACHE: *Acute Physiology and Chronic Health Evaluation*; gl: grados de libertad; MPM: *Mortality Prediction Model*; PBA H-L: prueba de bondad de ajuste de Hosmer-Lemeshow; SAPS: *Simplified Acute Physiology Score*.

**Tabla 4** Área bajo la curva con intervalo de confianza del 95%, con nivel de significación, para cada escala de gravedad evaluada

	Área	IC 95%		p
APACHE IV	0,805	0,751	0,858	0,00001
SAPS II	0,755	0,697	0,814	0,00001
MPM II	0,748	0,688	0,809	0,00001
SAPS III	0,737	0,675	0,799	0,00001
APACHE II	0,699	0,633	0,765	0,00001

APACHE: *Acute Physiology and Chronic Health Evaluation*; IC 95%: intervalo de confianza del 95%; MPM: *Mortality Prediction Model*; SAPS: *Simplified Acute Physiology Score*.

**Tabla 5** Determinación de la razón estandarizada de mortalidad y su intervalo de confianza del 95% para las escalas de gravedad evaluadas

	Número de pacientes	Supervivientes observados	Supervivientes esperados	Muertos observados	Muertos esperados	SMR	IC 95%
APACHE IV	471	293	343,6	178	93,3	1,9	1,78-2,02
APACHE II	498	313	328,2	185	169,8	1,1	1,07-1,13
SAPS III	257	157	165,2	100	91,8	1,1	1,06-1,14
SAPS II	258	158	161,2	100	96,8	1,03	1,01-1,05
MPM II	258	158	146,6	100	111,4	0,9	0,86-0,94

APACHE: *Acute Physiology and Chronic Health Evaluation*; IC 95%: intervalo de confianza del 95%; MPM: *Mortality Prediction Model*; SAPS: *Simplified Acute Physiology Score*; SMR: *standardized mortality ratio* («razón estandarizada de mortalidad»).

APACHE, SAPS y MPM son los 3 grupos de escalas más utilizados. Los problemas relacionados con la facilidad de utilización, el número de variables necesarias (142 en APACHE IV, 20 en SAPS III y 13 en MPM II 24 h, por ejemplo) y el pago por su utilización han limitado el uso de las versiones más recientes, mejoradas. Entre el conjunto de las diferentes versiones, la APACHE II es la más utilizada<sup>19</sup>. APACHE II manifestó sus imperfecciones en predecir subgrupos específicos de pacientes (fallo hepático, sepsis, etc.) y mostró que la mortalidad predicha es menor a la observada en los pacientes de UCI transferidos desde otras estancias hospitalarias (fenómeno conocido como *lead time bias*)<sup>20,21</sup> (por eso excluimos a los pacientes transferidos de otros centros). Algo similar pudiera ocurrir con SAPS II, que es más utilizada que SAPS III. Esto plantearía no solo las limitaciones propias de su menor complejidad (utilización de versiones «menos refinadas»), sino que utilizaríamos comparaciones de nuestros pacientes con otros de hace más de 2 décadas. Todas las escalas de gravedad requieren revisiones periódicas, recalibraciones y puestas al día en relación con la mejora de los cuidados (tratamiento y otros factores que influyen en la mortalidad, como los efectos de las nuevas tecnologías). Y estas revisiones deberían ser las utilizadas en los trabajos contemporáneos que se realizan. La no actualización de las escalas predictivas produciría una pérdida gradual de la calibración con el tiempo, resultando en una sobrestimación de la mortalidad para cualquier escala de gravedad dada<sup>22,23</sup>, fenómeno de sobrepronóstico que no hemos encontrado de forma clara en nuestro estudio.

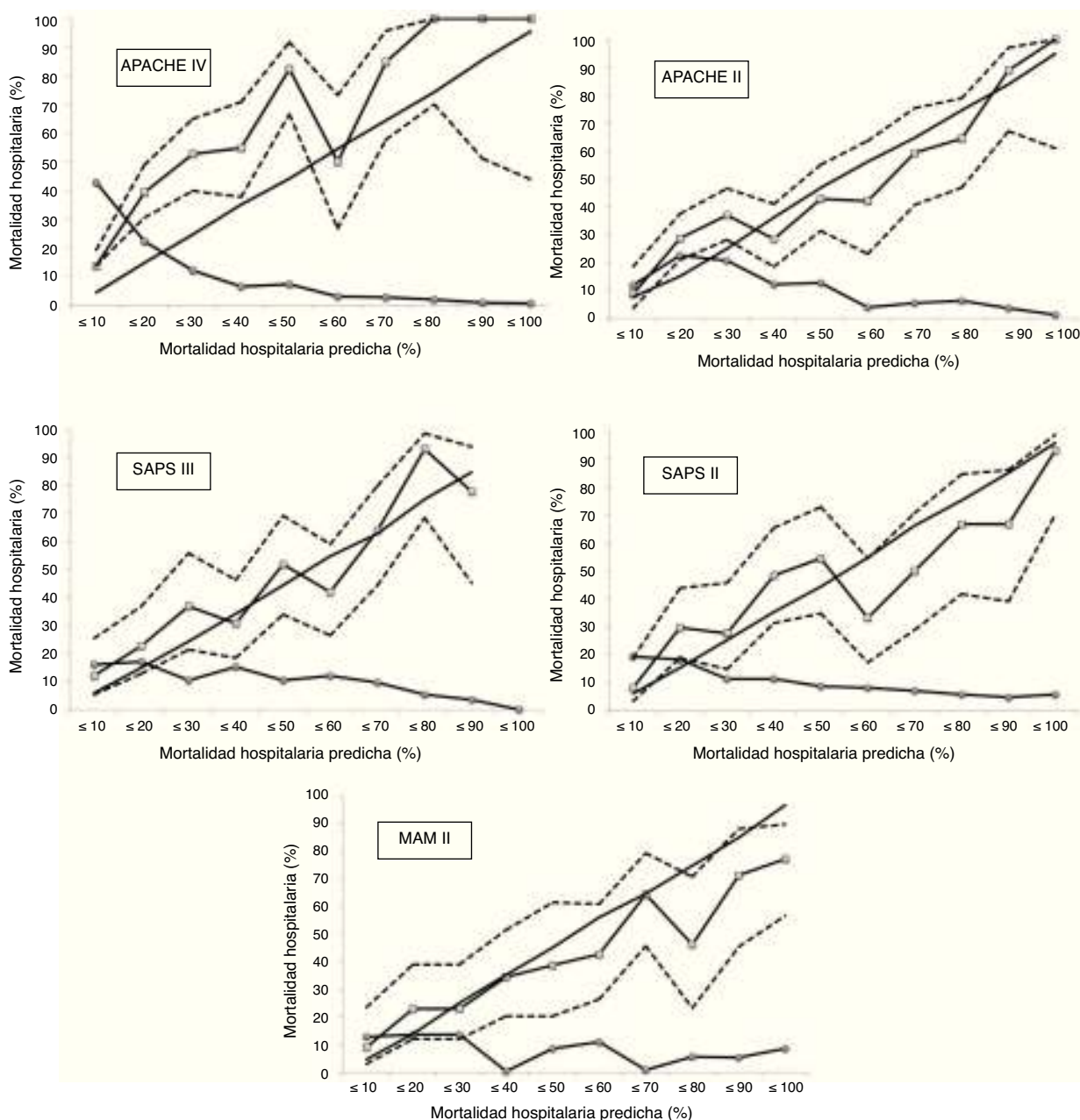
Hemos utilizado los principios más importantes para considerar en la valoración de una escala de gravedad, midiendo de forma eficaz mediante una adecuada discriminación y calibración el pronóstico en términos de mortalidad hospitalaria. Los sistemas predictivos han mostrado, en general,

una excelente discriminación, siendo sus problemas fundamentales a nivel de la calibración, sobre todo en el caso de la APACHE, pero que se mantiene en las demás escalas pronósticas (SAPS III, etc.)<sup>22</sup>.

Existen importantes diferencias entre las distintas escalas pronósticas en relación con la recogida de datos, el cálculo de la mortalidad, la eficacia y el coste<sup>24</sup>. Mientras que APACHE requiere la recogida de un amplio número de datos fisiológicos y generales, los otros instrumentos utilizan un menor número de variables y una codificación más sencilla, que facilita la recogida<sup>25</sup>. El paso de sistemas pronósticos simplificados a otros más complejos es un largo camino con numerosos problemas. Así, el desarrollo de la informática nos ha permitido la utilización de sistemas predictivos matemáticamente complejos, pero no se ha acompañado de desarrollos paralelos en la automatización del registro de los datos<sup>14</sup>. Esto nos dificulta la recogida en escalas pronósticas como la APACHE IV, donde la complejidad es máxima. Otros sistemas, como el MPM o el SAPS, en contraste, enfatizan su simplicidad. Existen pocos estudios que comparen las 3 escalas pronósticas; APACHE IV, en general, muestra la mejor precisión predictiva, aunque MPM (II y III) ha probado ser una buena alternativa eficaz cuando se considera un bajo coste y una menor complejidad de los datos recogidos<sup>25,26</sup>.

No existen estudios previos de utilización de escalas pronósticas en pacientes únicamente con SDMO, aunque forman parte, en mayor o menor proporción, de todas las series publicadas en UCI. De hecho, ocurre con frecuencia lo contrario, y así, por ejemplo, alrededor de dos tercios de la cohorte de pacientes de validación de APACHE IV están en bajo riesgo en el momento de la admisión (mortalidad < 10%)<sup>27,28</sup>.

La mejor discriminación la observamos en APACHE IV, obteniendo valores adecuados (0,805); SAPS II (0,755), MPM



**Figura 2** Curvas de calibración con el 95% del intervalo de confianza para las mortalidades observadas (línea con marcadores cuadrados con líneas discontinuas). La línea sin marcadores representa la mortalidad predicha; la línea con marcadores redondos representa el porcentaje de sujetos para cada decil de mortalidad.

II (0,748) y SAPS III (0,737) presentan una discriminación progresivamente peor, y ya en valores menos ajustados. APACHE II presenta los peores niveles de discriminación (0,699).

En nuestra serie, MPM II es el modelo pronóstico que mejor calibra, seguido por SAPS III. APACHE II, SAPS II y APACHE IV presentan una muy mala calibración, mostrando este último el peor resultado.

En relación con la SMR, todos presentan unos valores en torno a 1, menos APACHE IV, que con un valor de 1,9 (IC 95% 1,78-2,02) infraestima la mortalidad de forma más clara. De una forma general, APACHE II y SAPS III tienden a

subestimar la valoración, y MPM II a sobrestimarla, aunque todos, discretamente. La SMR es un parámetro de gestión clínica mandatorio en algunos países, tanto europeos como americanos (todas las UCI pertenecientes a la administración de veteranos), y se suele utilizar como una medida internacional de calidad<sup>29</sup>, aunque hay autores que argumentan una preocupación por su utilización en exclusiva como índice de calidad<sup>14</sup>. Muchos usuarios simplemente aspiran a una  $SMR < 1$ <sup>30</sup>. Pero es bien conocido, y a pesar de su amplia aceptación, que la SMR tiene algunas limitaciones intrínsecas<sup>30</sup>. Así, los pacientes de alto riesgo contribuyen desproporcionadamente a la SMR debido a que muchos de estos enfermos

fallecen. Por ello es importante comunicar por separado la SMR para pacientes de bajo, moderado y alto riesgo. Así, el impacto de una desproporción de pacientes de alto riesgo tendrá un efecto diluido y no dependerá tanto del tipo de enfermos visto en cada UCI. Esto es importante al ser cada vez más los trabajos que hablan sobre una menor precisión de la calibración en los extremos de severidad<sup>9,27,31</sup>.

Si analizamos las curvas de calibración, podemos establecer un grupo de mortalidad de moderado riesgo (entre 10 y 50% de mortalidad hospitalaria) y un grupo de alto riesgo (> 50% de mortalidad hospitalaria), que pueden ser analizados por separado<sup>27</sup>. APACHE IV es la escala pronóstica que presenta de una forma más intensa una infraestimación de la mortalidad predicha en el grupo de moderado riesgo; esto se mantiene en el grupo de alto riesgo, pero de una forma menos pronunciada. Observamos el mismo patrón, aunque menos intenso, en APACHE II y SAPS II, en ambos grupos de mortalidad. SAPS III solo tiende a infrapredicción en el grupo de moderado riesgo, comportándose bien en el grupo de alto riesgo. MPM II presenta un comportamiento diferente, mostrando una sobrepredicción, sobre todo en el grupo de alto riesgo. Es curioso que en la mayoría de las series de cohortes de enfermos, en las mayores probabilidades de muerte predicha, la mortalidad observada fue menor que la calculada<sup>18</sup>. En las curvas de calibración observamos que los IC 95% son muy amplios (menor precisión), por lo que probablemente necesitaríamos más pacientes para poder obtener resultados más ajustados y precisos.

Nuestro estudio tiene limitaciones. El tamaño muestral, aunque moderado, no es óptimo para estos estudios; ha sido realizado en una sola institución en España, fuera de la zona geográfica en la que algunas de estas escalas se validaron, lo cual repercute sobre la precisión del modelo<sup>32-34</sup>. Por otra parte, la utilización de escalas de gravedad que agrupen a los pacientes en enfermedades homogéneas, aunque tiene un interés que va creciendo con el tiempo, también es una limitación clara.

Es inevitable, y para muchos, deseable, que la predicción de escalas de riesgo se incremente con el tiempo, y lleve a ser juez de nuestras actividades clínicas<sup>35</sup>. Es importante buscar con un buen ajuste la máxima proximidad entre la realidad y el modelo propuesto. Para ello es vital evaluar las herramientas que utilizamos en nuestros pacientes, y hacerlo en los de mayor gravedad es algo de gran interés, pues tendemos a infravalorar la mortalidad real. Cada país debería asegurarse de que el modelo se ajusta a sus propios datos, pues una buena calibración del modelo produciría significados fiables y adecuadas probabilidades estimadas<sup>36</sup>. A falta de sistemas calibrados localmente, hemos realizado comparaciones indirectas de la evolución observada con el pronóstico esperado predicho por los diferentes modelos, aplicados en las poblaciones de enfermos utilizadas para desarrollar el modelo de riesgo. Esto sería importante realizarlo en la población de interés, para así poder implementar la valoración de la calidad utilizando la evolución en la UCI ajustada al riesgo, no solo valorando mortalidad, sino estancias, readmisiones en UCI u otros parámetros evolutivos<sup>37,38</sup>.

En nuestra serie de pacientes con SDMO corroboramos que APACHE IV tiene la mejor capacidad discriminativa, y confirmamos su mala calibración. MPM III no ha sido testado, pero comprobamos que MPM II tiene una buena discriminación y la mejor calibración de todos las escalas testadas.

En cuanto a SAPS II, mantiene la segunda mejor discriminación y una mala calibración. APACHE II muestra unos valores de calibración y discriminación que desaconsejarían su utilización en la actualidad, y SAPS III mantiene una adecuada calibración y una discriminación moderada. Pensamos que la valoración de estos resultados podría marcar el inicio de nuevos estudios más amplios y multicéntricos en determinadas poblaciones de pacientes críticos.

## Financiación

No se ha utilizado ningún tipo de financiación, parcial o total, para este artículo.

## Conflicto de intereses

No existen relaciones económicas ni personales que puedan sesgar o influir en nuestras actuaciones.

## Bibliografía

1. Power GS, Harrison DA. Why try to predict ICU outcomes? *Curr Opin Crit Care*. 2014;20:544-9.
2. Conry MC, Humphries N, Morgan K, McGowan Y, Montgomery A, Vedhara K, et al. A 10 year (2000-2010) systematic review of interventions to improve quality of care in hospitals. *BMC Health Serv Res*. 2012;12:275.
3. Kahn JM. Predicting outcome in critical care: Past, present and future. *Curr Opin Crit Care*. 2014;20:542-3.
4. Sikka P, Jaafar WM, Bozkanat E, El-Solh AA. A comparison of severity of illness scoring systems for elderly patients with severe pneumonia. *Intensive Care Med*. 2000;26:1803-10.
5. Salluh JI, Soares M. ICU severity of illness scores: APACHE, SAPS and MPM. *Curr Opin Crit Care*. 2014;20:557-65.
6. Scheffler RM, Knaus WA, Wagner DP, Zimmerman JE. Severity of illness and the relationship between intensive care and survival. *Am J Public Health*. 1982;72:449-54.
7. Ferreira AM, Sakr Y. Organ dysfunction: General approach, epidemiology, and organ failure scores. *Semin Respir Crit Care Med*. 2011;32:543-51.
8. Vincent JL, Moreno R, Takala J, Willatts S, de Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med*. 1996;22:707-10.
9. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. *Crit Care Med*. 2006;34:1297-310.
10. 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;270:2957-63.
11. Ledoux D, Canivet JL, Preiser JC, Lefrancq J, Damas P. SAPS 3 admission score: An external validation in a general intensive care population. *Intensive Care Med*. 2008;34:1873-7.
12. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. *JAMA*. 1993;270:2478-86.
13. Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol*. 1982;115:92-106.



14. Zimmerman JE, Kramer AA. A history of outcome prediction in the ICU. *Curr Opin Crit Care*. 2014;20:550–6.
15. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: A severity of disease-classification system. *Crit Care Med*. 1985;13:818–29.
16. Fry DE. Multiple organ dysfunction syndrome: Past, present and future. *Surg Infect (Larchmt)*. 2000;1:155–61.
17. Rehman MF, Siddiqui MS. Predicting death and disability, is it really possible? A medical ICU prognostication model study. *Crit Care Med*. 2014;42:2449–50.
18. Patel PA, Grant BJ. Application of mortality prediction to individual intensive care units. *Intensive Care Med*. 1999;25:977–82.
19. Cowen JS, Kelley MA. Errors and bias in using predictive scoring systems. *Crit Care Clin*. 1994;10:53–72.
20. Escarce JJ, Kelley MA. Admission source to the medical intensive care unit predicts hospital death independent of APACHE II score. *JAMA*. 1990;264:2389–94.
21. Capuzzo M, Valpioni V, Sgarbi A, Bortolazzi S, Pavoni V, Gilli G, et al. Validation of severity scoring systems SAPS II and APACHE II in a single-center population. *Intensive Care Med*. 2000;26:1779–85.
22. Nassar AP Jr, Mocelin AO, Nunes AL, Giannini FP, Brauer L, Andrade FM, et al. Caution when using prognostic models: A prospective comparison of 3 recent prognostic models. *J Crit Care*. 2012;27:423.
23. Kramer AA. Predictive mortality models are not like fine wine. *Crit Care*. 2005;9:636–7.
24. Glance LG, Osler TM, Dick A. Rating the quality of intensive care units: Is it a function of the intensive care unit scoring system? *Crit Care Med*. 2002;30:1976–82.
25. Kuzniewicz MW, Vasilevskis EE, Lane R, Dean ML, Trivedi NG, Rennie DJ, et al. Variation in ICU risk-adjusted mortality: Impact of methods of assessment and potential confounders. *Chest*. 2008;133:1319–27.
26. Rojas-Suarez J, Paternina-Caicedo AJ, Miranda J, Mendoza R, Dueñas-Castel C, Bourjeily G. Comparison of severity-of-illness scores in critically ill obstetric patients: A 6-year retrospective cohort. *Crit Care Med*. 2014;42:1047–54.
27. Breslow MJ, Badawi O. Severity scoring in the critically ill: Part 1-Interpretation and accuracy of outcome prediction scoring systems. *Chest*. 2012;141:245–52.
28. Niewiński G, Starczewska M, Kański A. Prognostic scoring systems for mortality in intensive care units-The APACHE model. *Anaesthesiol Intensive Ther*. 2014;46:46–9.
29. Flaatten H. The present use of quality indicators in the intensive care unit. *Acta Anaesthesiol Scand*. 2012;56:1078–83.
30. Breslow MJ, Badawi O. Severity scoring in the critically ill: Part 2: Maximizing value from outcome prediction scoring systems. *Chest*. 2012;141:518–27.
31. Beck DH, Smith GB, Taylor BL. The impact of low-risk intensive care admissions on mortality probabilities by SAPS II, APACHE II and APACHE III. *Anaesthesia*. 2002;57:21–6.
32. Harrison DA, Brady AR, Parry GJ, Carpenter JR, Rowan K. Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med*. 2006;34:1378–88.
33. Brinkman S, Abu-Hanna A, de Jonge E, de Keizer NF. Prediction of long-term mortality in ICU patients: Model validation and assessing the effect of using in-hospital versus long-term mortality on benchmarking. *Intensive Care Med*. 2013;39:1925–31.
34. Moran JL, Bristow P, Solomon PJ, George C, Hart GK. Mortality and length-of-stay outcomes, 1993-2003, in the binational Australian and New Zealand intensive care adult patient database. *Crit Care Med*. 2008;36:46–61.
35. Angus DC, Pinsky MR. Risk prediction: Judging the judges. *Intensive Care Med*. 1997;23:363–5.
36. Zimmerman JE, Kramer AA, Knaus WA. Changes in hospital mortality for United States intensive care unit admissions from 1988 to 2012. *Crit Care*. 2013;17:R81.
37. Kramer AA, Higgins TL, Zimmerman JE. Intensive care unit readmissions in U. S. hospitals: Patient characteristics, risk factors, and outcomes. *Crit Care Med*. 2012;40:3–10.
38. Kramer AA, Higgins TL, Zimmerman JE. The association between ICU readmission rate and patient outcomes. *Crit Care Med*. 2013;41:24–33.