



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE MEDICINA

División de Estudios Superiores de Posgrado
Hospital Español de México

**“Validación de dos modelos predictivos de
mortalidad: SAPS3 y MPM3.”**

TESIS DE POSGRADO

para obtener el título de Médico Especialista en
Medicina del Enfermo en Estado Crítico

Presenta:

Anaid Manzano Vergara

Asesor de Tesis:

Dr. Ulises W Cerón Díaz.

Noviembre 2013



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

ÌNDICE

Introducción	1
Marco teórico	2
Justificación	9
Objetivos	9
Hipótesis	9
Diseño	10
Material y Métodos	10
Resultados	11
Discusión	18
Conclusiones	19
Bibliografía	20

Introducción

Los sistemas de puntuación de la UCI fueron introducidos hace casi 30 años con el objetivo de la utilización de datos fisiológicos disponibles al ingreso en la UCI para predecir los resultados individuales de los pacientes. A pesar de estas predicciones tienen poca utilidad para el manejo de pacientes individuales, proporcionan un mecanismo para evaluar el desempeño de la UCI mediante la comparación de las cifras observadas en una población dada y los resultados observados en la población de referencia utilizado para su desarrollo algoritmos de predicción.

La adopción de un modelo predictivo requiere de un trabajo de validación local para asegurar que su aplicación se ajuste a la población de pacientes atendidos.

Uno de los grandes avances en la medicina crítica ha sido la creación de escalas generales de calificación de gravedad y modelos predictivos de mortalidad. Para que estos modelos cumplan su función de una manera adecuada, han sido necesarios estudios que confirmen su adecuado rendimiento en diferentes UTÍ's (Unidades de Cuidados Intensivos), ya que desde sus orígenes han sido creadas con grupos de pacientes con características demográficas diferentes, razón por la cual es necesario comprobar su validez con el grupo de enfermos que el médico intensivista maneja día con día.

Marco teórico

Los tres sistemas de puntuación más utilizados son APACHE, MPM y SAPS.³⁻⁷ Aunque las primeras versiones fueron introducidos hace muchos años, cada uno ha sido objeto de múltiples revisiones en el siguiente años.⁸⁻¹⁰ Además de los modelos actualizados regularmente introducidas por los desarrolladores, otros investigadores han creado modelos alternativos de pronóstico utilizando APACHE, SAPS, y elementos de datos MPM que pueden personalizarse para adaptarse a una población de paciente en específico.¹¹⁻¹⁵ Aunque algunos grupos han utilizado el mismas ponderaciones variables que se utilizaron en el modelo original y simplemente cambiaron las ecuaciones de regresión para adaptarse mejor a su población de pacientes, otros han obtenido nuevos pesos y creado nuevas ecuaciones de regresión. Es importante reconocer que estos nuevos modelos no son superiores a las originales, porque no han incluido todos los elementos nuevos de datos o refinamientos. Simplemente tienen un modelo modificado para ajustarse mejor a su propia población.

A pesar de muchas similitudes, hay diferencias importantes entre los tres sistemas. En primer lugar, APACHE IV es una herramienta propia, a pesar de Cerner Corporation ha hecho recientemente la mortalidad hospitalaria y en la UCI los algoritmos están disponibles gratuitamente ([https://apachefoundations.cernerworks.com / apachefoundations / login / auth](https://apachefoundations.cernerworks.com/login/auth)).

MPM0-III y SAPS 3 se han centrado en simplificar la recopilación de datos, el uso de un menor número de elementos y sólo los datos de la primera hora en la UCI, en comparación con APACHE, que utiliza los datos de todo el primer día. APACHE y MPM se desarrollaron principalmente con pacientes atendidos en la Unidades de Cuidados Intensivos de Estados Unidos de Norteamérica, mientras que el SAPS 3 incluyó pacientes

de 35 países.³⁻⁷ APACHE IV y MPM0-III utiliza grandes conjuntos de datos para su desarrollo (131 618 y 124 885, respectivamente), mientras que la SAPS 3 utilizó datos de 22 791 pacientes. APACHE es también el único sistema de puntuación que se ha demostrado una buena discriminación y calibración en la predicción de la Unidades de Cuidados Intensivos y la estancia hospitalaria (LOS) para las Unidades de Cuidados Intensivos (UCI) de Norteamérica. (MPM recientemente se ha demostrado predecir UCI adecuadamente la estancia en UCI en hospitales de California.¹³) Teniendo en cuenta el enfoque actual y futuro en la reducción de los costos de atención de la salud, la evaluación del desempeño LOS es de considerable importancia.

A pesar de una mayor atención en la seguridad de la atención sanitaria y el impulso de una mayor transparencia, sistemas de puntuación de la UCI no se utilizan en muchas UCIs norteamericanas. Aunque algunos países han convergido en una sola metodología y uso es obligatorio para todas las unidades de cuidados intensivos, no hay esfuerzos coordinados similares en los Estados Unidos para exigir reportes universales reportando calidad de UCI.^{11, 12}

APACHE se utiliza actualmente con más frecuencia que otros sistemas de puntuación en los Estados Unidos. Sin embargo, incluso con estas salvedades, parece probable que los sistemas de puntuación de la UCI se utilizan actualmente en un 10% a 15% de los pacientes de la UCI de Estados Unidos. El Foro Nacional de la Calidad también aprobó el uso de MPM a finales de 2010 (www.qualityforum.org). MPM se seleccionó sobre la base de tener un conjunto de datos requeridos más pequeño.

Antes de discutir los resultados de los diferentes sistemas, varios conceptos llevan breve revisión. Primero, es importante entender cómo se desarrollan y evalúan los sistemas de calificación de riesgo. Para generar un modelo predictivo, la mayoría de los investigadores dividen a su conjunto de datos en dos grupos. El primero se utiliza para desarrollar el modelo, por lo general a través de los métodos de regresión multivariable,

mientras que la segunda está reservada para la validación después de que se ha desarrollado el modelo final. Idealmente, el rendimiento en los conjuntos de desarrollo y validación es similar, y el modelo se puede aplicar a poblaciones más allá de la utilizada durante el desarrollo.

Para evaluar el desempeño, los investigadores se centran generalmente en la discriminación y la calibración. Es una práctica estándar cuando se describe la discriminación reportar el área bajo la curva característica de funcionamiento del receptor (ROC), que es una representación gráfica de la sensibilidad contra tasa de falsos positivos.¹⁹ Para los modelos de la UCI que predicen la mortalidad, este representa la probabilidad de que un paciente seleccionado al azar que muere tiene un riesgo estimado superior a un paciente seleccionado al azar que sobrevive.

Un ROC de 0,5 no es mejor que el azar, mientras que los valores. 0,7, 0,8 y 0,9 se consideran aceptables, excelente, y en circulación, respectivamente.²⁰ Para medir la calibración, esta examina qué tan bien los resultados reales se ajustan a su incidencia prevista, el estadístico de Hosmer-Lemeshow C es a menudo utilizado.²⁰ La forma más común de hacer esto es dividir la muestra en deciles de riesgo y evaluar el número real y el previsto de eventos en cada grupo. La prueba es esencialmente una prueba de χ^2 global que examina si existen diferencias significativas entre los resultados reales y previstos a través de grupos. No es de sorprender, que la influencia del tamaño de la muestra de entidades de crédito, las diferencias significativas son comunes cuando grandes tamaños de muestras son examinados.²¹ La calibración sólo es útil en el contexto de la discriminación, los modelos predictivos que se realizan no son mejores que el azar realizan de forma coherente en todos los estratos de riesgo y presentan una excelente calibración. La calibración también tiene relevancia en los diferentes tipos de UCI, diagnósticos de ingreso y regiones geográficas. Un sistema de puntuación que es bien realizado en un gran conjunto de datos heterogéneos, pero mal realizado en poblaciones

específicas de pacientes dando lugar a resultados poco fiables en la UCI. Las preocupaciones acerca de la calibración regional condujo a la creación de modelos de calificación del sistema por los gobiernos europeos utilizando sólo los datos de sus propios pacientes. Este enfoque era en gran medida el resultado de una mala calibración del los algoritmos de APACHE II cuando se utiliza en Gran Bretaña ICUs.²² Si el problema observado con la calibración refleja las diferencias regionales en la atención, la diferencia en la mezcla de la agudeza o el uso de algoritmos obsoletas refleja el rendimiento de poblaciones de mayor edad de pacientes determinados. Sin embargo, la decisión de adoptar modelos regionales impide la comparación de resultados a través de las fronteras nacionales y tiene dificultad para la identificación de los procesos de atención superiores, porque no hay una herramienta de comparación universal.

Al evaluar el rendimiento UCI, el enfoque más común consiste en calcular la relación del número de defunciones observadas con el número de muertes predichos por el sistema de puntuación de referencia. Esto se conoce como la tasa de mortalidad estandarizada (SMR). Del mismo modo, las tasas reales de predichos se utilizan para evaluar LOS desempeño. Aunque la mayoría de los usuarios se centran exclusivamente en las relaciones reales a predichos reportados, es importante tener en cuenta también las iniciativas comunitarias para estas tasas. Es poco probable que el punto de estimación SMR representa exactamente el valor verdadero, porque siempre hay un cierto grado de error aleatorio; IC proporciona información sobre el grado de precisión de la tasa observada.²³

Existe una extensa información en la literatura que evalúa la precisión de los tres sistemas principales UCI de puntuación y numerosos estudios que comparan una a otro.^{1, 2, 13, 14} Todos demuestran buena discriminación, con APACHE IV generalmente tiene un área ligeramente más alta del área bajo la curva ROC (Tabla 3). Algunos estudios sugieren que la calibración en los extremos de la gravedad puede ser menos accurate.^{3, 24}

Las variaciones en la duración de la atención recibida antes de la admisión a la UCI pueden introducir anticipadamente sesgo, que se ha demostrado que afectan la tasa de mortalidad ajustada.²

En la interpretación de los datos del sistema de puntuación, es importante considerar si los procesos de la UCI pueden diferir de la población de referencia de la UCI. Varios investigadores han sugerido que la UCI y las prácticas de alta hospitalaria pueden afectar el rendimiento de los scores en UCI, hay hospitales que pueden transferir fácilmente a los pacientes a otros hospitales o sitios alternativos de atención, tales como las unidades de cuidados agudos a largo plazo, tendrá estancias hospitalarias más cortas y potencialmente disminuido las tasas de mortalidad que las instalaciones que siguen la atención a estos pacientes. Vasilevskis et al encontraron una correlación entre las transferencias de la UCI a otros hospitales y observó SMR (cada aumento del 1% en las transferencias se asoció con una disminución de 0,02 SMR). Los pacientes trasladados a la UCI de otro hospital parecerían tener estancias más cortas y menos muertes que si se completara la estancia en la UCI originalmente. Esto ocurre con mayor frecuencia en los hospitales de la comunidad de escasos recursos de cuidados críticos y potencialmente podría resultar en una reducción sustancial de las tasas de mortalidad real a la pronosticada. Examinando los datos de ubicación de descarga puede ayudar al personal de calidad para evaluar si las transferencias de la UCI están bajando artificialmente las tasas de mortalidad.²⁴

Justificación

La adopción de un modelo predictivo requiere de un trabajo de validación local para asegurar que su rendimiento se ajusta a la población de pacientes atendidos.

Objetivos

Determinar la capacidad discriminativa y calibración de los sistemas predictores de mortalidad MPM3 y SAPS 3 en pacientes críticamente enfermos en la Unidad de Terapia Intensiva “Dr. Alberto Villazón Sahagún” del Hospital Español de México.

Hipótesis

Nula:

- El rendimiento de SAPS 3 y MPM3 es igual para predecir mortalidad hospitalaria en la UTI «Dr. Alberto Villazón Sahagún» del Hospital Español de México.

Alterna:

- El rendimiento de SAPS 3 para predecir mortalidad hospitalaria en la UTI «Dr. Alberto Villazón Sahagún» del Hospital Español de México es superior al rendimiento de MPM3.

Diseño

Es un estudio descriptivo transversal prospectivo de cohorte.

Material y métodos

El estudio incluyó a mil enfermos que ingresaron a la UTI. A todos los pacientes se les calculó la probabilidad de morir utilizando los modelos matemáticos SAPS 3 y MPM3, al ingreso.

- Criterios de inclusión:
 - Mayores de 18 años de edad.
 - Incluidos en el registro de la base de datos.
- Criterios de exclusión:
 - Operados de corazón
 - Enfermos coronarios
 - Traslados a otro hospital.
 - Reingresos

El recuento y análisis de los datos se hicieron a través de programas creados en Excel y mediante el programa estadístico SPSS versión 14.0. El rendimiento de ambos modelos se analizó a través de evaluar:

- Capacidad discriminativa
- Calibración.

Se evaluó la capacidad discriminativa entre vivos y muertos, a través de la construcción de curvas ROC (Receiver Operating Characteristic) mediante el programa estadístico SPSS.

- Se calculó el área por debajo de la curva y se estableció si existía diferencia estadística entre los modelos.

Para evaluar la calibración, se calculó los estadísticos «C y H» de la prueba de bondad de ajuste de Lemeshow y Hosmer, y se analizó la correlación entre la mortalidad predicha y la observada para cada uno de los modelos; los resultados se expresan en coeficiente de correlación de Pearson y r^2 .

Resultados

De la base de datos de nuestra Unidad se tomaron los primeros 1000 registros, de los cuales se excluyeron 175, debido a que cumplieron los criterios de exclusión (Ver tabla 1).

<i>Criterios de exclusión</i>	<i>No.</i>
Trasladados a otro hospital	47
Hospitalizado	1
Cirugía cardíaca	57
Menores de 18 años	5
Reingresos	65
Total	175

Tabla 1. Pacientes excluidos del estudio.

n=	825
Sexo (M/F)	416(50.4)/409(49.5)
Edad (X ± DE)	63+-19
Condición antes del ingreso □	
• Desconocida	10 (1.2)
• Encamado	86 (10.4)
• < 50%	39 (4.7)
• Sintomático	372 (45)
• Normal	318 (38.5)
Lugar de procedencia	
• Urgencias	265 (32.1)
• Hospitalización	265 (32.1)
• Quirófano	236 (28.6)
• Gineco-obstetricia	12 (1.4)
• Sala de Recuperación	15 (1.8)
• Unidad Coronaria	16 (1.9)
• Otro Hospital	7 (0.8)
• Otra UTI	6 (0.7)
• Diálisis	3 (0.3)
• Cirugía ambulatoria	1 (0.1)
Factor determinante de ingreso	
	149 (18)
• Amenaza de falla	39 (4.7)
• Trastorno Hidrometabolico	3 (0.3)
• Intoxicados	13 (1.5)
• Monitoreo Hemodinámico	214 (25.9)
• Posoperados	10 (1.2)
• Quirúrgicos de alto riesgo	38 (4.6)
• Post RCP	156 (18.9)
• Estado de Choque	203 (24.6)
• Ventilación mecánica	
Egreso de UTI	676 (81.9)
• Vivos	10 (1.2)
• RCP II	13 (1.5)
• RCP III	126 (15.2)
• Muertos	

Tabla 2. Datos Demográficos de la población.

El estudio se realizó con 825 registros de pacientes, 419 (50.4%) pacientes eran de género femenino y 426 (49.5) masculinos; con edad promedio de 63[±].19. Condicionantes del internamiento fue: normal en 318 (38.5%) pacientes, sintomático 372(45%) pacientes, deambula <50% 39 (4.7%) pacientes, encamado 86 (10.4%) pacientes y desconocido 10 (1.2%) pacientes. El lugar de procedencia de los pacientes: 265 (32.1%) pacientes del servicio de urgencias, 16 (1.9%) de la Unidad Coronaria, 15 (1.8%) de sala de Recuperación, 236 (28.6%) de quirófano, 265 (32.1%) de Piso de Hospitalización, 7 (0.8%) procedentes de otro hospital, 6 (0.7%) de otra UTI, 12(1.4%) del servicio de Ginecoobstetricia, 3 (0.3%) del servicio de Diálisis y uno (0.1%)de cirugía ambulatoria. El factor determinante de ingreso fue: amenaza de falla en 149 (18%) pacientes, 39 (4.7%) con trastorno Hidrometabólico, 3 (0.3%) intoxicados, 13 (1.5%) con necesidad de monitoreo hemodinámico, 214 (25.9%) posoperados, 10 (1.2%) quirúrgicos de alto riesgo, 38 (4.6%) post RCP, 156 (18.9%)en estado de choque y 203 (24.6%) con necesidad de ventilación mecánica. De estos pacientes, 676 (81.9%) fueron dados de alta vivos de UTI, 10 (1.2%) pacientes fueron dados de alta como RCPII, 13 (1.5%) pacientes como RCPIII y 126 (15.2%) pacientes murieron al alta de UTI (Ver Tabla 2).

En las tablas 3 y 4 se presentan los resultados de la prueba de Bondad de ajuste de Lemeshow y Hosmer, através del estadístico "C"; para ello se dividieron los enfermos en deciles de acuerdo a la probabilidad de morir. Para SAPS 3 (tabla 2) el estadístico C obtuvo un valor de 42.15, superior al valor crítico de Chi ²y por lo tanto una asociación estadísticamente significativa. En la tabla 3 se presentan los datos correspondientes para MPM3; observe que el valor del estadístico C obtuvo un valor de 140.3, superior al valor crítico de Chi ²y también existe una asociación estadísticamente significativa con un valor de p (0.05).

	"n"	Muertos	Vivos	Muer Esp	Viv Esp	Mort Act	Mort Esp	Mu-MuE	Vi-VivE
1	83	4	79	1.00	82.00	4.82	1.20	9.00	0.11
2	83	7	76	2.62	80.38	8.43	3.16	7.32	0.24
3	83	6	77	5.22	77.78	7.23	6.29	0.12	0.01
4	83	5	78	8.76	74.24	6.02	10.55	1.61	0.19
5	83	9	74	13.58	69.42	10.84	16.36	1.54	0.30
6	82	19	63	21.46	60.54	23.17	26.17	0.28	0.10
7	82	23	59	29.28	52.72	28.05	35.71	1.35	0.75
8	82	24	58	38.79	43.21	29.27	47.30	5.64	5.06
9	82	43	39	53.27	28.73	52.44	64.96	1.98	3.67
10	82	63	19	68.66	13.34	76.83	83.73	0.47	2.40
								29.31	12.83
								C	42.15
								gL	8
								X crítica	15

Tabla 3. Prueba de Bondad de Lemeshow y Hosmer para SAPS 3

	"n"	Muertos	Vivos	Muer Esp	Viv Esp	Mort Act	Mort Esp	Mu-MuE	Vi-VivE
1	83	4	79	0.19	82.81	4.82	0.23	76.00	0.18
2	83	7	76	1.21	81.79	8.43	1.46	27.72	0.41
3	83	6	77	2.35	80.65	7.23	2.83	5.66	0.17
4	83	5	78	3.93	79.07	7.00	4.74	0.29	0.01
5	83	9	74	5.84	77.16	15.00	7.03	1.71	0.13
6	82	19	63	8.93	73.07	22.00	10.89	11.34	1.39
7	82	23	59	13.33	68.67	30.00	16.26	7.01	1.36
8	82	24	58	23.03	58.97	26.00	28.08	0.04	0.02
9	82	43	39	43.31	38.69	46.00	52.81	0.00	0.00
10	82	63	19	71.08	10.92	75.00	86.69	0.92	5.98
								130.70	9.64
								C	140.35
								gL	8
								X crítica	15

Tabla 4. Prueba de bondad de Lemeshow y Hosmer para MPM3

En las Tablas 5 y 6 se presentan los resultados de la prueba de Bondad de ajuste de Lemeshow y Hosmer, a través del estadístico "H"; para ello se hicieron 10 grupos que incluyeron enfermos con probabilidad de morir en intervalos de clase progresivamente mayores de acuerdo a la probabilidad de morir. Para SAPS 3 (tabla 5) el estadístico H obtuvo un valor de 38.68, superior al valor crítico de χ^2 y por lo tanto una asociación estadísticamente significativa. En la tabla 6 se presentan los datos correspondientes para MPM3; observe que el estadístico H obtuvo un valor de 64.16 superior al valor crítico de

Chi ²el valor de Chi 2 es de 140 y también existe una asociación estadísticamente significativa con un valor de p (0.05).

	"n"	Muertos	Vivos	Muer Esp	Viv Esp	Mort Act	Mort Esp	Mu-MuE	Vi-VivE
1	268	18	250	10.53	257.47	6.72	3.93	5.30	0.22
2	138	12	126	18.85	119.15	8.70	13.66	2.49	0.39
3	76	14	62	18.68	57.32	18.42	24.58	1.17	0.38
4	84	26	58	28.66	55.34	30.95	34.12	0.25	0.13
5	63	19	44	27.23	35.77	30.16	43.22	2.49	1.89
6	44	15	29	23.66	20.34	34.09	53.77	3.17	3.69
7	51	24	27	32.66	18.34	47.06	64.04	2.30	4.09
8	44	32	12	32.98	11.02	72.73	74.95	0.03	0.09
9	45	31	14	38.36	6.64	68.89	85.24	1.41	8.16
10	12	12	0	11.03	0.97	100.00	91.92	0.09	0.97
								18.69	20.01
H								38.69	
gL								8.00	
X crítica								15.00	

Tabla 5. Estadístico H (SAPS3).

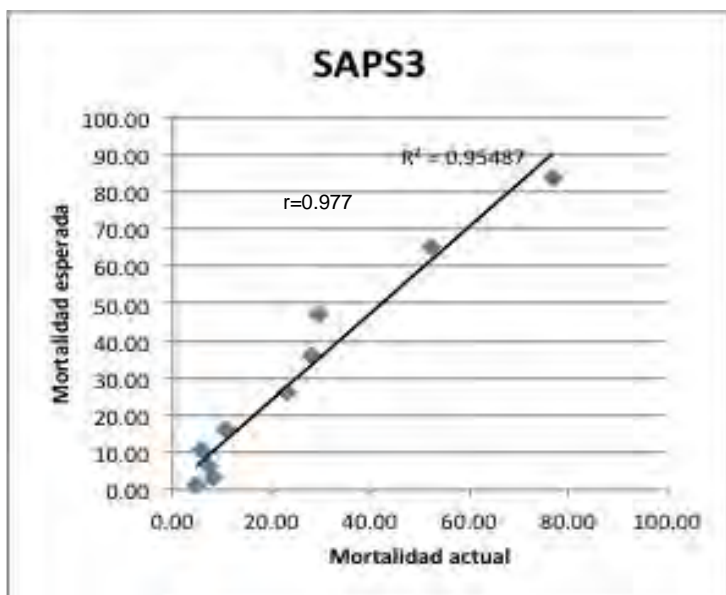
Tabla 6. Estadístico H (MPM3).

	"n"	Muertos	Vivos	Muer Esp	Viv Esp	Mort Act	Mort Esp	Mu-MuE	Vi-VivE
1	431	35	396	18.61	412.39	8.12	4.32	14.44	0.65
2	134	35	99	18.57	115.43	26.12	13.86	14.54	2.34
3	65	19	46	15.81	49.19	29.23	24.32	0.65	0.21
4	36	10	26	12.53	23.47	27.78	34.80	0.51	0.27
5	29	16	13	12.47	16.53	55.17	43.00	1.00	0.75
6	23	11	12	12.44	10.56	47.83	54.09	0.17	0.20
7	24	13	11	15.54	8.46	54.17	64.77	0.42	0.77
8	28	23	5	20.68	7.32	82.14	73.84	0.26	0.74
9	16	8	8	13.70	2.30	50.00	85.63	2.37	14.13
10	33	27	6	31.14	1.86	81.82	94.36	0.55	9.21
								34.90	29.26
H								64.16	
gL								8.00	
X crítica								15.00	

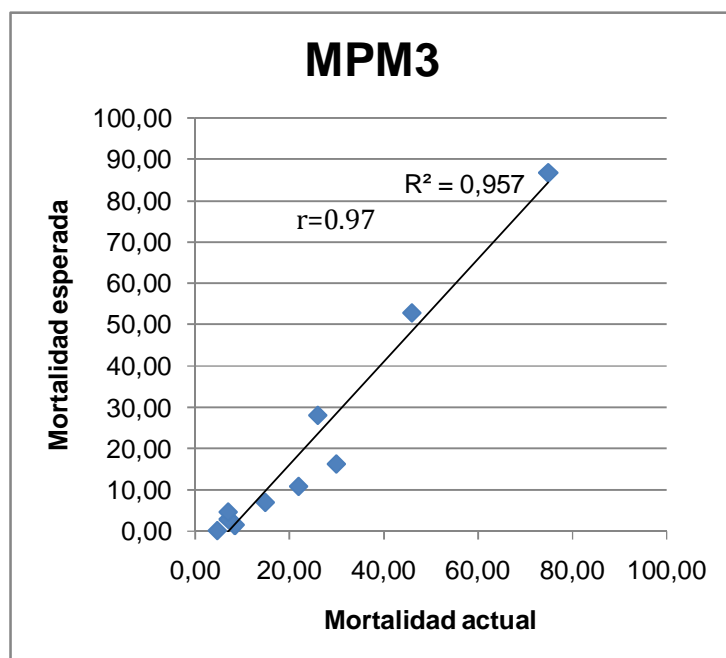
En las figuras 1 y 2 se presenta la correlación entre el promedio de la probabilidad de morir y la mortalidad actual para el total de registros agrupados en deciles de gravedad

progresivamente mayor. Para SAPS 3 el coeficiente de correlación fue de 0.97 y para MPM3 de 0.97.

Gráfica 1. Mortalidad esperada VS actual para SAPS 3



Gráfica 2. Mortalidad esperada VS actual para MPM3.



Al evaluar la capacidad discriminativa entre vivos y muertos, se construyeron curvas ROC (Receiver Operating Characteristic) mediante el programa estadístico SPSS. Calculándose así la sensibilidad y especificidad para el grupo de SAPS 3 y MPM3 así mismo se calculó el área por debajo de la curva . El área por debajo de la curva para SAPS 3 fue de 0.80 con un intervalo de confianza de 95 % (0.77-0.84). El área por debajo de la curva para MPM3 fue de 0.80 con un intervalo de confianza del 95% (0.77- 0.84) ver gráfico 3. Se presentan las curvas ROC para los dos modelos, observándose que la curva de SAPS 3 como de MPM3, su área por debajo de la curva es similar .

Gráfico 3. Curvas ROC para SAPS3 y MPM3

Discusión

Los modelos predictivos de mortalidad hospitalaria son usados de manera rutinaria en muchas unidades del mundo con la finalidad de hacer comparaciones de resultados entre países,^{25, 26} para asistir en la decisión de egreso de la UTI,²⁷ para evaluación de rendimiento de las unidades para la aleatorización de numerosos trabajos de investigación.

En este trabajo se evalúa el rendimiento de dos modelos matemáticos para predecir la mortalidad (MPM3 y SAPS 3) a través del análisis de datos capturados en forma prospectiva y recuperados de una base de datos. El rendimiento fue analizado en base a dos características: calibración y capacidad discriminativa.

Siguiendo las recomendaciones de Lemeshow y Hosmer se observó superioridad de SAPS 3 respecto a MPM3. En relación a la capacidad discriminativa se observó similitud para ambos modelos predictores de mortalidad.

En los artículos originales, los autores de SAPS 3 refieren ventajas de éste sobre los modelos previos, debido a que consideran en el desarrollo del diseño los avances en la ciencia médica, nuevas alternativas de tratamiento, así como el desarrollo de la computación, o dicho de otra manera, tratan de evitar la pérdida de la calibración propia del paso de los años. Otro punto a considerar es el hecho de que los pacientes y las prácticas médicas son diferentes debido a su distribución geográfica; esta podría ser una de las ventajas de SAPS 3, ya que este modelo se adapta a diferentes zonas geográficas.

Aguirre y cols evaluó el rendimiento de 2 modelos matemáticos para predecir la mortalidad (SAPS 3 y APACHE II), a través de análisis de datos obtenidos de forma

prospectiva, en una UTI, se observó superioridad de SAPS 3 respecto a APACHE II. En relación a la capacidad discriminativa se observó inferioridad de APACHE II.

Castella y cols.²⁸ reportaron la superioridad en rendimiento de los modelos de diseño más recientes, lo cual también fue demostrado por Moreno y cols.²⁹

Cerón y cols en un estudio evaluaron 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. Se observó superioridad de APACHE II y SAPS II respecto a MPM II-0 y MPM II-24.

Aunque en el trabajo original de SAPS 3 se desarrollaron modelos matemáticos de predicción regionales en este trabajo se utilizó el modelo matemático general lo cual permitiría hacer análisis de rendimiento en comparación con el comportamiento de todas las Unidades incluidas en el estudio original. Para este tipo de análisis era indispensable verificar el adecuado rendimiento de este modelo matemático tanto en su capacidad discriminativa como en su calibración. Los resultados de nuestro trabajo nos permiten afirmar que la aplicación de ambos modelos en estos pacientes tiene una buena capacidad discriminativa, aunque su calibración no es del todo satisfactoria fundamentalmente para MPM3, lo cual obligaría a hacer consideraciones especiales en ese tipo de análisis. Otra alternativa es hacer intentos para mejorar la calibración de ambos modos.

El origen de este problema de descalibración no es claro pero podría deberse a diferencias demográficas, diferencias en los sistemas de salud, tamaño de muestra, o

problemas el cálculo; sin embargo es importante señalar que la base de datos de donde proviene esta información es manejada por los médicos residentes de Medicina Crítica de un Hospital de Enseñanza donde se les capacita periódicamente en la obtención y captura de los datos, realizándose auditorías periódicas de la calidad, con el propósito de reducir al mínimo los errores de cálculo y captura de los mismos. Aunque el número de pacientes es considerable, no se descarta que también obedezca a un tamaño de muestra insuficiente.

Conclusiones

Este estudio demuestra que aunque MPM3 y SAPS3 tienen una aceptable capacidad discriminativa, su calibración no es del todo satisfactoria fundamentalmente para MPM3. Habrá que tomar en consideración estos resultados cuando se analicen sus aplicaciones.

Bibliografía

1. Vincent J-L, Moreno RP. Clinical review: scoring systems in the critically ill. *Crit Care*. 2010;14(2):311.
2. Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. *Crit Care Med*. 2011; 39(1):163-169.
3. 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(5):1297-1310.
4. Zimmerman JE, Kramer AA, McNair DS, Malila FM, Shaffer VL. Intensive care unit length of stay: benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV. *Crit Care Med*. 2006;34(10):2517-2529.

5. Metnitz PGH, Moreno RP, Almeida E, et al; SAPS 3 Investigators. SAPS 3—from evaluation of the patient to evaluation of the intensive care unit. Part 1: objectives, methods and cohort description. *Intensive Care Med.* 2005;31(10):1336-1344.
6. Moreno RP, Metnitz PG, Almeida E, et al; SAPS 3 Investigators. SAPS 3—from evaluation of the patient to evaluation of the intensive care unit. Part 2: development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med.* 2005;31(10):1345-1355.
7. Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). *Crit Care Med.* 2007;35(3):827-835.
8. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE—acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med.* 1981;9(8):591-597.
9. Le Gall JR, Loirat P, Alperovitch A, et al. A simplified acute physiology score for ICU patients. *Crit Care Med.* 1984; 12(11):975-977.
10. Lemeshow S, Teres D, Pastides H, Avrunin JS, Steingrub JS. A method for predicting survival and mortality of ICU patients using objectively derived weights. *Crit Care Med.* 1985; 13(7):519-525.
11. Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K. A new risk prediction model for critical care: the Intensive Care National Audit & Research Centre (ICNARC) model. *Crit Care Med.* 2007;35(4):1091-1098.
12. de Lange DW, Dusseljee J, Brinkman S, et al. Severity of illness and outcome in ICU patients in the Netherlands: results from the NICE registry 2006 – 2007. *Neth J Crit Care.* 2009;13(1):16-22.
13. Vasilevskis EE, Kuzniewicz MW, Cason BA, et al. Mortality probability model III and simplified acute physiology score II: assessing their value in predicting length of stay and comparison to APACHE IV. *Chest.* 2009;136(1):89-101.
14. Kuzniewicz MW, Vasilevskis EE, Lane R, et al. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. *Chest.* 2008;133(6):1319-1327.
15. Rowan KM, Kerr JH, Major E, McPherson K, Short A, Vessey MP. Intensive Care Society's APACHE II study in Britain and Ireland—II: outcome comparisons of intensive care units after adjustment for case mix by the American APACHE II method. *BMJ.* 1993;307(6910):977-981.
16. Render ML, Deddens J, Freyberg R, et al. Veterans Affairs intensive care unit risk adjustment model: validation, updating, recalibration. *Crit Care Med.* 2008;36(4):1031-1042.
17. Metnitz PG, Vesely H, Valentin A, et al. Evaluation of an interdisciplinary data set

- for national intensive care unit assessment. *Crit Care Med*. 1999;27(8):1486-1491.
18. Young MP, Birkmeyer JD. Potential reduction in mortality rates using an intensivists model to manage intensive care units. *Eff Clin Pract*. 2000;3(6):284-289.
 19. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29-36.
 20. Hosmer DW, Lemeshow S. Assessing the fit of the model. In: Hosmer DW, Lemeshow S, eds. *Applied Logistic Regression*. 2nd ed. New York, NY: Q Wiley-Interscience Publication; 2000:143-202.
 21. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. *Crit Care Med*. 2007;35(9):2052-2056.
 22. Wunsch H, Rowan KM, Angus DC. International comparisons in critical care: a necessity and challenge. *Curr Opin Crit Care*. 2007;13(6):725-731.
 23. Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott-Williams-Wilkins; 2008: 148-167.
 24. Michael J. Breslow, MD; and Omar Badawi, PharmD. *Severity Scoring in the Critically Ill Part 1—Interpretation and Accuracy of Outcome Prediction Scoring Systems* *Chest* 2012;141: 245-252.
 25. 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.
 26. 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.
 27. 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.
 28. 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.
 29. 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).