Artículo original

Revista de la Asociación Mexicana de
Medicina Crítica y TERAPIA INTENSIVA
Vol. XXI, Núm. 3 / Jul.-Sep. 2007
pp 119-124

Comparación del rendimiento de 2 modelos predictivos de mortalidad: SAPS 3 *vs* APACHE II, en una unidad de terapia intensiva mexicana

Dr. Carlos Alberto Aguirre Serrato,* Dr. Ulises W Cerón Díaz,† Dr. Alfredo Sierra Unzueta‡

RESUMEN

Introducción: 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.

Objetivo: Comparar el rendimiento de SAPS 3 y APACHE II para predecir la muerte hospitalaria.

Pacientes y métodos: Criterios de inclusión: enfermos que ingresaron a la UTI de enero a mayo de 2006. Criterios de exclusión: reingresos, enfermos coronarios, operados de corazón, alta voluntaria, altas a otro hospital. Se calculó la probabilidad de morir, por ambos sistemas, con los datos de ingreso a la UTI. El rendimiento de los modelos se evaluó en 2 áreas: a) capacidad discriminativa y b) calibración. La capacidad discriminativa se expresa por el área debajo de la curva (ADC) ROC (característica operativa del receptor) y la calibración se analiza a través de la prueba de bondad de ajuste de Lemeshow y Hosmer, así como la correlación entre mortalidad predicha y actual de cada uno de los grupos divididos por deciles. Se consideró estadísticamente significativo una $p < 0.05$ para las ADC y una adecuada calibración cuando el estadístico «C» era menor de la χ^2 crítica para 8 gL.

Resultados: Ciento diez enfermos registrados, 15 pacientes excluidos, 95 registros incluidos.

SUMMARY

Introduction: *The adoption of a predictive model requires a local validation work in order to ensure that its application is adjusted to the attended population.*

Objective: *To compare the performance of SAPS-3 and APACHE-II to predict in-hospital death.*

Patients and methods: *Inclusion criteria: patients who were admitted into the ICU from January until May 2006. Exclusion criteria: re-admittances, coronary patients, cardiac surgery patients, voluntary discharge, discharge to another hospital. Probability of death was calculated with both systems, with the UCI admittance data. The performance of the models was evaluated in 2 areas: a) discriminative capability and b) calibration. The discriminative capability is expressed by the AUC (area under the curve) ROC (receptor operative characteristic), and the calibration is analyzed through a Goodness of fit (Lemeshow and Hosmer), as well as the correlation between predicted and actual mortality for each group divided by deciles. A statistically significant probability of error (p) was considered of < 0.05 for the AUC, and an adequate calibration was considered when the C statistic tool was less than the critical χ^2 for 8 gL.*

Results: *110 registered patients, 15 of them were excluded, 95 included registries.*

* Médico residente de IV año.
† Médico adscrito.
‡ Jefe.

Unidad de Terapia Intensiva «Dr. Alberto Villazón Sahagún» del Hospital Español de México.

	SAPS 3	APACHE II	«p»
ADC * «C».	0.86 ± 0.018	0.79 ± 0.015	< 0.01
Lemeshow y Hosmer	6.54, p > 0.1	16.95, p < 0.05	

* ADC: Área por debajo la curva, ± EE Error estándar.

Conclusión: SAPS 3 tiene mejor rendimiento que APACHE II.

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

	SAPS 3	*APACHE II*	*«p»*
*AUC * «C».*	*0.86 ± 0.018*	*0.79 ± 0.015*	*< 0.01*
Lemeshow and Hosmer	*6.54, p > 0.1*	*16.95, p < 0.05*	

** AUC (area under the curve) ± EE.*

Conclusion: *SAPS 3 showed a better performance than APACHE II.*

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

INTRODUCCIÓN

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 UTI´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.

Desde la década de los 80 se han publicado una serie de escalas de calificación de gravedad y modelos predictivos de mortalidad (APACHE en 1981,[1] SAPS en 1984,[2] APACHE II en 1985,[3] MPM en 1981,[4] APACHE III en 1991,[5] SAPS II[6] y MPM II[7] en 1993), así como trabajos de validación, como el publicado por Sánchez y cols.,[8] el cual comparó la capacidad discriminativa de TISS, APACHE II, APACHE III, SAPS I, SAPS II, MPM II-0 y MPM II-24 con resultados a favor de APACHE II.

Recientemente, Moreno y cols. publicaron los resultados del diseño y validación de SAPS 3 a través de un estudio multicéntrico, multinacional (donde nuestra UTI colaboró) que incluyó un total de 16,784 pacientes admitidos de manera consecutiva en 303 UTI´s, con la característica innovadora de ser adaptado a cada región.[9,10] El último modelo en surgir fue APACHE IV,[11] en este año 2006.

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; estos trabajos ya han sido publicados con anterioridad en nuestro país, y como experiencia local contamos con el estudio publicado por Cerón y cols. que comparó de manera multicéntrica APACHE II, SAPS II, MPM II-0 y MPM II-24,[12] los cuales demostraron rendimiento aceptable.

El propósito de este trabajo es comparar el rendimiento de SAPS 3 y APACHE II para predecir mortalidad hospitalaria en la UTI «Dr. Alberto Villazón Sahagún» del Hospital Español de México.

PACIENTES Y MÉTODOS

El estudio incluyó a todos los enfermos que ingresaron a la UTI de nuestro hospital de manera consecutiva, de enero a mayo de 2006. La UTI cuenta con 12 camas y pertenece a un hospital de especialidades privado con residentes de especialización que incluye la de medicina del enfermo en estado crítico.

Todos los pacientes fueron calificados en su gravedad de acuerdo a las instrucciones publicadas originalmente para APACHE II[3] y SAPS 3.[9,10] Para este último caso se utilizó la herramienta que los autores han dejado disponible para los usuarios en Internet (*www.saps3.com*).

Para el cálculo de la probabilidad de morir se utilizaron los modelos matemáticos publicados por los autores; en el caso de SAPS 3 se eligió la ecuación general.

Para el análisis se excluyeron a menores de 18 años, operados de corazón, enfermos coronarios, traslados a otro hospital y reingresos.

El recuento y análisis de los datos se hicieron a través de programas creados en Excel.

El rendimiento de ambos modelos se analizó a través de evaluar tanto la capacidad discriminativa como la calibración. Para evaluar la capacidad discriminativa entre vivos y muertos, construimos curvas ROC (Receiver Operating Characteristic) para cada modelo, a través de calcular la sensibilidad y

especificidad en 10 puntos de corte de la probabilidad de morir.

Para establecer si hay diferencia estadísticamente significativa se calculó el área por debajo de la curva ROC[13] de cada modelo y se puso a prueba la hipótesis de nulidad de acuerdo a lo recomendado por Hanley y cols.[14] Se consideró significante una $p < 0.05$.

Para evaluar la calibración, se calculó el estadístico «C» de la prueba de bondad de ajuste de Lemeshow y Hosmer,[15] la cual nos permite evaluar la discrepancia entre el número de muertos observados y muertos esperados, así como sobrevivientes observados y sobrevivientes esperados, en 10 grupos de enfermos (deciles) de pronóstico vital progresivamente peor. Mientras menor es la discrepancia, menor es el valor del estadístico «C». Si el valor de «C» no supera el valor crítico de χ^2 para 8 gL y una $P < 0.05$ bimarginal, se considera que no se puede rechazar la hipótesis nula de no diferencia entre los eventos observados y los esperados; por lo tanto, la calibración es aceptable.

Se hizo un análisis de correlación entre la probabilidad hospitalaria de morir calculada con cada modelo y la mortalidad hospitalaria actual en los 10 grupos de pacientes (deciles) con pronóstico progresivamente mayor de morir. Los resultados se expresan a través de coeficiente de Pearson (r).

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

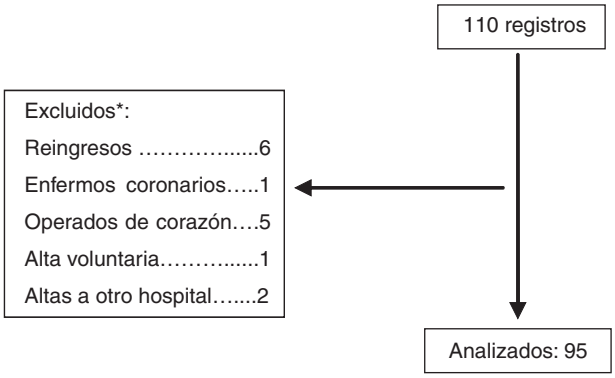
Los datos demográficos fueron extraídos de la base de datos de la UTI (BASUTI), que es llevada de manera prospectiva por los médicos de la unidad.

RESULTADOS

De un total de 110 ingresos, 15 fueron excluidos y 95 analizados, las causas de la exclusión se presentan en la *figura 1*, siendo la causa más frecuente el reingreso. Los datos demográficos de los pacientes seleccionados se presentan en el *cuadro I*.

La gran mayoría (67%) de los pacientes ingresaron por presentar falla de uno o más sistemas orgánicos vitales; con respecto a la categoría diagnóstica los pacientes no quirúrgicos fueron la gran mayoría (72%) con predominancia de la falla respiratoria con 25%.

La gravedad de los enfermos se expresa en el *cuadro II*. El promedio de APACHE II fue de 17 ± 9 y de SAPS 3 de 50 ± 17 puntos. La mortalidad hospitalaria fue de 20%, la probabilidad hospitalaria de morir por SAPS 3 y APACHE II fue de 24.7 y 28.6% respectivamente.



**Número de enfermos excluidos por cada uno de los criterios.*

Figura 1. *Total de enfermos y criterios de exclusión.*

Cuadro I. Datos demográficos de la totalidad de los enfermos.

	n = 95
Sexo (M/F):	49 (51)/46 (49)
Edad:	59 ±
FDI (A/B/C): (63/8/24)	63 (67)/8 (8)/24 (25)
Condición antes del ingreso:	
Desconocido:	0
Encamados:	10 (10)
Sintomático ambulatorio:	39 (41)
Asintomático:	46 (48)
Condición de ingreso:	
Estables:	32 (33)
Crítico inestables:	63 (66)
Moribundo:	1
Cirugía de urgencia:	13 (13)
Categoría diagnostica*	
Postoperados	27 (28)
No quirúrgico falla respiratoria	26 (27)
No quirúrgico falla cardiovascular:	13 (13)
No quirúrgico falla neurológica:	5 (5)
No quirúrgico otros:	24 (25)

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; Otros = cetoacidosis, sobredosis, sangrado de tubo digestivo.
Los resultados se expresan en frecuencia, entre paréntesis se presentan los porcentajes. La edad se expresa en promedio ± 1 desviación estándar.
* De acuerdo a clasificación usada en APACHE II.

La condición previa a la enfermedad que dio origen al ingreso fue normal en 48% de los casos; eran sintomáticos pero ambulatorios 41% de los casos. La mayoría (66%) ingresaron en una condición crítica e inestable.

El área por debajo de la curva ROC para APACHE II fue de 0.79 ± 0.015 (EE), para SAPS 3 de 0.86 ± 0.018 (EE); La diferencia de las áreas fue estadísticamente significativa ($p < 0.01$).

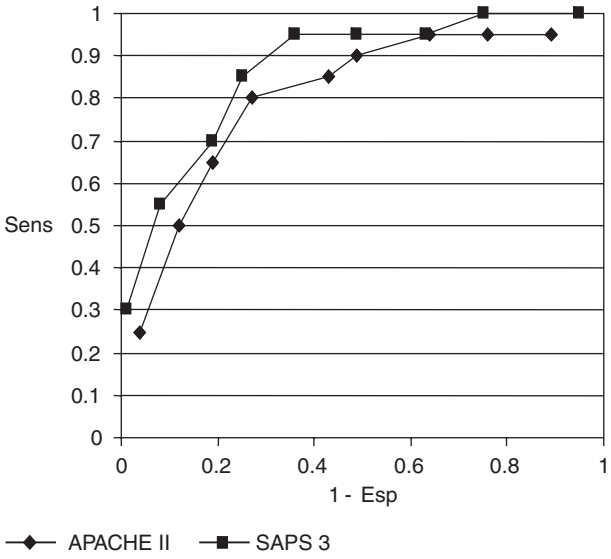
En la *figura 2* se presentan las curvas ROC para los dos modelos. Se puede observar que la curva de SAPS 3, tiene un área visualmente mayor, que APACHE II; esto se refleja en los cálculos del área y el significado estadístico de la comparación entre los dos modelos.

Los resultados para el estadístico «C» de la prueba de bondad de ajuste de Lemeshow y Hosmer se presentan en los *cuadros III y IV*. El estadístico «C» para SAPS 3 no superó el valor crítico de χ^2 ($p > 0.1$), a diferencia de APACHE II que sí lo superó ($p < 0.05$). La correlación entre la mortalidad esperada y la mortalidad actual demostró una «r» de 0.92 para ambos modelos *(figura 3)*. Es de notarse que existe una tendencia a mantener una mejor calibración por parte de SAPS 3 en probabilidades de morir mayores a 50%, mientras que APACHE II tiende a infraestimar a la mortalidad actual y existe una mayor dispersión en la probabilidad de morir de menor de 50%.

Cuadro II. Cuadro de gravedad de los enfermos.

n =	95 enfermos
APACHE II:	17.3 ± 9.5
SAPS 3:	50.0 ± 17.4
Mortalidad hospitalaria (%):	20
Probabilidad de morir por APACHE II (%):	28.6 ± 28.4
Probabilidad de morir por SAPS 3 (%):	24.7 ± 23.7

Se expresa en media ± desviación estándar.



ADC APACHE II ($\bar{X}$ ± EE) = 0.79 ± 0.015,
ADC SAPS 3 ($\bar{X}$ ± EE) = 0.86 ± 0.018
p < 0.01

Figura 2. *Curva ROC para los dos modelos.*

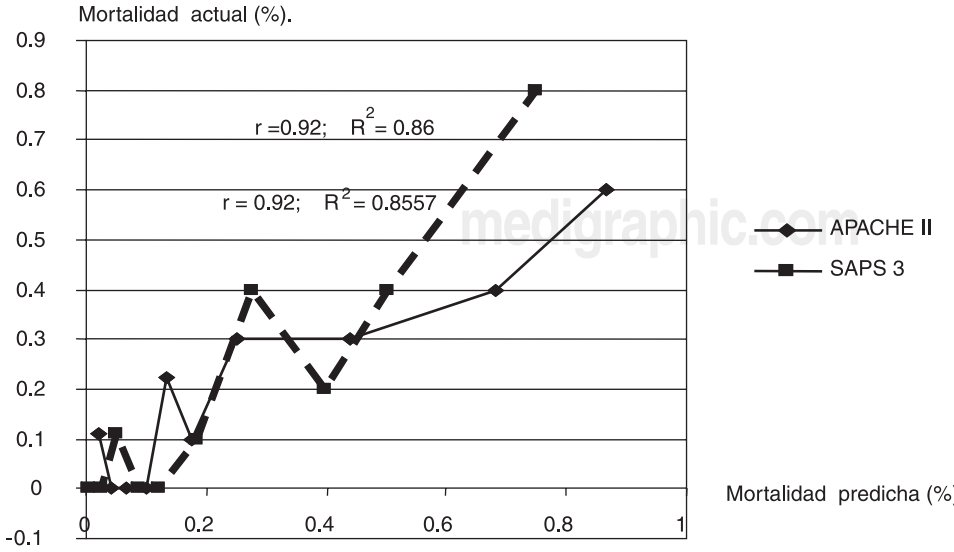


Figura 3. *Gráfico de correlación entre mortalidad predicha y mortalidad actual para SAPS3 y APACHE II.*

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,[16,17] para asistir en la decisión de egreso de la UTI,[18] para evaluación de rendimiento de las unidades,[19] para la aleatorización de numerosos trabajos de investigación. 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.

Basados en los comentarios hechos anteriormente, creemos en la necesidad de la búsqueda de nuevas herramientas que contribuyan a nuestro mejor ejercicio médico, pero para poder hacerlo es necesaria la validación de esas nuevas herramientas en nuestra población de trabajo.

En este trabajo se evaluó el rendimiento de 2 modelos matemáticos para predecir la mortalidad (SAPS 3 y APACHE II), a través de análisis de da-

Cuadro III. Prueba de bondad de ajuste de Lemeshow y Hosmer para SAPS 3.

Rango	Total	Probabilidad de morir (%)	Muertos actuales	Muertos esperados	Vivos actuales	Vivos esperados
0-0.1	9	0.006	0	0.06	9	8.94
0.1-0.3	9	0.024	0	0.22	9	8.78
0.4-0.6	9	0.05	1	0.45	8	8.55
0.6-0.10	9	0.087	0	0.79	9	8.21
0.10-0.15	9	0.123	0	1.11	9	7.89
0.15-0.22	10	0.185	1	1.85	9	8.15
0.22-0.34	10	0.276	4	2.76	6	7.24
0.34-0.46	10	0.398	2	3.98	8	6.02
0-46-0.58	10	0.5	4	5.0	6	5.0
0.58-0.89	10	0.749	8	7.49	2	2.51

C = 6.54, gL = 8, χ^2 crítica 15.5, $p > 0.1$

Cuadro IV. Prueba de bondad de ajuste usando la técnica de Lemeshow y Hosmer para APACHE II.

Rango	Total	Probabilidad de morir (%)	Muertos actuales	Muertos esperados	Vivos actuales	Vivos esperados
0.003-0.034	9	0.022	1	0.20	8	8.79
0.036-0.048	9	0.042	0	0.38	9	8.61
0.049-0.085	9	0.066	0	0.60	9	8.40
0.087-0.121	9	0.100	0	0.90	9	8.10
0.121-0.155	9	0.134	2	1.21	7	7.78
0.163-0.193	10	0.175	1	1.75	9	8.24
0.197-0.315	10	0.250	3	2.50	7	7.49
0.325-0.559	10	0.439	3	4.40	7	5.60
0.559-0.784	10	0.684	4	6.84	6	3.15
0.798-0.955	10	0.866	6	8.66	4	1.33

C = 16.95, gL = 8, χ^2 crítica 15.5, $p < 0.05$

gl: grados de libertad, C: estadístico «C» de la prueba de bondad de ajuste de Lemeshow y Hosmer; χ^2: chi cuadrada.

tos obtenidos de forma prospectiva, en una UTI. 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 APACHE II. En relación a la capacidad discriminativa se observó inferioridad de APACHE II. Estudios previos tales como el de Castella y cols.[20] reportaron la superioridad en rendimiento de los modelos de diseño más recientes, lo cual también fue demostrado por Moreno y cols.[8]

La razón por la cual SAPS 3 muestra mejor rendimiento que APACHE II podría radicar en las ventajas que da el ser un diseño más reciente; además son puntos a considerar que SAPS 3 tiene menos variables a evaluar (*www.saps3.org*), lo cual lo convierte en un modelo más práctico de utilizar, así como que puede ser utilizado con los datos recogidos en la primera hora de internamiento en la UTI, y considera los datos previos al ingreso del paciente (como uso de aminas, por ejemplo).

En conclusión, este trabajo demuestra que el modelo de SAPS 3, supera a APACHE II en su capacidad para predecir la mortalidad hospitalaria en los enfermos internados en nuestra Unidad de Cuidados Intensivos.

BIBLIOGRAFÍA

1. Knaus WA, Zimmerman JE, Wagner DP et al. APACHE-Acute physiology and chronic healt 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: Aseverity, 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. Moreno RP, Metnitz P, Almeida E et al. 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;31:1336-1344 DOI 10.1007/s00134-005-2762-2766.
10. Moreno RP, Metnitz P, Almeida E et al. 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;31:1345-1355 DOI 10.1007/s00134-005-2763-2765.
11. Zimmerman MD, Kramer A et al. Acute physiology and chronic health evaluation (APACHE) IV, *http://meeting.chestjournal.org/cgi/content/Abstract/128/4/297S*
12. Cerón DU, Esponda PJ, Dr. Borboya PM et al. 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. *Rev Asoc Mex Med Crit y Ter Int* 2000;14(2):50-59.
13. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
14. 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.
15. Lemeshow S, Hosmer DW. A review of goddness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;110:847-864.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.

Correspondencia:
Dr. Carlos Alberto Aguirre Serrato
Unidad de Terapia Intensiva «Dr. Alberto Villazón Sahagún» del Hospital Español de México. Tel: 01-55-55-79-29-52,
Tel. Celular: 01-55-18-63-63-18.
Correo electrónico:
sancar20002003@yahoo.com.mx