

Artículo original

Validación de dos modelos de predicción de mortalidad, PRISM y PIM2, en una Unidad de Cuidados Intensivos Pediátricos

Dres. Pablo Eulmesekian*, Augusto Pérez*, Pablo Mincos*, Hilario Ferrero* y Tomás Fiori Bimbi*

RESUMEN

Introducción. Pediatric Index of Mortality 2 (PIM2) y Pediatric Risk of Mortality (PRISM) son puntajes para predecir la probabilidad de mortalidad en unidades de cuidados intensivos pediátricos. El objetivo fue validar ambos modelos en la Unidad del Hospital Italiano de Buenos Aires.

Población, material y métodos. En forma prospectiva se recogió la información para obtener los puntajes de predicción de mortalidad para cada uno de los pacientes ingresados consecutivamente a la unidad entre 1/1 y 31/12 de 2004. Se emplearon en el cálculo las ecuaciones originales de cada modelo. La calibración se evaluó con la prueba de Hosmer y Lemeshow (HL) y los Índices de Mortalidad Estandarizados [Standardized Mortality Ratio, (SMR)] para la población completa; la discriminación con el área bajo la curva ROC (Receiver Operating Characteristics).

Resultados. Se incluyeron 765 pacientes. 53,6% fueron varones. La mediana de edad fue 4,1 años (1 mes-20 años). La mediana de la estadía en UCIP fue 2 días (1-154 días). Murieron en UCIP 22 pacientes (2,9%). La mortalidad estimada por PIM2 fue de 3,3%; SMR= 0,87 (IC 95% 0,54-1,21); HL $\chi^2 = 15,16$ (p= 0,01) y el área bajo la curva ROC fue de 0,92 (IC 95% 0,90-0,94). La mortalidad estimada por PRISM fue de 5,5%; SMR= 0,50 (IC 95% 0,28-0,73); HL $\chi^2 = 22$ (p= 0,0005) y el área bajo la curva ROC fue de 0,94 (IC 95% 0,92-0,96).

Conclusiones. PIM2 tuvo una mejor adecuación que PRISM en la población heterogénea de nuestra UCIP. Aunque la discriminación de ambos modelos fue muy buena, PRISM sobrestimó claramente la mortalidad.

Palabras clave: PIM2, PRISM, puntajes, mortalidad, cuidados intensivos pediátricos.

SUMMARY

Introduction. Pediatric Index of Mortality 2 (PIM2) and Pediatric Risk of Mortality (PRISM) are models of mortality probability prediction for Pediatric Intensive Care Units (PICU). We aimed to validate both scores in the multidisciplinary PICU of Hospital Italiano de Buenos Aires.

Population, material and methods. We prospectively collected all the information to obtain both scores of mortality prediction for each consecutive patient admitted to the PICU from January 1st to December 31st 2004. We used original equations for both models. Calibration was assessed with

Hosmer-Lemeshow (HL) goodness-of-fit tests and Standardized Mortality Ratio (SMR) for the whole population. Discrimination was assessed with the area under Receiver Operating Characteristic (ROC) curve.

Results. 765 patients were included over the 12 months period. 53.6% were male. Median age was 4.1 years (1 month-20 years). Median length of stay (LOS) was 2 days (1-154 days). 22 patients died in PICU (2.9%). PIM2 estimated mortality was 3.3%; SMR= 0.87 (95% CI 0.54-1.21); HL $\chi^2 = 15.16$ (p= 0.01) and ROC curve of 0.92 (95% CI 0.90-0.94). PRISM estimated mortality was 5.5%; SMR= 0.50 (95% CI 0.28 -0.73); HL $\chi^2 = 22$ (p= 0.0005) and ROC curve of 0.94 (95% CI 0.92-0.96).

Conclusions. PIM2 had a better performance than PRISM in the heterogeneous patient population from our PICU. Although both model discriminated very well, PRISM overestimated mortality remarkably.

Key words: PIM2, PRISM, scores, mortality, pediatric intensive care.

INTRODUCCIÓN

En las unidades de cuidados intensivos pediátricos (UCIP) es importante medir los resultados en forma objetiva para evaluar la calidad del servicio prestado a los pacientes y comparar el funcionamiento de la unidad en el tiempo y con otras unidades.¹

Los puntajes de predicción de mortalidad constituyen una herramienta útil para llevar a cabo estos objetivos. Permiten comparar la mortalidad observada en una población con la estimada por el modelo de predicción.

En cuidados intensivos pediátricos, los puntajes de predicción de mortalidad más utilizados son PRISM (Pediatric Risk of Mortality) y PIM (Pediatric Index of Mortality). En líneas generales, difieren en la cantidad de información necesaria

* Unidad de Cuidados Intensivos Pediátricos. Hospital Italiano de Buenos Aires.

Correspondencia:
Dr. Pablo Eulmesekian.
pablo.eulmesekian@hospitalitaliano.org.ar

Aclaración de intereses:
Ninguno para declarar.

para predecir el riesgo de muerte, el momento y el intervalo de tiempo en que ésta debe obtenerse.

PRISM² registra el peor valor de 14 variables fisiológicas evaluadas durante las primeras 24 horas desde el ingreso del paciente a UCIP para calcular un puntaje que se asocia a una probabilidad de muerte. Se desarrolló con datos obtenidos entre los años 1980 y 1985 en aproximadamente 2.700 pacientes en 9 UCIP de los EE.UU. y se ha utilizado ampliamente. Se realizó una revisión y actualización de PRISM, obteniéndose PRISM III³ con datos provenientes de 11.165 pacientes de 32 UCIP de EE.UU. durante los años 1993 y 1994. PRISM III utiliza 17 variables fisiológicas y también tiene en cuenta para determinar la probabilidad de mortalidad cinco categorías diagnósticas específicas y el tratamiento recibido antes del ingreso a UCIP. La ecuación utilizada para predecir la mortalidad está patentada.

PIM⁴ se desarrolló con datos de aproximadamente 6.000 pacientes de 7 UCIP de Australia y 1 de Inglaterra en los años 1994 a 1996. También se actualizó a PIM⁵ con datos de aproximadamente 21.000 pacientes de 14 UCIP de Australia, Nueva Zelanda e Inglaterra en los años 1997 a 1999. Tanto PIM como PIM2 se calculan con datos obtenidos al ingreso del paciente a la UCIP, con menor número de variables que PRISM. Su utilización es más simple y la ecuación para su cálculo es de dominio público.

Como parte del programa de control de calidad de nuestra UCIP decidimos trabajar con dos modelos de predicción de mortalidad para poder objetivar resultados, compararlos con otras unidades y definir el puntaje que mejor se adecua a la unidad. Elegimos PRISM, por ser el más comúnmente utilizado y PIM2 por haber sido recientemente desarrollado y reflejar el nivel de cuidados intensivos pediátricos brindado actualmente. Por lo tanto, el objetivo de este estudio fue validar ambos puntajes de predicción de mortalidad, según PRISM y PIM2, en la población de pacientes internados en la UCIP multidisciplinaria del Hospital Italiano de Buenos Aires.

POBLACIÓN, MATERIAL Y MÉTODOS

El estudio se realizó en la UCIP multidisciplinaria del Hospital Italiano de Buenos Aires, hospital general universitario con un Departamento de Pediatría cuya UCIP tiene 18 camas y aproximadamente 850 ingresos anuales.

Se incluyeron todos los pacientes ingresados a la unidad entre el 1 de enero de 2004 y el 31 de diciembre de 2004. Los datos se recolectaron en

forma prospectiva en planillas especialmente diseñadas. El cálculo de la probabilidad de muerte estimada por cada puntaje se realizó con las ecuaciones originales.^{2,5}

Se excluyeron aquellos pacientes que fueron ingresados a la unidad en forma electiva por un período menor a 12 horas (por ejemplo, cambio de traqueostoma, punción de médula ósea, admisión transitoria por falta de camas en las salas de pediatría general). Además de la información necesaria para calcular los puntajes se registraron edad, sexo, diagnóstico de ingreso, estadía en UCIP, condición clínica o posquirúrgica y fallecimiento en UCIP. Los autores recolectaron los datos. El autor principal los revisó al momento del cálculo de la probabilidad de muerte de cada paciente y finalmente se ingresaron a una hoja de cálculo (Excel). Esta base de datos se revisó periódicamente.

Análisis estadístico: La discriminación entre supervivencia y muerte se evaluó a través del análisis del área bajo la curva ROC (Receiver Operating Characteristic) y su correspondiente IC 95%. Un área bajo la curva ROC > 0,8 se consideró como discriminación adecuada.⁷ La calibración se evaluó con la prueba de Hosmer y Lemeshow. Para ello se dividió la población en cinco intervalos según el riesgo de muerte: <1%, 1-5%, 5-15%, 15-30% y >30%.^{2,6} y se comparó el número de muertes observado y estimado por el modelo en cada grupo. Se utilizó la fórmula $\sum (O - E) / E$; donde \sum es sumatoria de los resultados de cada grupo, O es número de muertes observado y E número de muertes estimado.⁷ Los valores de p se calcularon con 5 grados de libertad. Valores de p >0,05 sugieren una calibración adecuada, considerándose mejor a mayores valores.⁶ También se calculó el SMR (Standardized Mortality Ratio) con sus correspondientes IC 95%.⁸ en la población completa. El SMR se obtiene al dividir el número de muertes ocurridas sobre el número de muertes estimadas por el modelo. Las variables no categóricas paramétricas o no paramétricas se presentan con su medida central (media o mediana) y su correspondiente medida de dispersión (desvío estándar o rango). El análisis estadístico de los datos se realizó con el programa Stata 6.0 (StataCorp, USA).

RESULTADOS

Durante el período de estudio, entre el 1 de enero y el 31 de diciembre del 2004, ingresaron a la unidad 810 pacientes. Debido a permanencia en UCIP menor a 12 horas por razones electivas se excluyeron 45 pacientes, ninguno de los cuales falleció.

Se incluyeron en el análisis 765 pacientes, 22 de los cuales fallecieron (2,9%).

Aunque 18 pacientes permanecieron en UCIP menos de 24 horas se incluyeron en el análisis de PRISM ya que en este grupo se produjeron 2 muertes.

Ninguno de los 765 pacientes incluidos en el estudio egresó de la UCIP a los sectores pediátricos o a los hospitales de referencia en condiciones de riesgo de vida. La mediana de edad de los pacientes fue de 4,1 años (1 mes-20 años); la mediana de permanencia en UCIP fue de 2 días (1-154); requirieron asistencia respiratoria mecánica (ARM) el 20% de los pacientes, aunque no se tuvieron en

cuenta para este cálculo aquellos con ventilación no invasiva o ARM como soporte respiratorio de anestesia general para procedimientos realizados en UCIP (por ejemplo, colocación de catéteres o drenajes). El 40% de los días en UCIP corresponde a días en ARM (días de ARM/días totales de internación). La *Tabla 1* resume las características de la población.

Las medidas de calibración y discriminación para ambos puntajes se observan en la *Tabla 2*. Las *Tablas 3* y *4* muestran, respectivamente, la calibración de PIM2 y PRISM para cinco intervalos de riesgo preestablecidos.

El área bajo la curva ROC fue de 0,92 (IC 95% 0,90 - 0,94) para PIM2 y 0,94 (IC 95% 0,92 - 0,96) para PRISM.

La prueba de Hosmer-Lemeshow arrojó un valor de 15,16 para PIM2 ($p=0,01$) y de 22 para PRISM ($p=0,0005$).

El SMR fue de 0,87 (IC 95% 0,54 - 1,21) para PIM2 y 0,50 (IC 95% 0,28 - 0,73) para PRISM.

TABLA 1. Características de la población

Pacientes (n)	765
Mortalidad	22 (2,9%)
Sexo (M/F)	410/355 (53,6/46,4%)
Edad (mediana)	4,1 años (1 mes - 20 años)
Días en UCIP (mediana)	2 (1 - 154)
Pacientes ventilados**	154 (20%)
Pacientes clínicos	547 (71%)
Pacientes posquirúrgicos	218 (29%)
Diagnóstico de alto riesgo*	130 (17%)

* Definidos por PIM2 (paro cardiorrespiratorio, inmunodeficiencia combinada severa, leucemia-linfoma no en inducción, miocarditis-miocardiopatía, hemorragia cerebral espontánea, corazón izquierdo hipoplásico, VIH, insuficiencia hepática, enfermedad neurodegenerativa).

**n (%).

TABLA 2. Calibración y discriminación

Puntaje	Mortalidad estimada	SMR (95% IC)	Hosmer Lemeshow	P (df= 5)	AUC (95% IC)
PRISM	5,5%	0,50 (0,28 - 0,73)	22	0,0005	0,94 (0,92 - 0,96)
PIM2	3,3%	0,87 (0,54 - 1,21)	15,16	0,01	0,92 (0,90 - 0,94)

TABLA 3. PIM2: Calibración en cinco intervalos de riesgo

Intervalo de riesgo	Número	Muertes		Sobrevivientes	
		Observadas	Estimadas	Observados	Estimados
<1	417	0	1,87	417	415,13
1 - <5	232	4	4,66	228	227,34
5 - <15	79	12	5,94	67	73,06
15 - <30	25	1	5,09	24	19,91
>30	12	5	7,61	7	4,39
Total	765	22	25,17	743	739,83

TABLA 4. PRISM: Calibración en 5 intervalos de riesgo

Intervalo de riesgo	Puntaje Promedio	Número	Muertes		Sobrevivientes	
			Observadas	Estimadas	Observados	Estimados
<1	1,81	294	0	1,58	294	292,42
1 - <5	5,49	350	3	7,07	347	342,93
5 - <15	13,2	58	4	5,52	54	52,48
15 - <30	18,27	29	3	6,57	26	22,43
>30	27,94	34	12	22,37	22	11,63
Total	-	765	22	43,11	743	721,89

DISCUSIÓN

Los puntajes de predicción de mortalidad constituyen una herramienta ampliamente utilizada en cuidados intensivos. Pueden utilizarse para evaluar la gravedad de los pacientes en estudios clínicos, controlar la calidad de la UCIP y comparar resultados entre diferentes unidades.⁹

Mediante la validación de PRISM y PIM2 en nuestra población observamos que ambos modelos discriminan bien entre muerte y supervivencia aunque no calibran adecuadamente si tenemos en cuenta los valores de p obtenidos en la prueba de Hosmer y Lemeshow. No obstante, al analizar los SMR consideramos que la calibración de PIM2 es adecuada y claramente superior a la de PRISM.

El SMR constituye una forma simple de comparar el desempeño de una unidad con aquellas unidades en las que se desarrolló el modelo. Se obtiene dividiendo el número de muertes ocurridas en una población sobre el número de muertes estimadas de ocurrir por el modelo. Un $SMR < 1$ sugiere que el desempeño es mejor que lo estimado mientras que, por el contrario, un $SMR > 1$ sugiere que el desempeño es peor que lo estimado por el modelo. Dicho de otro modo, el SMR permite comparar el número de muertes ocurridas en una población con una estimación del número de muertes ocurridas en ésta si hubiera sido tratada en las unidades donde se desarrolló el modelo.¹⁰

Los SMR de la población estudiada fueron menores a 1. Esto significa que para PRISM, con un SMR de 0,50 (IC 95% 0,28 - 0,73), se produjeron entre 72 y 27% menos muertes que las que hubieran ocurrido si hubiera sido tratada en EE.UU. a principios de los 80. Al aplicar el mismo razonamiento

con PIM2, un SMR de 0,87 (IC 95% 0,54 - 1,21) significa que en nuestra población ocurrieron entre 46% menos de muertes y 21% más de muertes que si hubiera sido tratada a fines de los 90 en aquellas UCIP donde se desarrolló PIM2.

Otro método de evaluar la calibración de modelos de predicción es la prueba de Hosmer y Lemeshow. Demasiado énfasis se deposita habitualmente en el valor absoluto de esta prueba y su correspondiente valor de p. Generalmente se considera inadecuada la calibración de un modelo cuando el valor de p es $< 0,05$. No obstante, este concepto debe interpretarse con cautela, ya que, el valor de p es poco confiable en muestras pequeñas o con escaso número de muertes¹⁰ y además, con esta prueba se pueden obtener muchos valores diferentes de p en la misma población cuando se altera la disposición de los sujetos.¹¹

En nuestra población, el valor de p obtenido con la prueba de Hosmer y Lemeshow fue menor a 0,05 para los dos modelos ¿Significa esto que ambos puntajes son inapropiados para utilizarlos en nuestra población y deben descartarse o adaptarse con un factor de corrección? Creemos que no, porque algunos, como Shann,¹⁰ sostienen que el valor de p es menos importante que el análisis de muertes observadas y estimadas en cada intervalo de riesgo (Figuras 1 y 2). Consideramos lógico que el valor de p sea menor de 0,05 para PRISM, ya que el IC 95% del SMR está por debajo de 1 y la mortalidad estimada en cada intervalo de riesgo es mayor que la observada. Una posible interpretación de esto es que los cuidados brindados en nuestra unidad son diferentes de los brindados en las unidades donde se desarrolló PRISM hace 20-25 años y no que el

FIGURA 1. PIM2: Muertes según intervalo de riesgo

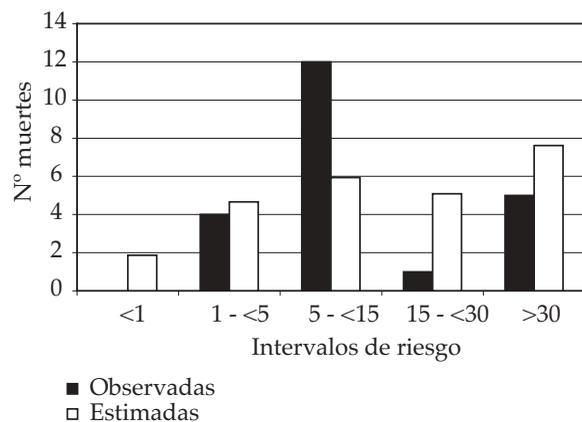
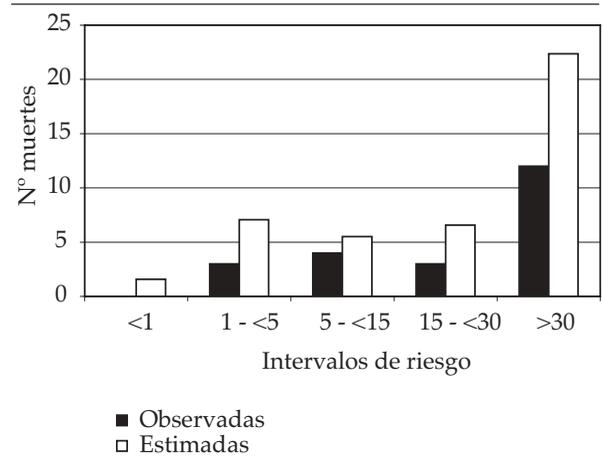


FIGURA 2. PRISM: Muertes según intervalo de riesgo



modelo es defectuoso y necesita ser adaptado a esta población. Algo similar ocurre con los resultados obtenidos de la calibración de PIM2. El valor de p de la prueba de Hosmer-Lemeshow fue $< 0,05$ aunque la mortalidad observada en los diferentes intervalos de riesgo es menor a la estimada (excepto en el intervalo de riesgo medio que va de 5 a 15%) y el SMR es cercano a 1, lo que sugiere que la calibración podría no ser inadecuada y el modelo, útil en esta población. Por ello y luego de este estudio, en la Unidad no seguimos utilizando PRISM y continuamos usando PIM2 como herramienta de control de calidad.

Las limitaciones más importantes de este estudio son el escaso número de pacientes, el escaso número de muertes y el hecho de provenir de sólo una UCIP. Con respecto al número de pacientes creemos que los valores de las pruebas de calibración de PIM2 serían diferentes si la muestra fuera mayor. Además por ser ésta una población procedente de una sola UCIP, los resultados obtenidos tal vez no sean generalizables a otras unidades.

Una posible limitación del estudio es el hecho de que más del 80% de los pacientes tienen una mortalidad estimada menor al 5%. No obstante, PRISM y PIM2 son modelos de predicción de mortalidad y no de gravedad de enfermedad;¹⁰ esto significa que no siempre pacientes con bajo riesgo de mortalidad estimada por el modelo tienen poca gravedad; por ejemplo, los pacientes con bronquiolitis en asistencia respiratoria mecánica tienen una mortalidad estimada por PIM2 menor al 1% y esto no significa que no estén gravemente enfermos. Una distribución de riesgo de mortalidad similar a la de nuestra población ha sido descripta en otros estudios.^{4,5,7}

Desde el año 1997 no se publican en Argentina estudios que evalúen el desempeño de las unidades de cuidados intensivos pediátricos. Ese año, y concretando uno de los pocos trabajos multicéntricos de nuestra especialidad, Saporiti y col.,¹² lograron describir las características y el desempeño de 13 UCIP del país, logrando reflejar la realidad nacional. Varios años han pasado y creemos que el funcionamiento de nuestras UCIP debe haber mejorado. Consideramos, por lo tanto, de suma utilidad la realización de un estudio similar que incluya el mayor número posible de unidades para poder evaluar el modelo de predicción de mortalidad más apropiado a nuestra realidad y de este modo ser capaces de fijar estándares nacionales de cuidados y resultados, a los que todos debemos tender para luego mejorar progresivamente año a año.

CONCLUSIONES

Tanto PRISM como PIM2 muestran una adecuada discriminación entre muerte y supervivencia en esta población. No obstante, PIM2 muestra una calibración superior a PRISM, motivo por el cual es el modelo de predicción de mortalidad que decidimos continuar utilizando en nuestra unidad como una de las estrategias de control de calidad y herramienta de comparación con otras unidades. ■

BIBLIOGRAFÍA

1. Gemke RBB, Bonsel GJ, van Vught AJ. Outcome assessment and quality assurance in pediatric intensive care. En: Update in Intensive Care and Emergency Medicine. Tibboel D, van der Voort. Springer-Verlag, Berlín, Heidelberg, 1996.
2. Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality score. Crit Care Med 1988; 16:1110-6.
3. Pollack MM, Patel K, Ruttimann UE. PRISM III: an updated Pediatric Risk of Mortality score. Crit Care Med 1996; 24:743-52.
4. Shann F, Pearson G, Slater A, Wilkinson K. Pediatric Index of Mortality: a mortality prediction model for children in intensive care. Intensive Care Med 1997; 23:201-7.
5. Slater A, Shann F, Pearson G. PIM2: a revised version of the Pediatric Index of Mortality. Intensive Care Med 2003; 29:278-85.
6. Leteurtre S, Leclerc F, Wirth J, Noizet O, Magnenant E, Sadik A, Fourier C, Cremer R. Can generic pediatric mortality scores calculated 4 hours after admission be used as inclusion criteria for clinical trials? Critical Care 2004; 8:185-193.
7. Slater A, Shann F. The suitability of PIM, PIM2, PRISM and PRISM III for monitoring the quality of pediatric intensive care in Australia and New Zealand. Pediatr Crit Care Med 2004; 5:447-54.
8. 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; 22:1385-91.
9. Gemke RBB, van Vught AJ. Scoring systems in pediatric intensive care: PRISM III versus PIM. Intensive Care Med 2002; 28:204-7.
10. Shann F. Are we doing a good job? PRISM, PIM and all that. Intensive Care Med 2002; 28:105-7.
11. Bertolini G, D'Amico R, Nardo D, Tiñáis A, Apolone G. One model, several results: the paradox of the Hosmer-Lemeshow goodness-of-fit test for the logistic regression model. J Epidemiol Biostat 2000; 5:251-3.
12. Saporiti A. Cuidados intensivos pediátricos: trabajo colaborativo sobre características de población y grupos de riesgo. Arch Argent Pediatr 1997; 95:66-74.
13. Marcin JP, Pollack MM. Review of the methodologies and applications of scoring systems in neonatal and pediatric intensive care. Pediatr Crit Care 2000; 1:20-27.
14. Pearson GA, Stickley J, Shann F. Calibration of the pediatric index of mortality in UK pediatric intensive care units. Arch Dis Child 2001; 84: 125-8.
15. Bertolini G, Ripamonti D, Cattaneo A, Apolone G. Pediatric Risk of Mortality: an assessment of its performance in a sample of 26 Italian intensive care units. Crit Care Med 1998; 26: 1427-32.
16. Tibby SM, Taylor D, Festa M, Hanna S, Hatherhill M, Jones G, Habibi P, Durward A, Murdoch IA. A comparison of

- three scoring systems for mortality risk among retrieved intensive care patients. Arch Dis Child 2002; 87: 421-5.
17. Lins R, Elseviers M, Daelemans R, De Broe M. Problems in the development, validation and adaptation of prognostic models for acute renal failure. Nephrol Dial Transplant 2001; 16:1098-1101.
 18. Justice A, Covinsky K, Berlin J. Assessing the generalizability of prognostic information. Ann Intern Med 1999; 130:515-24.
 19. Zimmerman J, Wagner D. Prognostic systems in intensive care: how do you interpret an observed mortality that is higher than expected. Crit Care Med 2000; 28:258-60.
 20. Pollack M. Quality of care: national or international standards? Crit Care Med 1998; 26:1313-14.
 21. Tilford J, Roberson P, Lensing S, Fiser D. Improvement in pediatric critical care outcomes. Crit Care Med 2000; 28:602-3.
 22. Teres D, Lemeshow S. Why severity models should be used with caution? Crit Care Clin 1994; 10:93-110.