

Validación prospectiva de siete sistemas locales e internacionales de evaluación del riesgo en cirugía cardíaca

RAÚL A. BORRACCI^{MTSAC, 1, 2, 3}, MIGUEL RUBIO^{MTSAC, 1, 2, 3}, GONZALO CORTÉS Y TRISTÁN^{1, 3}, MARIANO GIORGI^{†, 2}, RODOLFO A. AHUAD GUERRERO^{MTSAC, 2}

Recibido: 26/07/2006
Aceptado: 17/10/2006

Dirección para separatas:
Raúl A. Borracci
La Pampa 3030, 1° B
(1428) Buenos Aires, Argentina
e-mail: borracci@universia.com.ar

RESUMEN

Introducción

En la Argentina se elaboraron varios modelos de ajuste por riesgo (*scores*) de mortalidad en cirugía cardíaca que en general no fueron validados en poblaciones diferentes de las que les dieron origen. Asimismo, distintos cirujanos utilizan diferentes *scores* para comparar y presentar sus resultados.

Objetivos

El presente trabajo se llevó a cabo con el propósito de evaluar el desempeño de siete *scores* de riesgo de mortalidad en cirugía cardíaca de uso corriente en nuestro medio.

Material y métodos

Se analizaron prospectivamente 502 pacientes sometidos a cirugía cardíaca entre 2004 y 2005. Se identificaron las variables preoperatorias para el cálculo de siete *scores* de riesgo de mortalidad, cuatro internacionales, Parsonnet, Ontario, EuroSCORE y Cleveland, y tres locales, CONAREC, Español y Favaloro. Para determinar la discriminación, la precisión y la calibración de cada *score*, se calculó el riesgo esperado promedio de cada uno, la bondad de ajuste de Hosmer y Lemeshow, el coeficiente Q , las áreas ROC, el índice de Shannon y tres métodos gráficos.

Resultados

La comparación de las áreas ROC para todo tipo de cirugía cardíaca mostró que el EuroSCORE tuvo la mejor área (78,7%), junto con el *score* del Español (78,6%). Aunque estos modelos tuvieron el valor predictivo más alto, no se encontraron diferencias estadísticas al compararlos con el resto de los modelos. Para la cirugía coronaria, el EuroSCORE nuevamente mostró la mejor predicción (76,3%), y luego el de Cleveland (76,0%), a pesar de que tampoco se hallaron diferencias estadísticas con el resto de los *scores*. Para todo tipo de cirugía, el Ontario tuvo el mejor índice de Shannon global (0,913), pero el peor índice de fallecidos (0,074), lo cual indica que otorgó riesgos esperados más bajos a los sobrevivientes, pero también riesgos bajos a los que murieron. Por su parte, el EuroSCORE se ubicó en segundo lugar para el índice global (0,905) y el de fallecidos (0,106). Para la cirugía coronaria, el Español tuvo el índice global más alto (0,945), pero también el índice de fallecidos más bajo (0,050), lo que señala una buena precisión en el primer caso y una mala precisión cuando sólo se consideran los fallecidos. En el EuroSCORE, los índices de Shannon de 0,926 y 0,073 lo ubican ligeramente en mejor posición al evaluar los dos tipos de valores.

Conclusiones

La comparación de siete *scores* para estratificar el riesgo de mortalidad en la cirugía cardíaca demostró que el EuroSCORE tuvo el mejor desempeño, tanto para los procedimientos valvulares como para los coronarios. Por su parte, el *score* del Español tuvo un desempeño similar cuando se consideraron todos los tipos de cirugía, mientras que el de Cleveland fue mejor para estratificar el riesgo de los coronarios. De acuerdo con el índice de Shannon, la precisión para predecir mortalidad a nivel del paciente individual fue pobre con todos los modelos.

REV ARGENT CARDIOL 2006;74:458-464.

Palabras clave > Cirugía cardíaca - Riesgo - Mortalidad

INTRODUCCIÓN

Los modelos de ajuste por riesgo (*scores*) para predecir la mortalidad de la cirugía cardíaca son impres-

cindibles para evaluar la calidad de los resultados y comparar grupos quirúrgicos distintos. (1) La aplicación adecuada de estos *scores* de predicción requiere la validación externa, o sea, la evaluación del desem-

Servicios de Cirugía Cardíaca y Cardiología de la Clínica Bazterrica¹, Buenos Aires, de la Corporación Médica de Gral. San Martín², San Martín, y del Sanatorio Prof. Itoiz³, Avellaneda, Buenos Aires, Argentina

^{MTSAC} Miembro Titular de la Sociedad Argentina de Cardiología

[†] Para optar a Miembro Titular de la Sociedad Argentina de Cardiología

peño del modelo en la población en la que se pretende aplicar. A su vez, este desempeño se basa en la bondad de ajuste y en el poder de discriminación del método, según el riesgo de ocurrencia del evento. (2)

En la Argentina se elaboraron varios *scores* de mortalidad que no fueron validados en poblaciones y grupos quirúrgicos diferentes de los que les dieron origen. (3-7) Asimismo, distintos cirujanos utilizan diferentes *scores* locales o internacionales para comparar y presentar sus resultados.

El objetivo de este trabajo fue evaluar el desempeño de siete *scores* de riesgo de mortalidad en cirugía cardíaca, de uso corriente en nuestro medio.

MATERIAL Y MÉTODOS

Se analizaron en forma prospectiva los datos de 502 pacientes sometidos a cirugía cardíaca en forma consecutiva entre 2004 y 2005, en tres servicios quirúrgicos asociados. En una base de datos *ad hoc* se incorporaron las variables preoperatorias necesarias para el cálculo de siete *scores* de riesgo de mortalidad quirúrgica, cuatro de uso internacional, Parsonnet, (8) Ontario, (9) EuroSCORE logístico (10) y Cleveland; (11) y tres *scores* de desarrollo local, CONAREC, (3) Español (4, 12) y Favaloro. (6) La selección de estos modelos de ajuste por riesgo dependió de su aceptación general y de la accesibilidad a las técnicas para su cálculo. Para el caso de los *scores* desarrollados exclusivamente para la cirugía coronaria (Cleveland, CONAREC y Favaloro), sus puntajes se determinaron y compararon en una muestra de 347 coronarios. Una vez calculada la mortalidad esperada con cada *score*, se cotejó con la mortalidad observada para el grupo total de cirugías y para la muestra individual de coronarios, respectivamente.

Método de validación

Se usaron distintos modelos de análisis para determinar la discriminación, la precisión (*accuracy*) y la calibración de cada *score*. En primer lugar se calculó el riesgo esperado promedio de cada *score*, junto con su intervalo de confianza del 95% ($IC_{95\%}$), la prueba de bondad de ajuste chi cuadrado de Hosmer y Lemeshow y el coeficiente Q (*likelihood*). (13)

Curvas ROC (*receiver operating characteristic curve*)

Se determinaron las áreas bajo las curvas ROC de cada uno de los *scores*, junto con sus respectivos $IC_{95\%}$. La comparación de las áreas ROC se realizó con el método de Hanley-McNeil, (14) para lo cual fue necesario calcular el coeficiente de correlación r para cada par de *scores* a confrontar.

Índice de Shannon (S)

La precisión se evaluó con el índice de Shannon, el cual es un método basado en la teoría de la información, de acuerdo con el valor de entropía (cantidad de información aportada por el modelo de ajuste por riesgo). El grado de precisión se calculó en cada paciente mediante la comparación entre el valor predicho por cada *score* y el resultado observado, de acuerdo con la siguiente fórmula:

$$S = \{[(1 + o) \ln(1 + e) + 2o] \ln(2 - e) - \ln 2\} / \ln 2$$

donde \ln es el logaritmo natural, o es la presencia ($o = 1$) o la ausencia ($o = 0$) del evento muerte y e es el valor esperado obtenido con cada *score* para cada paciente. A partir de cada índice individual se calcularon el promedio más una desviación estándar (DE) para el total de los pacientes (índice de

Shannon global) y el promedio más una DE de los fallecidos solamente (índice de Shannon para fallecidos); de esta forma se pudo determinar la precisión de cada modelo en una y otra situación. Este índice de precisión puede oscilar entre 0 y 1 y este último valor es el de precisión perfecta. (15-17)

Métodos gráficos

Adicionalmente se usaron tres métodos gráficos para comparar el desempeño de los distintos *scores*. Por un lado se comparó en un gráfico la mortalidad observada (x) versus la esperada (y) para los diferentes modelos de predicción de riesgo separados por quintiles. Lo mismo se realizó para mostrar la variación del índice de Shannon con los distintos *scores* de riesgo. Finalmente se construyeron curvas VLAD (*variable life-adjusted display*), según fueran comunicadas previamente. (18, 19)

Análisis estadístico

Los datos se resumieron como promedios, desviación estándar (DE) o error estándar (EE). El uso de intervalos de confianza del 95% se indicó oportunamente en las tablas. La comparación de múltiples medias, como en el caso de los índices de Shannon, se realizó con ANOVA de medidas repetidas. Dado que no existen modelos apropiados para el cálculo del tamaño de la muestra para comparar curvas ROC, la elección se basó en el número elegido por otros autores en comparaciones múltiples similares. (17, 20)

RESULTADOS

Las características preoperatorias de la población y los resultados quirúrgicos se resumen en la Tabla 1. La comparación de las áreas ROC para todo tipo de cirugía cardíaca (Tabla 2a) mostró que el EuroSCORE tuvo la mejor área (78,7%), junto con el *score* local del Español (78,6%). Aunque estos dos modelos tuvieron el valor predictivo más alto, no se encontraron diferencias estadísticas al compararlos con el resto de los modelos (Tabla 3). Para el caso de la cirugía coronaria (Tabla 2b), el EuroSCORE nuevamente mostró la mejor predicción (76,3%), y en segundo lugar el de la Cleveland (76,0%), a pesar de que tampoco se hallaron diferencias estadísticas con el resto de los *scores* (Tabla 3). La inclusión del EuroSCORE entre los modelos de predicción en cirugía coronaria se debió a que éste tuvo el mejor desempeño global.

La precisión de los modelos evaluada con el índice de Shannon se consideró en función de su índice global que incluye vivos y muertos, y de su índice exclusivo para predecir muertes (índice fallecidos) (Tabla 2). En este estudio se halló que en el análisis de todo tipo de cirugía el *score* de Ontario tuvo el mejor índice global (0,913), pero el peor índice de fallecidos (0,074); esto indica que le otorgó riesgos esperados más bajos a los que sobrevivieron, pero también riesgos bajos a los que murieron. Por su parte, el EuroSCORE se ubicó en el segundo lugar tanto para el mejor valor del índice global (0,905) como el de fallecidos (0,106). Para el caso de la cirugía coronaria, el *score* del Español tuvo el índice global más alto (0,945), pero también el índice de fallecidos más bajo (0,050), lo cual señala una buena precisión en el primer caso y una mala precisión cuando sólo se conside-

Tabla 1. Características de la población (n = 502)

VARIABLES	n (%)
Preoperatorio	
Edad (media ± DE)	63,5 ± 10,21 años
Sexo femenino	120 (23,9)
Diabetes	99 (19,7)
Hipertensión	326 (64,9)
Obesidad	95 (18,9)
Insuficiencia cardíaca	58 (11,5)
ACV	24 (4,7)
Enfermedad pulmonar	55 (10,9)
Disfunción renal	30 (5,9)
Anemia	12 (2,4)
Angor inestable	249 (49,6)
Infarto reciente	41 (8,2)
Reoperación	15 (2,9)
Disfunción VI moderado-severo	166 (33,0)
Tipo de cirugía:	
Coronario	347 (69,1)
Valvular	93 (18,5)
Combinado	62 (12,4)
Cirugía de urgencia	90 (17,9)
Posoperatorio	
Mortalidad	24 (4,8)
Complicaciones:	
- Reoperación por sangrado	9 (1,8)
- Infarto (tipo Q)*	9 (2,6)
- ACV	10 (1,9)
- Diálisis	9 (1,8)
- Mediastinitis	7 (1,4)

ACV: Accidente cerebrovascular. VI: Ventrículo izquierdo. DE: Desviación estándar.

* Calculado sobre las cirugías coronarias exclusivamente (n = 347).

ran los fallecidos. En el caso del EuroSCORE, los valores de 0,926 y 0,073 lo ubican en una posición ligeramente mejor al evaluar los dos tipos de índices. De todas formas, los bajos índices de Shannon encontrados en los grupos de fallecidos demuestran la poca precisión de todos los modelos con este método. En el gráfico de la Figura 1 se muestra la variación del índice de Shannon por quintil en uno de los grupos (todo tipo de cirugía). En este caso se observa cómo disminuye la precisión de este índice para todos los modelos a medida que se avanza a los quintiles con mayor número entre mortalidad observada y esperada para cada quintil. Por su parte, los *scores* de Parsonnet, Español y CONAREC parecen sobrestimar el riesgo, especialmente en los quintiles con mayor riesgo. En

cambio, los *scores* de Ontario, Cleveland y Favaloro subestiman el riesgo de la misma forma.

Por último, los gráficos de la Figura 3 representan las curvas VLAD de cada modelo de ajuste por riesgo. La variación del valor de vida neta ganada a través del tiempo muestra que la curva del EuroSCORE es la que más se acerca al valor "cero" ideal. De la misma manera que en los gráficos de la Figura 2 de mortalidad observada *versus* esperada, los *scores* de Parsonnet, Español y CONAREC sobrestiman el riesgo, mientras que los de Ontario, Cleveland y Favaloro lo subestiman.

DISCUSIÓN

La comparación de los distintos *scores* de riesgo sobre la base de las curvas ROC, del índice de Shannon y de dos métodos gráficos demostró que el EuroSCORE tuvo en todos los casos una tendencia al mejor desempeño para predecir la mortalidad hospitalaria de la cirugía cardíaca, tanto para los procedimientos valvulares como para los coronarios. Por su parte, el *score* local del Español tuvo un desempeño similar cuando se consideraron todos los tipos de cirugía en conjunto, mientras que el de Cleveland fue mejor para la cirugía coronaria. Aunque en el caso de las curvas ROC no se observaron diferencias estadísticas, el peso en conjunto de todos los métodos de evaluación favoreció en primer lugar al EuroSCORE y luego al *score* del Español.

En particular, las áreas de las curvas ROC de cada *score* se hallaron entre 0,71 y 0,79, lo cual indicaría un nivel de predicción "aceptable" en todos los casos. En especial para el EuroSCORE, la validación realizada en otros países mostró un área ROC de entre 0,73 y 0,83, con dependencia del tipo de cirugía. (17, 20-23)

Pocos estudios han comparado varios *scores* a la vez. Geissler y colaboradores (20) confrontaron seis *scores* de riesgo (entre ellos, Parsonnet, Ontario, Cleveland y EuroSCORE) y encontraron que el EuroSCORE tuvo el valor predictivo más alto, con un área ROC similar a la de nuestro estudio (0,786), a pesar de la falta de diferencia estadística con los otros modelos, debido posiblemente a la baja potencia de la muestra (n = 505 pacientes). Estos mismos autores manifiestan la dificultad de calcular un tamaño de muestra adecuado, habida cuenta de que los *scores* comparados no dan información sobre el tamaño mínimo de muestra para obtener la mejor predicción posible. Por su parte, Pinna-Pintor y colaboradores (17) cotejaron cuatro modelos de ajuste por riesgo en cirugía coronaria exclusivamente: Parsonnet, Cleveland, French y EuroSCORE. En este caso, las áreas ROC fueron 0,60, 0,86, 0,82 y 0,81, respectivamente, con el *score* de la Cleveland en el mejor lugar (n = 418 pacientes).

Como ya se señaló, la precisión evaluada con el índice de Shannon debe considerarse en función de su índice global de vivos y muertos, y de su índice exclusivo para predecir muertes. Un *score* que otorga valores de riesgo muy bajos para los que sobrevive-

Tabla 2. Áreas ROC e índices de precisión para los distintos *scores* de riesgo aplicados a todo tipo de cirugía cardíaca (a) y a cirugía coronaria exclusivamente (b)

(a)	EuroSCORE	Español	Ontario	Parsonnet
Riesgo promedio (%)	6,25	9,16	4,41	9,20
IC _{95%}	5,64 a 6,85	8,31 a 9,99	4,01 a 4,81	8,36 a 10,0
Hosmer-Lemeshow (<i>p</i>)	0,176	0,508	0,337	0,336
(g.l.)	(8)	(8)	(6)	(3)
Q (<i>likelihood</i>)	0,697	0,674	0,676	0,658
Área ROC	0,787	0,786	0,747	0,708
EE	0,0391	0,0428	0,0469	0,0471
IC _{95%}	0,710 a 0,863	0,701 a 0,869	0,655 a 0,838	0,615 a 0,800
Índices de Shannon				
Índice global	0,905 ± 0,225	0,884 ± 0,226	0,913 ± 0,232	0,877 ± 0,230*
Índice fallecidos	0,106 ± 0,049	0,112 ± 0,061	0,074 ± 0,053	0,095 ± 0,062*

*ANOVA de medidas repetidas: Índice global, *p* = 0,059; índice fallecidos, *p* = 0,004.

(b)	EuroSCORE	Cleveland	CONAREC	Favaloro
Riesgo promedio (%)	5,28	1,86	14,9	2,24
IC _{95%}	4,67 a 5,89	1,59 a 2,13	13,4 a 16,4	2,04 a 2,43
Hosmer-Lemeshow (<i>p</i>)	0,634	0,556	0,442	0,091
(g.l.)	(8)	(4)	(3)	(4)
Q (<i>likelihood</i>)	0,737	0,721	0,626	0,729
Área ROC	0,763	0,760	0,751	0,709
EE	0,0607	0,0619	0,0549	0,0675
IC _{95%}	0,643 a 0,881	0,638 a 0,880	0,643 a 0,858	0,576 a 0,841
Índices de Shannon				
Índice global	0,926 ± 0,198	0,945 ± 0,203	0,839 ± 0,228	0,943 ± 0,203*
Índice fallecidos	0,073 ± 0,056	0,050 ± 0,031	0,109 ± 0,025	0,053 ± 0,053*

*ANOVA de medidas repetidas: Índice global, *p* = 0,000; índice fallecidos, *p* = 0,000.

ROC: Receiver-operating-characteristic curve.

EE: Error estándar. IC_{95%}: Intervalo de confianza del 95%. g.l.: Grados de libertad.

Tabla 3. Comparación de las áreas ROC (Hanley-McNeil [14])

	EuroSCORE		
	r	z	p
Parsonnet	0,629	1,290	0,099
Ontario	0,581	0,655	0,256
Español	0,489	0,024	0,491
Cleveland	0,466	0,812	0,208
CONAREC	0,361	0,180	0,429
Favaloro	0,437	0,050	0,480

ROC: Receiver-operating-characteristic curve.

ron y probabilidades de muerte muy altas a los que fallecieron tendrá índices de Shannon globales y de fallecidos muy altos (buena precisión) y viceversa. De acuerdo con este índice, la precisión de la mortalidad global (sobrevivientes y fallecidos) se halló en todos los modelos entre 0,88 y 0,95, lo cual indica un buen

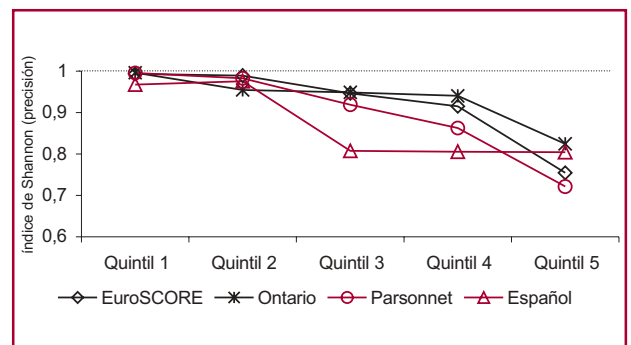


Fig. 1. Variación del índice de Shannon (precisión) por quintil.

grado de desempeño de cada *score*. De todas maneras, dadas las características del cálculo de este índice, estos valores pueden tener una variación importante según se consideren grupos con mayor o menor riesgo (mayor o menor número de fallecidos). El gráfico de la Figura 1 muestra cómo el índice de Shannon

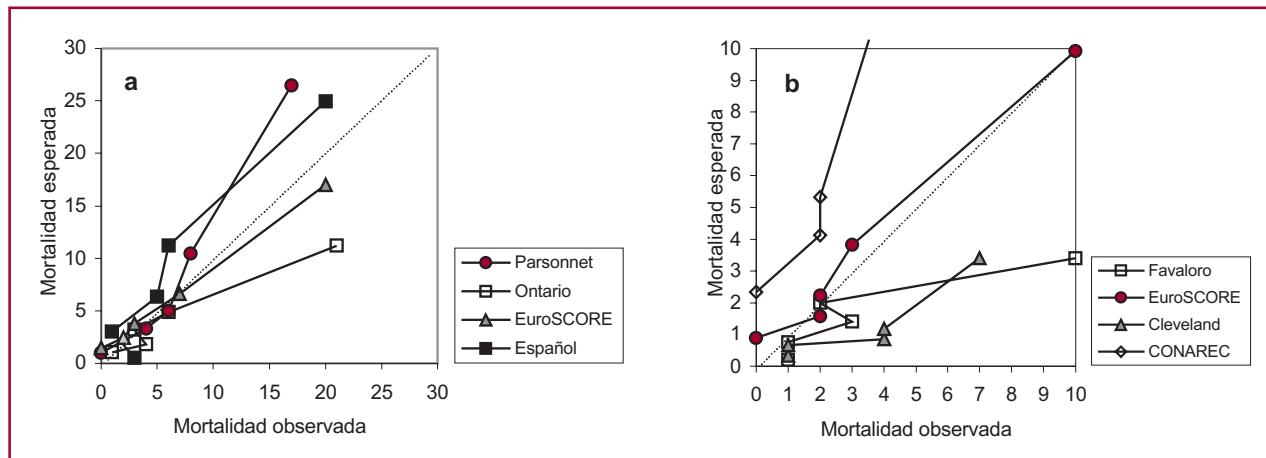


Fig. 2. Comparación de la mortalidad observada versus la esperada para diferentes modelos de predicción de riesgo separados por quintiles. a. Todo tipo de cirugía. b. Sólo coronarios.

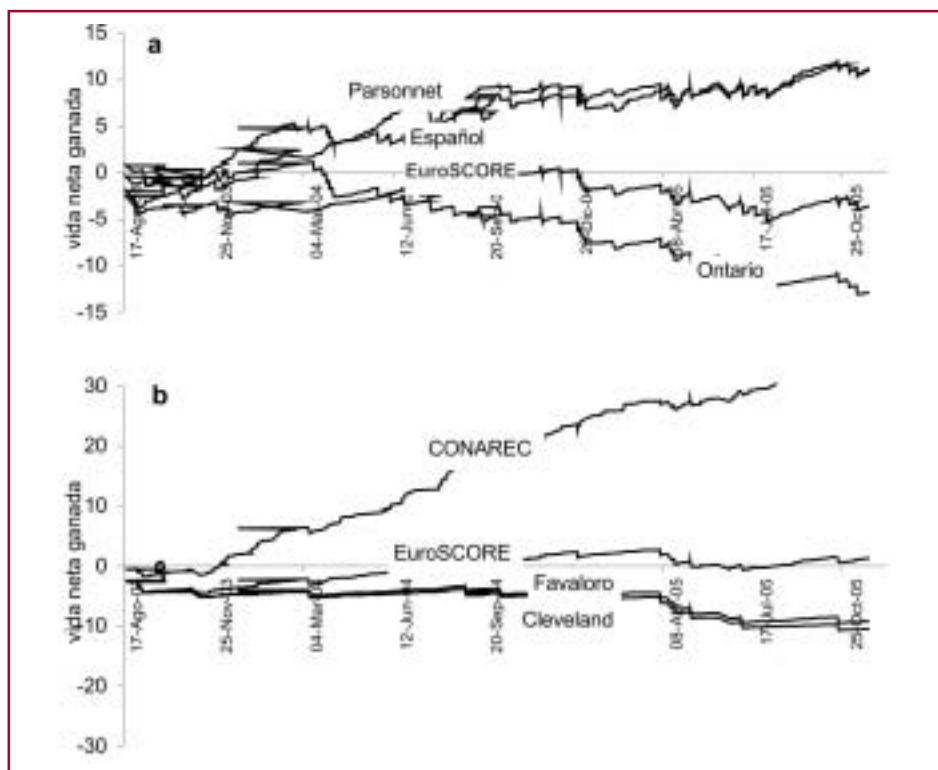


Fig. 3. Curvas VLAD comparativas entre distintos scores de riesgo para todo tipo de cirugía cardíaca (a) y para cirugía coronaria (b). VLAD: Variable life-adjusted display.

puede cambiar desde valores óptimos de precisión cercanos a 1,0 hasta aproximadamente 0,7, según el nivel de riesgo (quintil) que se considere, con la consiguiente pérdida de precisión. Asimismo, cuando se calculó el índice de Shannon exclusivamente para los fallecidos (Tabla 2), dichos valores se redujeron hasta 0,05 y 0,112, indicativos de una precisión muy baja. Datos similares se encontraron en Pinna-Pintor y colaboradores, (17) que calcularon el índice de Shannon para sus cuatro modelos. En ese caso, el índice global estuvo entre 0,97 y 0,98, mientras que el índice exclusivo para los que fallecieron varió entre 0,07 y 0,15.

Estas características especiales del índice de Shannon se deben a que el método adjudica el valor 0 (cero) al paciente que sobrevive y 1 (uno) al que fallece, mientras que, en forma ideal, todos los scores adjudicarían 0 o un riesgo muy bajo (cercano a 0) al que sobrevive y un puntaje que excepcionalmente superaría 0,5 para el paciente con mayor riesgo de fallecer en el posoperatorio. En resumen, aquellos scores que adjudiquen el riesgo más cercano a 0 a quienes sobrevivan y más cercano a 1 a quienes fallezcan tendrán el mejor grado de precisión basado en este índice. En este sentido, el índice de Shannon es el mejor método para

determinar la precisión de un *score* para predecir el riesgo de mortalidad a nivel del paciente individual, mientras que el resto de los modelos de análisis enfocan la predicción sobre poblaciones.

Para interpretar los gráficos de la Figura 2, que comparan las mortalidades observadas y esperadas en dos ejes, debe tenerse en cuenta que para cada *score* de riesgo se indican los quintiles correspondientes a la mortalidad observada (x) y la esperada (y). La línea diagonal de la figura representa la coincidencia o precisión óptima; los puntos por encima de la diagonal indican una sobrestimación del riesgo por el modelo, mientras que los puntos por debajo señalan que se subestima el riesgo con dicho *score*. Del análisis visual de los gráficos se observa que el EuroSCORE parece tener la mejor relación entre mortalidad observada y esperada para cada quintil.

La validez de determinado *score* para ser usado en una población distinta de la que fue creado depende de múltiples factores relacionados con el diseño del modelo de ajuste. Por ejemplo, la población basal de la que se generó el *score* podría tener características epidemiológicas diferentes. Además, los pacientes podrían pertenecer a uno o múltiples centros, y sus datos haber sido recolectados en forma prospectiva o retrospectiva. Asimismo, el modelo de ajuste puede haber sido validado o no en forma prospectiva. En el caso del EuroSCORE, su diseño fue multicéntrico y prospectivo sobre casi 20.000 pacientes y su validación incluye varias publicaciones de distintos países fuera de Europa. Por su parte, el *score* del Español fue desarrollado por un solo centro y en forma retrospectiva, aunque su validación se realizó recientemente en forma prospectiva en un centro quirúrgico diferente. (5)

La investigación de los resultados de la cirugía cardíaca tiene una importancia fundamental, en la medida en que los sistemas de salud demandan datos estadísticos confiables sobre los riesgos y el pronóstico con esta terapéutica. Aunque el análisis sistemático de varios *scores* de mortalidad realizado en este estudio tiene un interesante valor comparativo, la confrontación en trabajos futuros de modelos predictores de morbilidad generará más información sobre otros tipos de resultados, como son la calidad de vida posterior al procedimiento y los costos.

CONCLUSIONES

La comparación de siete *scores* locales e internacionales para estratificar el riesgo de mortalidad en la cirugía cardíaca demostró que el EuroSCORE tuvo el mejor desempeño, tanto para los procedimientos valvulares como para los coronarios. Por su parte, el *score* local del Español tuvo un desempeño similar cuando se consideraron todos los tipos de cirugía en conjunto, mientras que el de Cleveland fue mejor para la estratificación del riesgo de la cirugía coronaria. Por su parte, y de acuerdo con el índice de Shannon, la precisión para predecir el riesgo de mortalidad a nivel del paciente individual fue pobre con todos los modelos de ajuste.

SUMMARY

Prospective Validation of Seven Local and International Systems for Risk Assessment in Cardiac Surgery

Background

In Argentina, several models of risk adjustment (scores) in cardiac surgery have been proposed, which generally have not been validated in populations other than those originally described. Additionally, different surgeons utilize different scores to compare and present their results.

Objectives

This study was performed with the goal of assessing the performance of seven scores of mortality risk in cardiac surgery, which are commonly used in our country.

Material and Methods

A group of 502 patients who underwent surgery between 2004 and 2005 were analyzed prospectively. Preoperative variables were identified for the calculation of seven mortality risk scores: four international scores: Parsonnet, Ontario, EuroSCORE and Cleveland, and three local scores: CONAREC, Español and Favaloro. To assess the discrimination, accuracy and calibration of each score, the following were calculated: expected average risk of each score, Hosmer and Lemeshow goodness of fit, coefficient Q, ROC areas, Shannon index and three graphic methods.

Results

A comparison of the ROC areas for all types of cardiac surgery showed that the EuroSCORE had the best area (78.7%), together with the Español score (78.6%).

Although these models had the higher predictive value, no statistical differences were found when they were compared to the other models. For coronary surgery, the EuroSCORE again was the most predictive (76.3%), followed by the Cleveland (76%), although here too, there were no statistical difference with the other scores. For all types of surgery, the Ontario score had the best global Shannon index (0.913), but had the worst index of deaths (0.074), which shows that it assigned lower expected risks to survivors, but it also assigned low risks to those who died. The EuroSCORE ranked second for the global index (0.905) and for deaths (0.106). For coronary surgery, the Español score had the higher global index (0.945), but also the lowest index of deaths (0.050), which indicates good accuracy in the first case and poor accuracy only when the deceased are taken into account. The EuroSCORE is positioned slightly better by the Shannon indices of 0.926 and 0.073, when assessing both types of values.

Conclusions

The comparison of seven scores to stratify mortality risk in cardiac surgery showed that the EuroSCORE had the best performance, for valvular as well as for coronary procedures. As to the Español score, it had a similar performance when all types of surgery were considered, whereas the Cleveland score was better to stratify risk in coronary disease. According to Shannon's index, the accuracy to predict mortality in the individual patient was poor with all the models assessed.

Key words > Cardiac Surgery - Risk - Mortality

BIBLIOGRAFÍA

1. Little WC, Constantinescu M, Applegate RJ, Kutcher MA, Burrows MT, Kahl FR, et al. Can coronary angiography predict the site of a

- subsequent myocardial infarction in patients with mild-to-moderate coronary artery disease? *Circulation* 1988;78:1157-66.
2. Ferrante D. Scores de predicción de mortalidad en cirugía cardíaca. *Rev Argent Cardiol* 2005;73:326 (editorial).
 3. Ciruzzi M, Aranda G, Bozovich G, Henquin R, Vallejos J, Cherro A y col. CONAREC III. Predicción del riesgo de mortalidad intrahospitalaria en pacientes sometidos a cirugía coronaria. *Rev Argent Cardiol* 1996;64:79-90.
 4. Carosella VC, Cárdenas C, Sellanes M, Russo M, Pensa C, Nojek C. Score de predicción de riesgo preoperatorio de muerte intrahospitalaria en 3.611 pacientes sometidos a cirugía cardíaca. *Rev Argent Cardiol* 2001;69:94 (abstract).
 5. Carosella VC, Sellanes M, Cáceres M, Rodríguez W, Pensa C, Cárdenas C y col. Validación de un score de predicción de fácil aplicación de mortalidad en cirugía cardíaca. *Rev Argent Cardiol* 2005;73:139 (abstract).
 6. Gabe E, Giunta G, Macchia A, Raffaelli H, Abud J, Dulbecco E y col. Score para estimar mortalidad en cirugía coronaria. *Rev Argent Cardiol* 2003;71:124 (abstract).
 7. Nani S, Tenorio Núñez O, Damonte M, Devoro M, Rivas R, Mancini B. Score quirúrgico. Validación del análisis de riesgo en una población de pacientes sometidos a cirugía cardíaca. *Rev Argent Cardiol* 2003;71 (Supl 3):177 (abstract).
 8. Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;79:13-12.
 9. Tu JV, Jaglal SB, Naylor CD. Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery. Steering Committee of the Provincial Adult Cardiac Care Network of Ontario. *Circulation* 1995;91:677-84.
 10. Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999;16:9-13.
 11. Higgins TL, Estafanous FG, Loop FD, Beck GJ, Blum JM, Paranandi L. Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients. A clinical severity score. *JAMA* 1992;267:2344-8.
 12. Nojek CA, Eguren Maxwell A, Cárdenas C. Estenosis aórtica. Tratamiento quirúrgico. En: Trongé JE y col. *Emergencias cardiovasculares en el gerente*. Buenos Aires: Cesarini Hnos Editores; 2002. p. 213-9.
 13. Shapiro AR. The evaluation of clinical predictions. A method and initial application. *N Engl J Med* 1977;296:1509-14.
 14. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839-43.
 15. Diamond GA, Hirsch M, Forrester JS, Staniloff HM, Vas R, Halpern SW, et al. Application of information theory to clinical diagnostic testing. The electrocardiographic stress test. *Circulation* 1981;63:915-21.
 16. Hlatky M, Botvinick E, Brundage B. Diagnostic accuracy of cardiologists compared with probability calculations using Bayes' rule. *Am J Cardiol* 1982;49:1927-31.
 17. Pinna-Pintor P, Bobbio M, Colangelo S, Veglia F, Giammaria M, Cuni D, et al. Inaccuracy of four coronary surgery risk-adjusted models to predict mortality in individual patients. *Eur J Cardiothorac Surg* 2002;21:199-204.
 18. Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life-adjusted display. *Lancet* 1997;350:1128-30.
 19. Borracci RA, Rubio M, Cortés y Tristán G, Mémoli R, Giorgi M, Ahuad Guerrero RA. Validez temporal de los sistemas de estratificación por riesgo para la monitorización continua de los resultados de la cirugía cardíaca. *Rev Argent Cardiol* 2005;73:341-5.
 20. Geissler HJ, Holzl P, Marohl S, Kuhn-Regnier F, Mehlhorn U, Sudkamp M, et al. Risk stratification in heart surgery: comparison of six score systems. *Eur J Cardiothorac Surg* 2000;17:400-6.
 21. Nashef SA, Roques F, Hammill BG, Peterson ED, Michel P, Grover FL, et al; EuroSCORE Project Group. Validation of European System for Cardiac Operative Risk Evaluation (EuroSCORE) in North American cardiac surgery. *Eur J Cardiothorac Surg* 2002;22:101-5.
 22. Karthik S, Srinivasan AK, Grayson AD, Jackson M, Sharpe DA, Keenan DJ, et al. Limitations of additive EuroSCORE for measuring risk stratified mortality in combined coronary and valve surgery. *Eur J Cardiothorac Surg* 2004;26:318-22.
 23. Yap CH, Reid C, Yui M, Rowland MA, Mohajeri M, Skillington PD, et al. Validation of the EuroSCORE model in Australia. *Eur J Cardiothorac Surg* 2006;29:441-6.