

Original

## Validación del EuroSCORE II: ¿funciona en nuestro medio?

Jacobo Silva, Manuel Carnero, Fernando Reguillo, Javier Cobiella, Enrique Villagrán, Lorena Montes, Zaadi Garcés, Ali Ayaon, Luis Maroto, Ali Alswies y Enrique Rodríguez

Servicio de Cirugía Cardíaca, Hospital Clínico San Carlos, Madrid, España

### INFORMACIÓN DEL ARTÍCULO

#### Historia del artículo:

Recibido el 11 de abril de 2012  
Aceptado el 16 de marzo de 2013

Palabras clave:  
EuroSCORE II  
EuroSCORE  
Validación

### RESUMEN

**Objetivos:** Validar la nueva escala European System for Cardiac Operation Risk Evaluation II (EuroSCORE II) (ESII) en términos de calibración y capacidad discriminatoria, comparar la distribución de las variables entre nuestra base de datos y el ESII, así como comparar estos resultados con el EuroSCORE logístico clásico (ES).

**Material y métodos:** Se analizó una población de 4.166 pacientes intervenidos durante un período de 7 años. Se estudió tanto para el ES como el ESII, la discriminación mediante curvas ROC y la calibración mediante el test de Hosmer-Lemeshow.

**Resultados:** La puntuación media del ESII fue superior a la del ES: 9,1 (desviación estándar:10,4) vs. 3,46 (desviación estándar: 4,3);  $p < 0,001$ . El área bajo la curva ROC para el ES fue 0,75 (intervalo de confianza 95% 0,72-0,78) y 0,78 (intervalo de confianza 95% 0,75-0,81) para el ESII ( $p = 0,233$ ). El test de Hosmer-Lemeshow mostró una pobre calibración para las 2 escalas: ES ( $\chi^2 = 26,6$ ,  $p = < 0,001$ ) y ESII ( $\chi^2 = 58,19$ ,  $p < 0,0001$ ). El porcentaje de cirugía coronaria (46,7 vs. 28,9%), la edad media (67 vs. 64,6), así como el sexo femenino (37,7 vs. 30,9%) mostraron una distribución diferente al comparar ambas bases de datos, sin embargo no se pudo establecer ningún tipo de análisis estadístico.

**Conclusiones:** El nuevo sistema de estratificación de riesgo ESII obtiene una mortalidad esperada inferior al ES y tiene un valor predictivo adecuado en términos de capacidad discriminatoria; sin embargo, muestra una pobre calibración. Se observa una tendencia a una mejor capacidad discriminatoria respecto al ES, aunque dicho hallazgo no muestra significación estadística. Se observa una distribución diferente entre ciertas variables al comparar nuestros datos con la base de datos del ESII.

© 2012 Sociedad Española de Cirugía Torácica-Cardiovascular. Publicado por Elsevier España, S.L. Todos los derechos reservados.

## EuroSCORE II validation. Does it work in our patients?

### ABSTRACT

Keywords  
EuroSCORE II  
EuroSCORE  
Validation

**Objectives:** Validate the new EuroSCORE (ESII) risk model in terms of discriminative and calibration power and compare this results with the classic EuroSCORE (ES). We also compare our data distribution with the ESII database.

**Methods:** A 4166 patient population operated during a 7 year period was analyzed. The model was then tested on the validation data set for calibration (by comparing the observed and predicted mortality) and for discrimination (using the area under the ROC curve).

**Results:** The predicted mortality by the ESII was higher than the ES: 9.1(SD: 10.4) vs 3.46 (SD: 4.3):  $p < 0.001$ . The Hosmer-Lemeshow test showed a poor calibration for both models: ES ( $\chi^2 = 26.6$ ,  $p = < 0.001$ ), ESII ( $\chi^2 = 58.19$ ,  $p < 0.0001$ ). Areas under ROC curves were 0.75 (IC95% 0.72-0.78) for ES and 0.78 (IC95% 0.75-0.81) for ESII ( $p < 0.233$ ).

**Conclusions:** The new EuroSCORE II risk model has a predicted mortality lower than EuroSCORE and a good predictive value in terms of calibration and poor discrimination. A non significant better discrimination power was observer in the ESII. The distribution of some variables was different between our data and ESII.

© 2012 Sociedad Española de Cirugía Torácica-Cardiovascular. Published by Elsevier España, S.L. All rights reserved.

\*Autor para correspondencia. Servicio de Cirugía Cardíaca, Hospital Clínico San Carlos, C/ Profesor Martín Lagos s/n, 7.<sup>a</sup> Planta Norte, 28040 Madrid, España. Tel.: 913303691  
Correo electrónico: jsilva8252@yahoo.es (J. Silva).

## Introducción

La escala de riesgo European System for Cardiac Operation Risk Evaluation<sup>1</sup> (EuroSCORE) (ES) es un modelo de riesgo que permite predecir la mortalidad postoperatoria después de una cirugía cardíaca mayor. Esta escala fue publicada en 1999<sup>2</sup>. Desde entonces, ha sido ampliamente utilizada en nuestro medio para predecir la mortalidad postoperatoria<sup>3</sup> y como criterio de evaluación de la calidad de los resultados de mortalidad hospitalaria. Esta escala ha sido validada en incontables publicaciones demostrando una excelente capacidad discriminatoria para distintos tipos de cirugía cardíaca<sup>4</sup>. Sin embargo, en los últimos años, distintas publicaciones han puesto en evidencia que esta escala pudiera estar sobreestimando el riesgo postoperatorio en algunos subgrupos como en pacientes octogenarios, cirugía valvular aórtica y cirugía coronaria sin circulación extracorpórea<sup>5-7</sup>. Esta mala calibración pudiera explicarse por los avances técnicos y tecnológicos en cirugía cardíaca, anestesiología y perfusión, aspectos que han conseguido una disminución de la mortalidad ajustada al riesgo.

En febrero de 2012, fue publicada la nueva escala EuroSCORE II (ESII)<sup>8</sup>. Este modelo predictivo de mortalidad postoperatoria se construyó con base en los resultados quirúrgicos observados en 22.381 pacientes intervenidos en 154 hospitales de todo el mundo durante un período de 12 semanas (mayo-julio 2010). Este nuevo modelo ha demostrado una capacidad discriminatoria similar al ES, con una mejor calibración<sup>8</sup>. Como aspectos novedosos, cabe destacar que el nuevo modelo de ESII incluye nuevas variables predictoras como la clase funcional New York Heart Association, la diabetes mellitus insulíndependiente (DMID) e introduce el aclaramiento de creatinina estimada como medida de la función renal en lugar de la creatinina sérica<sup>8</sup>. Además, elimina la rotura septal postinfarto e introduce una nueva escala de complejidad quirúrgica con base en el número de procedimientos realizados. Al igual que con el modelo original, es necesario, desde este momento, comprobar la validez externa de la nueva escala para garantizar su aplicabilidad. Nuestro grupo publicó en octubre de 2012 la validación de esta nueva escala en un grupo de 3.800 pacientes<sup>9</sup> y observamos una adecuada capacidad discriminatoria y una pobre calibración.

El objetivo del presente estudio fue validar la nueva escala ESII con una muestra mayor de pacientes en términos de calibración y capacidad discriminatoria, así como comparar estos resultados con el ES en los pacientes sometidos a cirugía cardíaca mayor en nuestro centro. También tratamos de analizar las diferencias poblacionales entre la base de datos del ESII y nuestro medio.

## Materiales y métodos

### Población a estudio

Se evaluó de forma ambispectiva y consecutiva toda la cohorte de pacientes intervenida de cirugía cardíaca mayor en nuestro centro desde enero de 2005 hasta enero de 2012 (ambos inclusive). Fueron incluidos todos aquellos sujetos adultos sometidos a procedimientos quirúrgicos mayores: cirugía coronaria aislada (con y sin circulación extracorpórea), cirugía valvular aislada (fuesen sustituciones por prótesis o reparaciones, por vía convencional o mínimamente invasiva), cirugía combinada valvular y coronaria, cirugía de la aorta torácica y otros procedimientos (cardiopatías congénitas del adulto, complicaciones mecánicas del infarto, ablación quirúrgica de arritmias auriculares o ventriculares, pericardiectomías, cirugía tumoral y traumatismos cardíacos).

Se incluyeron todas estas cirugías independientemente del nivel de prioridad (electivo/urgente/emergente). Se excluyeron los procedimientos menores, la implantación transcatéter de prótesis valvulares y los trasplantes cardíacos. Fueron excluidos del estudio todos

aquellos pacientes sometidos a otros procedimientos quirúrgicos cardíacos o aquellos a quien no fue posible estimar alguna de las puntuaciones de las escalas a estudio (ES o ESII).

### Recogida de datos, variables y eventos analizados

De forma prospectiva, se recogió información pre, intra y postoperatoria en la base de datos SICCS (Sistema Informático Cirugía Cardíaca Sorin) (Biomenco, Madrid, España) codificada a través de más de 100 variables. Dicha base de datos incluye el ES<sup>3</sup> en toda la muestra de pacientes de forma previa al momento de la cirugía. Los resultados fueron revisados antes del alta por un miembro del equipo como control de calidad interna de la recogida de estos datos, que fueron volcados posteriormente al paquete estadístico SPSS 18.0 (SPSS Inc., Chicago, IL) para Windows. De forma retrospectiva, se calculó el ESII<sup>8</sup> en la misma cohorte de sujetos recodificando las nuevas variables que este introducía. El cálculo de la puntuación para cada paciente fue llevado a cabo mediante la fórmula de regresión logística con base en los coeficientes beta ( $\beta$ ) del modelo ESII (<http://www.euroscore.org>):

$$\text{mortalidad predicha (\%)} = \frac{e^{(\beta_0 + \sum \beta_i X_i)}}{1 + e^{(\beta_0 + \sum \beta_i X_i)}}$$

Se calculó el índice de mortalidad ajustado por riesgo (IMAR) dividiendo la mortalidad observada por la predicha por ambas escalas, y se analizó la capacidad discriminatoria y la bondad de ajuste de ambas escalas logísticas en la estimación de la mortalidad postoperatoria (entendida como la hospitalaria o la acaecida en el período de 30 días siguientes a la cirugía).

Se obtuvo un consentimiento firmado del paciente previo a la intervención para en análisis de la puntuación ES. Posteriormente, el comité ético de nuestra institución proveyó del consentimiento para la estimación del ESII en la misma cohorte de pacientes y la realización de este estudio.

### Análisis estadístico

Las variables cuantitativas se expresaron mediante media y desviación estándar o intervalo de confianza 95%, salvo distribución no normal, en cuyo caso, se resumieron mediante mediana y rango intercuartil. Las variables cualitativas se expresaron en frecuencias absolutas y porcentajes. Las variables cuantitativas se compararon mediante prueba de t de Student o Wilcoxon (según la normalidad de la distribución). Las variables cualitativas se compararon mediante  $\chi^2$  o la prueba exacta de Fischer.

Se calculó, asimismo, la calibración de ambas escalas para los eventos mediante la prueba de bondad de ajuste de  $\chi^2$  de Hosmer-Lemeshow. Esta prueba compara la frecuencia de eventos (en este caso, mortalidad postoperatoria) esperada frente a la observada con una prueba de  $\chi^2$ , de suerte que valores de  $\chi^2$  altos con p bajas indican discrepancias relevantes entre la frecuencia observada y esperada del evento. La discriminación mide la capacidad de un modelo (en este caso, ES y ESII) de diferenciar entre los individuos de una muestra que sufren un evento (en este caso, muerte) y los que no. La capacidad discriminatoria de las escalas ES y ESII para los eventos estudiados se estimó mediante curvas ROC (*receiver operating characteristics*)<sup>10</sup>. Se calcularon sus áreas bajo la curva (ABC), y se compararon con el estadístico z según una ley normal (N(0,1)) con el método propuesto por Hanley y McNeil<sup>11</sup>.

$$z = \frac{ABC_1 - ABC_2}{\sqrt{(SE(ABC_1)^2 + SE(ABC_2)^2)}}$$

Para el resto del análisis, se empleó el paquete estadístico SPSS 18.0 (SPSS Inc., Chicago, IL) para Windows. Se consideró significativo un valor  $p < 0,05$  (bilateral).

Se analizaron la distribución cruda de las variables y los resultados de forma comparativa entre nuestra base de datos y los resultados publicados en el ESII.

## Resultados

Entre enero de 2005 y enero de 2012 fueron incluidas en el estudio un total de 4.166 cirugías cardíacas mayores: 1.202 revascularizaciones miocárdicas (28,9%), 1.678 cirugías valvulares aisladas (40,3%), 293 procedimientos valvulares y coronarios combinados (7%), 411 cirugías de la aorta torácica (9,9%) y 582 cirugías cardíacas mayores no correspondientes a ninguno de los anteriores grupos (14%).

**Tabla 1**  
Distribución de las distintas variables de EuroSCORE y EuroSCORE II

EuroSCORE II	EuroSCORE
NYHA	
II 1.254 (30,1%)	
III 1.379 (33,1%)	
IV 233 (5,6%)	
CCS4 198 (4,8%)	
DMID 379 (9,1%)	
Edad 67 (SD 12,2)	
Sexo femenino 1.571 (37,7%)	
Arteriopatía periférica 613 (14,7%)	
EPOC 430 (10,3%)	
AFS 262 (6,3%)	
Reoperación 734 (17,6%)	
Función renal	
Diálisis 133 (3,2%)	
CC (ml/min) ≤ 50 959 (23%)	
CC (ml/min) > 50-85 1.957 (47%)	
Endocarditis activa 128 (3,1%)	
Estado crítico 289 (6,9%)	
FEV	
> 50% 2.980 (71,5%)	
31-50% 1.012 (24,3%)	
21-30% 136 (3,3%)	
≤ 20% 38 (0,9%)	
IAM reciente 594 (14,3%)	
PSAP	
31-55 mmHg 710 (17%)	
≥ 55 mmHg 461 (11,1%)	
Urgencia	
Urgencia 257 (6,2%)	
Emergente 223 (5,4%)	
Savage 7 (0,2%)	
Aorta torácica 411 (9,9%)	
Peso del procedimiento	
1 No CABG 1.583 (38%)	
2 1.007 (24,17%)	
3+ 172 (4,13%)	
Datos expresados en n (%) y media (SD). La definición de cada uno de los ítems se acoge a las previamente publicadas para ES (1) y ESII (2).	
Cr: creatina; DMID: diabetes mellitus insulíndependiente; EPOC: enfermedad pulmonar obstructiva crónica; FEV1: fracción de eyeción del ventrículo izquierdo; IAM: infarto agudo de miocardio; NYHA: New York Heart Association.	

## Distribución de variables de las escalas EuroSCORE y EuroSCORE II

En la tabla 1 se resume la distribución de las distintas variables de ambas escalas. Nueve de las 17 variables del ES originario<sup>1</sup> permanecen en el ESII<sup>8</sup> (edad, sexo, arteriopatía periférica, enfermedad pulmonar obstructiva crónica, reintervención, endocarditis activa, estado crítico, infarto reciente y cirugía de la aorta torácica). Se observó un porcentaje de sexo femenino del 37,7%, con una edad media de 67 años (desviación estándar 10,2). Asimismo, se evidenció una importante prevalencia de comorbilidades graves como arteriopatía periférica (14,7%), enfermedad pulmonar obstructiva crónica (10,3%), reoperación (17,6%) o estado crítico (6,9%). De acuerdo con la nueva codificación introducida en el ESII se han modificado otras variables como la función renal (calculada mediante la fórmula de Cockcroft-Gault), la fracción de eyeción del ventrículo izquierdo, la hipertensión pulmonar, la prioridad quirúrgica o el peso del procedimiento. Con la nueva codificación, se observó una prevalencia de pacientes con nefropatía moderada o grave ( $FG < 85 \text{ ml/min}$ ) del 73,2% en la escala ESII frente a creatina  $< 2,26 \text{ mg/dL}$  en el ES (3,4%). En la nueva escala ha desaparecido la codificación de la rotura septal postinfarto (cuya incidencia en nuestra muestra fue 0,16%). Finalmente, se han introducido nuevas variables: diabetes mellitus insulíndependiente (prevalencia de 9,1%) y New York Heart Association (con una alta prevalencia de clases funcionales III y IV: 33,1 y 5,6%, respectivamente).

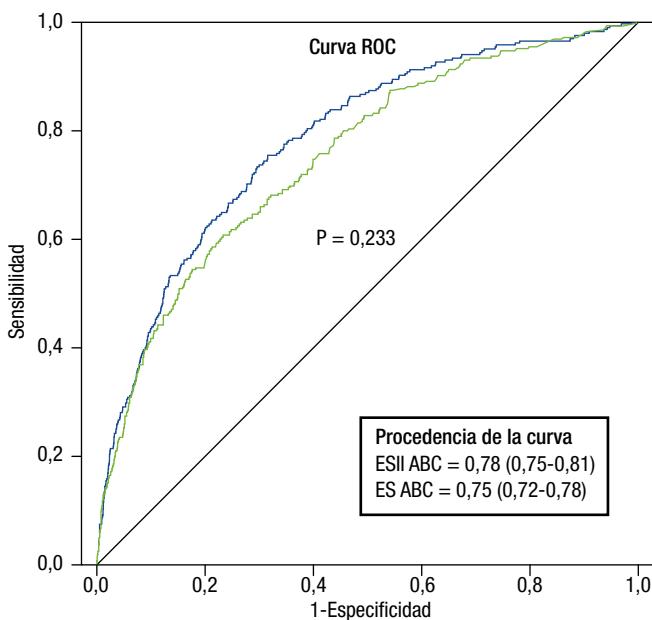
## Mortalidad observada, estimada y ajustada

La mortalidad global, estimada y ajustada para toda la cohorte de pacientes y subgrupos de patologías se resume en la tabla 2. La mortalidad observada en esta muestra de pacientes fue 6,3%. La mortalidad media predicha por la escala ES y ESII fue 9,1% (intervalo de confianza [IC] 95% 8,8-9,4) y 3,46% (IC 95% 3,3-3,6), respectivamente ( $p < 0,001$ ). La mortalidad ajustada: (observada/predicha) fue 0,69 y 1,82 para las escalas ES y ESII, respectivamente. Se observó una disminución estadísticamente significativa de la mortalidad esperada por la escala ESII respecto de la escala ES para todos los subgrupos de patología (tabla 2): 2,8 vs. 5,86% ( $p < 0,0001$ ) en el grupo de cirugía coronaria; 3,2 vs. 9,12% ( $p < 0,0001$ ) en el subgrupo de cirugía valvular; 15,9 vs. 5,2% ( $p < 0,001$ ) en patología de aorta torácica; 5,1 vs. 9,7% ( $p < 0,0001$ ) en cirugía mixta; y 3,6 vs. 10,45% ( $p < 0,001$ ) en otras intervenciones cardíacas mayores. Esta disminución en la mortalidad estimada se tradujo en un incremento de la mortalidad ajustada (IMAR) para el ESII; de suerte que en el subgrupo de patología coronaria pasó de ser en el ES 0,72 a 1,5 en el ESII; en los subgrupos de cirugía valvular aislada de 0,76 a 2,16, aorta 0,57 a 1,75, y otros tipos de cirugía cardíaca mayor pasó de 0,47 a 1,36. En el subgrupo de cirugía combinada mixta, se produjo un incremento de la mortalidad ajustada de 1,03 a 1,96.

**Tabla 2**  
Mortalidad observada, estimada y ajustada para la cohorte global y subgrupos de patologías quirúrgicas

	n (%)	Mortalidad observada (%)	ES medio (%) (IC 95%)	ESII medio (IC 95%)	p	IMAR (ES)	IMAR (ESII)
Global	4.166	6,3	9,1 (8,8-9,4)	3,46 (3,3-3,6)	< 0,001	0,69	1,82
Coronario	1.202 (28,9)	4,2	5,86 (5,4-6,3)	2,8 (2,6-3)	< 0,0001	0,72	1,5
Valvular	1.678 (40,3)	6,9	9,12 (8,6-9,6)	3,2 (3,02-3,4)	< 0,0001	0,76	2,16
Mixto	293 (7)	10	9,7 (8,6-10,9)	5,1 (4,46-5,7)	< 0,001	1,03	1,96
Aorta	411 (9,9)	9,1	15,9 (14,7-17,2)	5,2 (4,7-5,76)	< 0,0001	0,57	1,75
Otras	582 (14)	4,9	10,45 (9,4-11,5)	3,6 (3,2-4)	< 0,001	0,47	1,36

Mortalidad observada, predicha por escalas EuroSCORE y EuroSCORE II y ajustada (cociente entre mortalidad observada y predicha por cada escala). Comparación de proporciones. ES: EuroSCORE; ESII: EuroSCORE II; IC 95%: intervalo de confianza del 95%; IMAR: índice de mortalidad ajustada por el riesgo; p: comparación de proporciones. Se considera significativa  $p < 0,05$ .



**Figura 1.** Curvas ROC para la muestra global.  
p es la probabilidad para  $z \geq z_i$  (2). Se considera significativa  $p < 0,05$ . ES: EuroSCORE.  
ESII: EuroSCORE II.

**Tabla 3**  
Capacidad discriminatoria (curvas ROC) de EuroSCORE y EuroSCORE II

	Discriminación (curvas ROC)			
	ABC (IC 95%)		Test Hanley y McNeil	
	ES	ESII	z	p
Global	0,75 (0,72-0,78)	0,78 (0,75-0,81)	0,732	0,233
Coronario	0,79 (0,71-0,86)	0,80 (0,74-0,86)	0,403	0,345
Valvular	0,73 (0,68-0,77)	0,76 (0,71-0,8)	1,019	0,154
Mixto	0,78 (0,70-0,86)	0,80 (0,73-0,87)	0,133	0,448
Aorta	0,70 (0,62-0,79)	0,72 (0,63-0,81)	0,464	0,326
Otras	0,73 (0,61-0,84)	0,85 (0,75-0,94)	0,331	0,371

ABC: área bajo la curva; ES: EuroSCORE; ESII: EuroSCORE II; ROC: Receiver Operating Characteristics.  
Se compararon las ABC con el test de Hanley y McNeil. Se muestran los valores z y el valor de p correspondiente a  $P(z \geq z_i)$ . Se consideró significativa  $p < 0,05$ .

#### Análisis de la calibración y la capacidad discriminatoria

Las 2 escalas mostraron una buena capacidad de discriminación (ABC por encima de 0,7) en la muestra global de pacientes, con ABC de 0,75 (IC 95% 0,72-0,78) para el ES y 0,78 (IC 95% 0,75-0,81) para el ESII, sin que existiesen diferencias estadísticamente significativas ( $z = 0,732$ ,  $p = 0,233$ ) (fig. 1, tabla 3). La tabla 3 muestra la comparación de las ABC de las curvas ROC para los subgrupos de patologías. No se observó ninguna diferencia estadísticamente significativa en la capacidad de discriminación, con curvas ROC y ABC muy similares para todos los subgrupos de pacientes ( $p > 0,05$  en test de Hanley y McNeil). En la tabla 3 se observó que todos los límites inferiores de los IC 95% de las ABC de las curvas ROC de las 2 escalas para los distintos grupos de patología y para la muestra global superaban 0,7, lo cual supone una buena capacidad discriminativa. Se observó una mayor capacidad discriminatoria de la escala ESII para todos los grupos quirúrgicos y la muestra global, sin llegar a ser estas diferencias estadísticamente significativas.

Sin embargo, la bondad de ajuste (calibración) fue mala para ambas escalas en el grupo global, siendo la correspondiente al ES ( $\chi^2 = 26,6$ ,  $p = < 0,001$ ) y ESII ( $\chi^2 = 58,19$ ,  $p < 0,0001$ ).

**Tabla 4**  
Distribución de mortalidad y variables entre la serie EuroSCORE II y nuestro grupo

	ESII	HCSC
Puntuación media ESII	3,9%	3,46%
Puntuación media ES logístico	7,57%	9,1%
Mortalidad hospitalaria	4,63%	6,3%
Mortalidad 90 días	6,023%	No disponible
Edad media (años)	64,6	67
Sexo (% femenino)	30,9%	37,7%
Cirugía coronaria (%)	46,7%	28,9%
Estado crítico	4,1%	6,9%
Diálisis	1,1%	3,2%

ES: EuroSCORE clásico; ESII: EuroSCORE II; HCSC: Hospital Clínico San Carlos.

#### Comparación cruda de nuestros datos con la base de datos EuroSCORE II

La tabla 4 muestra los resultados comparativos de la mortalidad observada, predicha por el ES y ESII, y la distribución de ciertas variables en nuestra base de datos y la del ESII. Se observa una mortalidad predicha por el ES logístico en nuestra serie del 9,1% frente a un 7,57% en la base de datos del ESII. Sin embargo, el ESII medio en nuestra serie fue 3,46% frente a 3,9% en la base de datos del ESII. Encuentramos en nuestra serie mayor edad media (67 años vs. 64,6), mayor prevalencia de sexo femenino (37,7 vs. 30,9%), diálisis (3,2 vs. 1,1%), estado crítico (6,9 vs. 4,1%) y una menor proporción de cirugía coronaria aislada (28,9 vs. 46,7%).

#### Discusión

La escala de riesgo ES ha sido una herramienta muy útil durante los últimos años como método de estimación de mortalidad y control de calidad de los resultados en cirugía cardíaca. Las características parsimoniosas del modelo, de fácil aplicación, permitieron su rápida divulgación entre la comunidad de cirujanos y cardiólogos. Sin embargo, numerosos estudios han mostrado una sobreestimación del riesgo en ciertos subgrupos de pacientes<sup>5-7</sup>. La escala ESII ha sido recientemente publicada. El desarrollo de esta nueva escala obedece a múltiples deficiencias observadas a lo largo de años de la aplicación de ES a distintas muestras de pacientes como: baja prevalencia de octogenarios (< 2%) o cirugía valvular (< 30%) en la cohorte en la que fue estimada, elevado número de pacientes coronarios, análisis del impacto de la función renal en la estimación de la mortalidad, disminución de la capacidad discriminatoria a medida que mejoraron los resultados de la cirugía cardíaca. En su desarrollo han participado 154 centros de 43 países. EuroSCORE II ha sido calculado en una subcohorte consecutiva de 16.828 pacientes y estimada su validez en otra subcohorte de 5.553 sujetos<sup>8</sup>. EuroSCORE II fue capaz de predecir la mortalidad hospitalaria después de una cirugía cardíaca mayor con una excelente capacidad discriminativa (ABC = 0,81; IC 95% 0,782-0,836)<sup>8</sup>. La escala originaria ES, aplicada en dicha cohorte de 5.553 sujetos<sup>8</sup>, también mostró buena capacidad discriminatoria con un ABC de 0,789. En el presente trabajo, con 4.166 pacientes, observamos que ambas escalas poseen buena capacidad de discriminación con ABC de 0,75 (IC 95% 0,72-0,78) para el ES y 0,78 (IC 95% 0,75-0,81) para el ESII, respectivamente, si bien no llegaron a alcanzar las descritas en el trabajo de Nashef et al.<sup>8</sup>.

Al aplicar en nuestra serie ambos modelos logísticos en los subgrupos de patologías quirúrgicas (tabla 3), la escala ESII mostró en todos los grupos una tendencia a una mejor capacidad discriminatoria, aunque las diferencias entre las ABC de cada escala para cada patología no alcanzaron diferencias estadísticamente significativas. La peor discriminación de ambos modelos se observó en los pacientes con cirugía aorta, con ABC de 0,70 y 0,72 para ES y ESII, respectivamente ( $p = 0,326$ ). En el grupo de revascularizaciones

miocárdicas con 1.202 pacientes (28,9%), las 2 escalas discriminaron de manera bastante precisa: ABC = 0,78 y ABC = 0,8 para ES y ESII, respectivamente ( $p = 0,345$ ). Estas ABC son comparables a las obtenidas al aplicar ES y ESII en la muestra de 5.553 pacientes del trabajo de Nashef et al.<sup>7</sup> citadas anteriormente. En el grupo de cirugía valvular, que representa el 40,3% de los pacientes de nuestra muestra, la capacidad de discriminación (ABC = 0,73 y ABC = 0,76 para ES y ESII) fue menor a la hallada en la muestra global, en la cirugía coronaria y en la descrita anteriormente por Nashef et al.<sup>8</sup>. Sin embargo, fue sensiblemente mayor a la descrita por Basraon et al.<sup>5</sup>, quienes hallaron un ABC de 0,62 para una muestra de 537 sujetos sometidos a una sustitución valvular aórtica aislada. El ABC de ES y ESII para la muestra global y cada uno de los subgrupos del presente trabajo fue similar al hallado en un trabajo reciente publicado por Siregar et al.<sup>4</sup>. En este estudio, que analizó la bondad de ajuste y discriminación de ES en más de 400.000 pacientes, las ABC halladas oscilaron entre 0,7 y 0,8. En definitiva, a tenor de los resultados de este estudio y del análisis de Nashef et al.<sup>8</sup>, la escala ES ya mostraba una excelente capacidad discriminativa que ha sido discretamente mejorada con la nueva versión ESII.

Múltiples estudios previos han detectado una sobreestimación de la mortalidad hospitalaria al aplicar ES a distintos subgrupos de pacientes<sup>4-6,12</sup>. Por ejemplo, en un metanálisis publicado por Parolari et al.<sup>12</sup>, en 26.621 pacientes sometidos a cirugía valvular, el IMAR osciló entre 0,45 y 0,89. En una revisión sistemática de Siregar et al.<sup>4</sup>, ES sobreestimaba la mortalidad en pacientes sometidos a cirugía coronaria, valvular y combinada, con IMAR entre 0,43 y 0,62. Cuando se aplicó la escala ES I logística a los 5.553 pacientes de la subcohorte de la validación de ESII<sup>8</sup>, se observó una mortalidad esperada de 7,57% frente a una observada de 3,9%. La mala calibración del ES original ha sido el principal motivo que ha impulsado el desarrollo de un nuevo modelo ES<sup>8</sup>. Sin embargo, al analizar la discriminación del nuevo modelo en la muestra de validación del ESII<sup>8</sup>, se observó que, a pesar de no existir una gran discrepancia entre la mortalidad observada (3,95%) y la estimada (4,18%) (cambio < 10%), el test de Hosmer-Lemeshow resultó casi significativo ( $\chi^2 = 15,48$ ,  $p = 0,051$ ), lo que denota diferencias (sin llegar a ser estadísticamente significativas) entre la mortalidad estimada y observada.

En nuestra muestra, ambas escalas demostraron una mala calibración global, con una mortalidad observada de 6,3% y esperadas/ajustadas (IMAR) de 9,1%/0,69 y 3,6%/1,82 para ES y ESII, respectivamente; y resultados significativos en el test de Hosmer-Lemeshow para ambas escalas (ES:  $\chi^2 = 26,667$ ,  $p < 0,001$ ; y ESII:  $\chi^2 = 58,914$ ,  $p < 0,0001$ ). Sin embargo, la utilización de la prueba de Hosmer-Lemeshow ha sido criticada por algunos autores y sus resultados deben ser interpretados con cautela, sobre todo en grandes series, donde un resultado positivo no implica necesariamente una mala calibración<sup>13</sup>.

La mala calibración en el grupo global podría ser explicada por las diferencias entre nuestra muestra de pacientes y la muestra de sujetos con la que se diseñó la escala ESII<sup>8</sup>. En un estudio publicado en 2000 por Roques et al., sobre la validación del ES en distintos países europeos, la serie española mostró la peor calibración y capacidad discriminatoria; y este hallazgo fue atribuido a las diferencias poblacionales encontradas entre los distintos países<sup>14</sup>. Así, encontramos en nuestra serie (tabla 4) mayor edad media (67 años vs. 64,6), mayor prevalencia de sexo femenino (37,7 vs. 30,9%), diálisis (3,2 vs. 1,1%), estado crítico (6,9 vs. 4,1%) y una menor proporción de cirugía coronaria aislada (28,9 vs. 46,7%). Además, en la muestra de pacientes de este estudio, se incluyeron sujetos intervenidos hace más de 5 años (desde 2004), donde ESII pueda perder validez externa (ESII se estimó a partir de pacientes intervenidos entre mayo y julio de 2010)<sup>8</sup>. Llama la atención que la mortalidad predicha por el ES logístico en nuestra serie fue de 9,1% frente a un 7,57% en la base de datos del ESII. Sin embargo, el ESII medio en nuestra serie fue 3,46% frente a 3,9% en la base de datos ESII. Estos datos apuntan que determina-

das variables como la edad y el sexo han perdido peso a la hora de estimar la mortalidad en el ESII.

Es claro que toda escala de riesgo es potencialmente mejorable, sobre todo en lo referente a la calibración, es decir, cómo se ajustan la mortalidad observada y la estimada. En el estudio del ESII, a pesar de no existir diferencias significativas entre la mortalidad estimada (3,95%) y observada (4,18%) en la cohorte de validación de ESII<sup>8</sup>, el resultado de la prueba de Hosmer-Lemeshow resultó casi significativo ( $p = 0,0505$ ).

Los problemas en la calibración del nuevo modelo observados tanto en nuestra muestra de pacientes como en el propio diseño de la escala<sup>8</sup> pueden deberse a varios motivos: en primer lugar, no fueron incluidas en el ESII variables de interacción o modificadoras del efecto, que sin duda hubieran mejorado la precisión del modelo<sup>15</sup>. En segundo lugar, la codificación de ciertas variables: es llamativo el mayor peso en la mortalidad del filtrado glomerular estimado por debajo de 50 ml/min ( $\beta = 0,8592256$ ) frente a la diálisis ( $\beta = 0,6421508$ ), lo que puede explicarse por la baja prevalencia de pacientes en diálisis (1,1%). Además, la función renal se ha estimado por el filtrado glomerular calculado por la fórmula Cockcroft-Gault, lo que sin duda alguna supone una aportación importante con respecto de la escala previa<sup>16</sup>. Sin embargo, se ha demostrado que existen mejores estimadores de la función renal como el calculado con la fórmula MDRD<sup>17</sup>. Por último, esta variable en realidad es una variable de interacción, ya que incluye la edad, la creatinina y el sexo, y en el modelo debiera haber sido tratada como tal, lo cual hubiera supuesto respetar el principio jerárquico.

En tercer lugar, no existió ningún control de calidad mediante una auditoría externa en la recogida en los datos. Un estudio reciente<sup>18</sup> demostró la existencia de discrepancias interobservador en un 26,3% de 1.719 pacientes. En dicho estudio se demostró que la mayoría de la variabilidad se debía a la puntuación de 5 de las 17 variables. La escala ESII incluye 1 variable y 9 categorías más. Esto podría disminuir la concordancia. Además, muchos de estos centros que participaron en su elaboración pudieron derivar ciertos pacientes a otros hospitales secundarios donde finalmente fallecieron, alterando de esta forma la mortalidad real.

La nueva escala ESII supone una actualización necesaria del modelo de predicción del riesgo quirúrgico más extensamente usado en el mundo: el ES. Las modificaciones incluidas han incrementado su discriminación con respecto a la escala previa para valorar el riesgo de los pacientes sometidos a una cirugía cardíaca mayor hoy en día. Los resultados de este estudio demuestran que, a pesar de poseer muy buena capacidad de discriminación, la bondad de ajuste de ESII es peor que la de su antecesor. Es necesario aplicar este nuevo modelo en muestras de pacientes más grandes y en múltiples centros para investigar con más precisión su validez externa. Como posible mejora se podría realizar un registro prospectivo que se vaya actualizando de forma continua, como se realiza en la base de datos de la STS, lo que conseguiría sin duda una mejor capacidad discriminatoria. Por último, la aplicación de programas de inteligencia artificial basados en modelos de simulación podrían mejorar la exactitud de esta escala de riesgo<sup>19</sup>.

## Conclusiones

El nuevo sistema de estratificación de riesgo ESII obtiene una mortalidad esperada inferior al ES clásico y tiene un valor predictivo adecuado en términos de discriminación; sin embargo, la calibración observada es pobre. Se observa una tendencia a una mejor capacidad discriminatoria respecto al ES clásico, aunque dicho hallazgo no muestra significación estadística. Se observa en nuestro medio una distribución de variables sensiblemente diferente frente al registro de ESII, aunque no se pudieron aplicar pruebas estadísticas para compararlas.

## Bibliografía

1. Nashef SAM, Roques F, Michel P, Gauduchea E, Lemeshow S, Salamon R; the EuroSCORE Study Group. European System for Cardiac Operative Risk Evaluation (EuroSCORE). Eur J Cardiothorac Surg. 1999;16:9-13.
2. Roques F, Nashef SAM, Michel P, Gauduchea E, de Vincenti C, Baudet E, et al. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. Eur J Cardiothorac Surg. 1999;15:816-23.
3. Roques F, Michel P, Goldstone A, Nashef SAM. The logistic EuroSCORE. Eur Heart J. 2003;24:881-2.
4. Siregar S, Groenewold RHH, de Heer F, Bots ML, van der Graaf Y, van Herwerden LA. Performance of the original EuroScore. Eur J Cardiothorac Surg. 2012;41:746-54.
5. Basraon J, Chandrashekhar YS, John R, Agnihotri A, Kelly R, Ward H, et al. Comparison of risk scores to estimate perioperative mortality in aortic valve replacement surgery. Ann Thorac Surg. 2011;92:535-40.
6. Parolari A, Pesce LL, Trezzi M, Loardi C, Kasseem S, Brambillasca C, et al. Performance of EuroSCORE in CABG and off-pump coronary artery bypass grafting: single institution experience and meta-analysis. Eur Heart J. 2009;30:297-304.
7. Yap CH, Reid C, Yii M, Rowland MA, Mohajeri M, Skillington PD, et al. Validation of the EuroSCORE model in Australia. Eur J Cardiothorac Surg. 2006;29:441-6.
8. Nashef SAM, Roques F, Sharples L, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. Eur J Cardiothorac Surg. 2012;41:1-12.
9. Carnero-Alcazar M, Silva JA, Reguillo FJ, Maroto LC, Cobiella J, Villagrán E, et al. Validation of EuroSCORE II on a single-centre 3800 patient cohort. ICVTS. 2012;16:293-300.
10. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. BMJ. 2009;338:b604.
11. 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.
12. Parolari A, Pesci LL, Trezzi M, Cavallotti L, Kasseem S, Loardi C, et al. EuroSCORE performance in valve surgery: a meta-analysis. Ann Thorac Surg. 2010;89:787-93.
13. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Critical Care Medicine. 2007;35:2052-6.
14. Roques F, Nashef SAM, Michel P, Pinna Pintor P, David M, Baudet E, and The EuroSCORE Study Group Does EuroSCORE work in individual European countries? Eur J Cardiothorac Surg. 2000;18:27-30.
15. Miettinen OS. Confounding and effect-modification. Am J Epidemiol. 1974;100: 350-3.
16. Van Gameren M, Klieverik LM, Struijs A, Venema AC, Kappetein AP, Bogers AJ, et al. Impact of the definition of renal dysfunction on EuroSCORE performance. J Cardiovasc Surg. 2009;50:703-9.
17. Silva J, Ridao-Cano N, Segura A, Maroto LC, Cobiella J, Carnero M, et al. Can estimated glomerular filtration rate improve the EuroSCORE? Interact Cardiovasc Thorac Surg. 2008;7:1054-7.
18. Lebreton G, Merle S, Inamo J, Hennequin JL, Sanchez B, Riols Z, et al. Limitation in the inter-observer reliability of EuroSCORE: what should change in EuroSCORE II? Eur J Cardiothorac Surg. 2011;40:1304-8.
19. Nilsson J, Ohlsson M, Thulin L, Hoglund P, Nashef S, Brandt J. Risk factor identification and mortality prediction in cardiac surgery using artificial neural networks. J Thorac Cardiovasc Surg. 2006;132:12-9.



# BIO MED



unidix

# Especialistas en cirugía cardiovascular

desde 1977 al cuidado de tu salud



91 803 28 02



info@biomed.es

