



CIMAT

Centro de Investigación en Matemáticas, A.C.

VALIDACIÓN DEL MODELO EUROSCORE II EN EL HOSPITAL REGIONAL DE ALTA ESPECIALIDAD DEL BAJIO

T E S I N A

Que para obtener el grado de
Especialidad en Métodos Estadísticos

Presenta

Raúl Teniente Valente

Director

Dr. Humberto Martínez Bautista

Autorización de la versión final

“Lo que sabemos es una gota de agua; lo que ignoramos es un océano”

“Si piensas que la aventura es peligrosa, prueba la rutina. Es mortal”

Paulo Coelho

“En medio de la dificultad reside la oportunidad”

Albert Einstein

“La mente que se abre a una nueva idea jamás volverá a su tamaño original”

Albert Einstein

Dedicatoria y agradecimientos

“El que da, no debe volver a acordarse; pero el que recibe nunca debe olvidar”.

A mis queridos Padres (QEPD) quienes con su amor y ejemplo me enseñaron los valores que han guiado mi camino.

A mi esposa Ma del Pilar, por su amor, apoyo y comprensión.

A mis hijos, Gabriela, Alejandro y Omar por su apoyo y solidaridad.

A mis nietos: Renata, Alejandro, Ares Gabriel, Regina, Diego, Ricardo y Beatriz Elena por ser fuente inagotable de energía para seguir mi camino.

A mis compañeros del servicio de Cardiología y Cirugía de corazón, especialmente a los doctores Miguel Ángel Chagoya Santillán, Iliana Acevedo Bañuelos, Iván García Muñoz y Ricardo Romo Escamilla, por su ayuda en la captura de datos.

“El talento gana partidos, pero el trabajo en equipo y la inteligencia gana campeonatos”.

Michel Jordán

A las autoridades del HRAEB, por el esfuerzo y visión de convertir a nuestro hospital en una institución de salud de excelencia. En especial a la directora Médica, Dra. Esperanza García Moreno por su apoyo.

“Los grandes líderes encuentran formas de conectarse con su gente y ayudarla a alcanzar todo su potencial”.

Steven J. Stowell

Al Dr. Edgard Efrén Lozada Hernández, por su apoyo.

A los profesores del Centro de Investigaciones en Matemáticas, A.C. (CIMAT) Unidad Aguascalientes, maestros del curso de la “Especialidad en Métodos Estadísticos” por sus enseñanzas.

“Enseñar no es transferir conocimiento, es crear la posibilidad de producirlo”

Paulo Freire

Al Maestro Ignacio Méndez Gómez Humaran y al Doctor Rafael Pérez Abreu Carrión por sus enseñanzas y la revisión del presente documento.

Al Doctor Humberto Martínez Bautista por sus enseñanzas, la dirección en la elaboración y corrección del presente documento.

“La observación científica ha establecido que la educación no es lo que un profesor da, es un proceso natural llevado a cabo espontáneamente por el individuo y no se adquiere escuchando palabras, sino a través de experiencias en el entorno”.

María Montessori

ÍNDICE

I. Resumen	5
II. Antecedentes	6
II.1 Modelos predictivos en cirugía cardíaca	9
II.2 EuroSCORE I	11
II.3 EuroSCORE II	14
III. Justificación	17
IV. Planteamiento del problema	18
V. Hipótesis	18
VI. Objetivos	19
VII. Material y métodos	19
VIII. Análisis estadístico	24
IX. Consideraciones éticas	25
X. Resultados	25
XI. Discusión	45
XII. Conclusión	49
XIII. Bibliografía	50

I. Resumen

Antecedentes:

El modelo pronóstico EuroSCORE para mortalidad perioperatoria en cirugía cardíaca publicado en 1999 ha sido utilizado ampliamente en diversas latitudes. En los últimos lustros mostró deficiencias en su calibración, aunque siguió conservando buena discriminación. El modelo fue actualizado y publicado en el 2012 y fue llamado EuroSCORE II.

Objetivo:

Evaluar el desempeño del EuroSCORE II en pacientes sometidos a cirugía cardíaca con y sin bomba de circulación extracorpórea en el HRAEB de León, Gto.

Material y métodos:

Estudio observacional, transversal, retrospectivo. En pacientes ≥ 16 años operados del 01/01/2008 al 31/12/2013. Se registró la mortalidad hospitalaria y se calculó el puntaje EuroSCORE II mediante la calculadora en línea disponible en www.euroscore.org. Se valoró la discriminación con el área bajo la curva ROC. La calibración fue evaluada mediante la prueba Ch2 de bondad de ajuste de Hosmer-Lemeshow (H-L) y la razón de mortalidad ajustada al riesgo (RAMR).

Resultados:

Un total de 338 pacientes, edad 49.87 ± 16.61 años (16-80), 176 hombres (52.1%) y 162 mujeres (47.9%). Tipos de cirugía: valvular 108 (31.95%), revascularización aorto-coronaria 101 (29.88%), congénitos 51 (15.08%) y otros 78 (23.07%). El EuroSCORE II fue de 4.1 ± 5.3 . La mortalidad observada fue de 10.9%. El área bajo la curva ROC fue de 0.806 (IC 95% 0.739-0.872) compatible con buena discriminación. La Ch2 de H-L fue 13.2 con $p = 0.104$, sugestiva de calibración adecuada. La RAMR fue de 2.65, compatible con infraestimación global del modelo, principalmente en los deciles de alto riesgo.

Conclusión:

EuroSCORE II es confiable para estimar las probabilidades de muerte en pacientes en el HRAEB. La discriminación es buena. La calibración global adecuada pero el RAMR sugiere que el modelo infraestima el riesgo en pacientes ubicados en los deciles de alto riesgo.

II. Antecedentes

Las condiciones generales, tipo de cardiopatía y su gravedad, así como el procedimiento quirúrgico requerido son diferentes en cada paciente sometido a cirugía cardíaca mayor. Por lo tanto, el riesgo de mortalidad y morbilidad difiere también. En base a este concepto, en los años setentas y ochentas del siglo pasado inició el desarrollo de sistemas de puntajes o escalas de riesgo encaminados a dotar a cardiólogos y cirujanos de corazón de herramientas útiles para pronosticar el riesgo que conlleva un paciente en particular al ser sometido a una cirugía cardíaca mayor como revascularización aorto-coronaria, plastia y/o sustitución valvular, corrección de alguna malformación congénita, entre otros tipos de procedimientos. La mortalidad perioperatoria es el desenlace de interés más frecuentemente usado en estas escalas de riesgo, aunque algunas escalas consideran tanto la mortalidad como la morbilidad. Esta última generalmente comprende las complicaciones más frecuentemente observadas en este tipo de cirugías: insuficiencia renal aguda, intubación prolongada, infección profunda de la herida quirúrgica, entre otras. La construcción de las escalas de riesgo conlleva la identificación de las condiciones generales, tipo de cardiopatía y gravedad así como el procedimiento o procedimientos quirúrgicos asociados con el desenlace de interés. A estas condiciones se les llama factores de riesgo o variables predictivas y al evento de desenlace o de interés variable de resultado o de respuesta. En estadística, a las primeras se les conoce como variables independientes o variables x , a la segunda como variable dependiente o variable de respuesta y [1,2].

La variable dependiente (y) y la(s) variable(s) independientes (x) pueden ser numéricas o categóricas. Puede existir una relación causal o no entre variables. “*La asociación estadística mide la similitud estructural entre dos variables, identificando cierta correspondencia entre valores de una y otra*” [3]. Los métodos más conocidos y utilizados para establecer asociación estadística entre dos variables numéricas son la gráfica de dispersión, el coeficiente de correlación de Pearson, el coeficiente de correlación de Spearman y la regresión lineal. Para establecer asociación estadística entre dos variables categóricas los métodos más conocidos y utilizados son la Ji-cuadrada (χ^2), tasas relativas, razón de momios (*Odds Ratio*) y el análisis de correspondencia [3].

Las escalas de riesgo o también llamados *scores* pronósticos más usados en cirugía de corazón (EuroSCORE, EuroSCORE II y STS) utilizan como variable dependiente la mortalidad, que es una variable categórica nominal dicotómica (muerto/vivo) y una serie de variables independientes que pueden ser numéricas y/o categóricas. En algunos modelos, algunas variables numéricas son categorizadas, como en el EuroSCORE y el EuroSCORE II, por ejemplo la fracción de expulsión del ventrículo izquierdo, la presión sistólica de la arteria pulmonar, la depuración de creatinina, entre otras. Para la construcción de modelos o modelación de escalas pronosticas existen métodos para explorar la asociación estadística. Los más conocidos son los modelos lineales generalizados, los modelos jerárquicos, los modelos longitudinales y los modelos de ecuaciones estructurales, entre otros [3]. La selección de cualquiera de esos modelos está en función del tipo de variables utilizadas. En la construcción y modelación de las escalas EuroSCORE y EuroSCORE II las herramientas utilizadas fueron la Regresión Logística Univariada (RLU) para identificar las variables con valor pronóstico y la Regresión Logística Múltiple para modelar la escala.

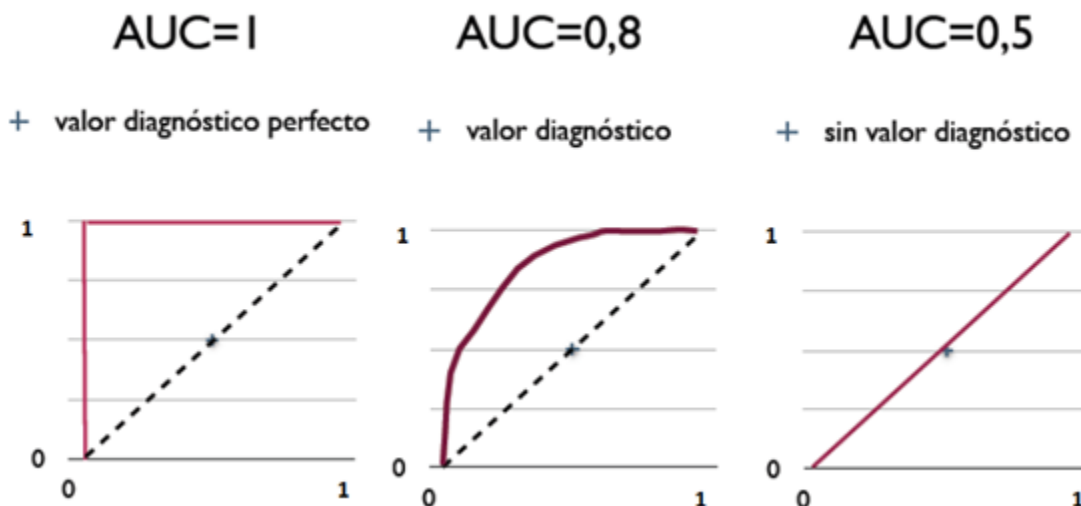
Una vez construido un modelo pronóstico, este debe ser sometido a validación interna, es decir, debe valorarse su rendimiento en la misma población donde fue desarrollado. La validación considera principalmente dos aspectos, la calibración y la discriminación. Si cumple bien estos dos aspectos, luego puede ser sometido a validación externa, es decir, evaluar su desempeño en poblaciones diferentes de la población donde fue desarrollado.

La calibración se define como el grado de acuerdo o concordancia entre los resultados esperados y observados; es decir, para modelos de predicción de mortalidad perioperatoria, valora si la mortalidad media pronosticada por el modelo se apega a la mortalidad que observamos en la población de estudio [4]. Es la comparación de los episodios esperados con los episodios observados a través de todo el rango de riesgo [4]. La prueba estadística utilizada generalmente para hacer dicha comparación recibe el nombre de prueba de bondad de ajuste de Hosmer-Lemeshow, ya que estos bioestadísticos (David W. Hosmer y Stanley Lemeshow) fueron quienes la desarrollaron junto a una serie de pruebas estadísticas para medir la bondad de ajuste en modelos de regresión logística. De las pruebas desarrolladas por estos autores, el estadístico C y el estadístico H son los más usados. La diferencia entre ellos es la manera de agrupar las probabilidades estimadas. El más conocido es el estadístico C ya que está integrado en la mayoría de los *softwares estadísticos*. La base del estadístico

C es la agrupación de las probabilidades estimadas bajo el modelo de regresión. El grupo de valores se divide en 10 intervalos llamados deciles de riesgo, iniciando desde el decil cuyas probabilidades son las más pequeñas, con aumento progresivo del riesgo en los deciles superiores [5,6].

Como se señaló en el párrafo anterior, la prueba calcula un estadístico C a partir de la diferencia entre los valores de mortalidad observados y los valores de mortalidad esperados por el modelo en distintos segmentos del grupo en estudio. Cuanto menor sea el valor del estadístico C, mejor será la calibración; la mortalidad predicha y la observada se acercan [4,5,6]. Un valor de $p > 0.05$ (que indica no diferencias estadísticas) sugiere que el modelo ajusta bien los datos (los valores no difieren significativamente entre si).

Por otra parte, la discriminación consiste en la valoración de la sensibilidad y la especificidad de la escala. Es una medida de la capacidad de un modelo de diferenciar a los individuos que sufrirán el evento de interés de los individuos que no sufrirán dicho evento. Se realiza mediante la valoración del área bajo la curva (ABC) (AUC en inglés) del *receiver operating characteristic* o curva ROC. La curva ROC representa gráficamente la relación entre la sensibilidad (la tasa de verdaderos positivos) y 1 menos la especificidad (la tasa de falsos positivos). Se calculan pares de valores de sensibilidad y especificidad para todos los posibles valores de corte para las probabilidades predichas del riesgo [4,7].



En la Figura anterior, valores de 1 indican discriminación perfecta. Por lo tanto, cuanto más se acerquen los valores a 1, mejor será su capacidad discriminatoria, como en el caso del lado izquierdo del gráfico de arriba. Se consideran adecuados los valores superiores a 0.7, y buenos los valores de 0.8 como los que muestra el centro del gráfico. En cambio, valores de 0.5 hablan de que el modelo no discrimina mejor que el azar, como se observa en la imagen derecha del gráfico [7].

II.1 Modelos predictivos en cirugía cardíaca

En 1980, Kennedy en Estados Unidos de América (EUA) publicó el desarrollo de uno de los primeros modelos predictivos conocidos en cirugía cardíaca. Este modelo fue desarrollado para ser aplicado solo en pacientes sometidos a cirugía de revascularización aorto-coronaria. Kennedy identificó 21 variables o factores de riesgo relacionados con la muerte [2].

En 1989 en Newark, EUA, Parsonnet desarrolló la escala que lleva su nombre y que en su momento fue la más difundida y utilizada a nivel mundial. El Score de Parsonnet fue actualizado en 1993, 1995, y 1997. La versión aditiva de 1995 Identificó 13 variables predictivas o factores de riesgo relacionadas con la mortalidad. A diferencia de la escala desarrollada por Kennedy, el score de Parsonnet podía utilizarse en cualquier procedimiento quirúrgico cardíaco mayor [2]. En la siguiente tabla aparecen las variables y el puntaje asignado a cada una de ellas en el score de Parsonnet, versión aditiva de 1995:

Tabla 1. Variables, escala y el puntaje de riesgo de Parsonnet (versión aditiva 1995)

Factor de riesgo		Puntuación
Mujer		1
Edad	70-74	2
	75-79	3
	>79	1
Fracción de expulsión	30-49%	1
	< 30%	3
Obesidad mórbida	Peso mayor 1.5 veces el ideal	1
Diabetes	Diabético o historia de diabetes	1
Hipertensión	Mayor de 140/90 o historia de hipertensión o medicación antihipertensiva	1
Reoperación	Primera re-operación	4
	Segunda o más re-operaciones	8
Balón preoperatorio	Presente en el momento de la cirugía	3
Aneurisma del ventrículo izquierdo	Resecada en la cirugía	3
Cirugía sobre la válvula mitral		4
Cirugía sobre la válvula aórtica		5
Cirugía sobre la válvula tricúspide		3
Cirugía valvular + coronaria		3

Fuente: Escala de riesgo de Parsonnet (versión 1995) [2].

En 1989 la Sociedad Americana de Cirugía Torácica (STS) inició la construcción de un modelo predictivo. Inicialmente la escala consideraba 31 factores de riesgo o variables predictivas para muerte peri-operatoria en pacientes sometidos a cirugía de revascularización coronaria. Desde el año 2011 el cálculo se obtiene en línea, a través de su página de internet. Actualmente la base de datos de la STS recibe los registros de los pacientes introducidos diariamente por más de 250 centros de cirugía cardíaca de los Estados Unidos y Canadá lo que representaría alrededor del 94% de todas las cirugías cardíacas mayores efectuadas en dichos países de Norte América. Se sabe que en la actualidad hay más de 5.000.000 de registros. El score STS es actualizado y ajustado periódicamente. Si bien en su inicio la escala de riesgo del STS fue usada sólo para calcular el riesgo quirúrgico en pacientes sometidos a cirugía de revascularización coronaria, actualmente el score puede utilizarse para predecir

el riesgo de mortalidad en todos los tipos de cirugía cardíaca mayor; así como también para predecir la morbilidad grave como la insuficiencia renal aguda, accidente cerebrovascular, ventilación mecánica prolongada, infección profunda de la herida esternal y re-operación. Con el pasar de los años, la escala ha ido afinándose y ha mostrado muy buen desempeño en todo tipo de cirugías cardíacas mayores. Sin embargo, su complejidad ha ido en aumento, por lo que su uso ha resultado menos amigable que otros modelos [4].

II.2 EuroSCORE I

En 1999 fue publicado el análisis del perfil de riesgo de 19030 pacientes europeos sometidos a cirugía cardíaca general con bomba de circulación extracorpórea entre Septiembre y Noviembre de 1995 en 132 hospitales de ocho países europeos, a saber: Alemania, Francia, Reino Unido, Italia, España, Finlandia, Suecia y Suiza. La edad media del grupo de pacientes estudiados fue 62.5 ± 10.7 años (rango 17-94 años), el 28% fueron mujeres, el índice de masa corporal [IMC] fue de 26.3 ± 3.9 . La incidencia de factores de riesgo comunes fue la siguiente: hipertensión arterial 43.6%, diabetes mellitus 16.7%, arteriopatía extracardíaca 2.9%, insuficiencia renal crónica 3.5%, enfermedad pulmonar crónica 3.9%, cirugía cardíaca previa 7.3% y deterioro de la función ventricular izquierda 31.4%. La cirugía coronaria aislada representó el 63.6% de todos los procedimientos, y el 29.8% de los pacientes tuvieron operaciones valvulares. La mortalidad hospitalaria general fue del 4.8%. Para la construcción del modelo originalmente se consideraron 97 factores de riesgo potenciales de los cuales finalmente se identificaron solo 17 factores asociados estadísticamente a la mortalidad hospitalaria o a 30 días después de la cirugía [4,8]. De los 17 factores de riesgo identificados, nueve estaban relacionados con las *condiciones generales del paciente* (edad, género femenino, creatinina sérica mayor de $200 \mu\text{mol/L}$ [$> 2.26 \text{ mg/dl}$], arteriopatía extracardíaca, enfermedad pulmonar obstructiva crónica [EPOC], disfunción neurológica, cirugía cardíaca previa, endocarditis activa, estado preoperatorio crítico), cuatro con *factores cardíacos* (angina inestable, fracción de expulsión del ventrículo izquierdo [FEVI] disminuida, infarto de miocardio reciente [< 90 días] y presión sistólica de la arteria pulmonar $> 60 \text{ mm Hg}$) y otros cuatro factores de riesgo estaban *asociados al tipo de procedimiento quirúrgico* (cirugía distinta a revascularización aorta-coronaria, cirugía de urgencia, rotura del septum interventricular y cirugía de aorta torácica). Esta base de datos sirvió para que el grupo de trabajo desarrollara el modelo pronóstico llamado EuroSCORE [8,9].

La selección de variables para elaborar el modelo EuroSCORE fue realizada en base a la identificación de su asociación estadística con el fallecimiento de pacientes a través de análisis de regresión logística univariada así como a su objetividad, credibilidad, disponibilidad, resistencia a la falsificación, tener valor de $p < 0.2$ y estar presentes en al menos el 2% de los pacientes de la muestra. Luego el modelo fue elaborado mediante regresión logística múltiple. De la base de datos de 19030 pacientes se tomaron cerca de 15,000 pacientes (14799) los cuales fueron divididos al azar en dos subgrupos, uno formado por cerca del 90% de ellos ($n=13302$), destinado al desarrollo del modelo EuroSCORE y otro subgrupo de aproximadamente el 10% restante ($n=1497$) para validarlo [9]. Las variables seleccionadas, su definición, la *razón de momios* (OR), los puntos asignados a cada variable para la elaboración de una versión aditiva y los coeficientes β para calcular la versión logística aparecen en la tabla siguiente:

Tabla 2. Variables incluidas y pesos obtenidos en el modelo EuroSCORE

Variable	Definición	OR	Puntos	β
Factores del Paciente				
Edad	Por cada 5 años a partir de los 60	1.1	1	0.066635
Género	Femenino	1.4	1	0.330405
Creatinina sérica	Mayor de 200 $\mu\text{mol/L}$	1.9	2	0.652165
Arteriopatía extracardíaca	Claudicación de miembros inferiores Oclusión carotídea o estenosis >50% Cirugía vascular previa o prevista sobre la Aorta abdominal, carótidas o arterias periféricas	1.9	2	0.655892
Enfermedad pulmonar crónica	Requiere tratamiento prolongado con broncodilatadores o esteroides	1.6	1	0.493134
Disfunción neurológica	Daño neurológico que afecte severamente a la deambulación o a la actividad cotidiana	2.3	2	0.841626
Cirugía cardíaca previa	Cirugía cardíaca que haya requerido apertura del pericardio	2.6	3	1.002625
Endocarditis activa	El paciente está con tratamiento antibiótico en el momento de la cirugía	2.5	3	1.101265
Estado preoperatorio crítico	Una o más de las siguientes condiciones: taquicardia/fibrilación ventricular/muerte súbita recuperada, reanimación cardio-respiratoria previa, ventilación mecánica previa a la anestesia, inotrópicos preoperatorios, balón de contrapulsación aórtico preoperatorio, falla renal aguda preoperatoria (oliguria $\leq \text{ml/h}$)	2.2	3	0.905813

Variable	Definición	OR	Puntos	β
Factores del Paciente				
Angina inestable	Angina de reposo que requiere nitratos I.V. hasta la llegada a quirófano	1.5	2	0.567708
Infarto de miocardio reciente	Infarto de miocardio en los últimos 90 días	1.6	2	0.546022
Fracción expulsión	Menor de 30%	2.5	3	1.094443
Fracción expulsión	30-50%	1.5	1	0.419643
Presión sistólica arteria pulmonar	>60 mm Hg	2	2	0.767692
Cirugía de emergencia	Requerida antes del día siguiente laborable	2.8	2	0.712795
Ruptura del septum	Ruptura del septum interventricular por infarto de miocardio	3.8	4	1.462009
Cirugía diferente al by pass coronario	Cirugía cardíaca diferente a la revascularización coronaria o además de ella	1.6	2	0.542036
Cirugía de aorta torácica	Cirugía de aorta ascendente, arco o aorta descendente	3.2	3	1.159787

Fuente: Información tomada de las referencias [8] y [9]

El rendimiento del modelo desarrollado EuroSCORE aditivo evaluado en la subpoblación para tal fin (n=1497), mostró una calibración con valor del estadístico C de 7.5, con $p=0.68$, mientras la discriminación mostró una área bajo la curva de 0.76. Es decir, el modelo tuvo buena calibración y aceptable discriminación. La mortalidad promedio esperada con intervalo de confianza del 95% fue de 4.72-4.95. La mortalidad observada fue de 4.7, con IC 95% de 4.37-5.06%. Dada la sencillez de la versión aditiva, esta fue la primera en ser dada a conocer. Esta versión se puede aplicar a la cabecera del paciente recogiendo la información de las 17 variables del score. Luego, se suman los números enteros de cada factor de riesgo del score y la suma de dichos valores proporciona un número que nos permite ubicar el riesgo del paciente en uno de los tres grupos de riesgo siguientes: *bajo riesgo* (0-2 puntos), con mortalidad esperada con IC 95% de 1.27-1.29 y mortalidad observada de 0.8% con IC 95% de 0.56-1.10%; *mediano riesgo* (3-5 puntos) con mortalidad esperada con IC 95% de 2.90-2.94 y mortalidad observada de 3% con IC de 95% 2.62-3.51% y *alto riesgo* (≥ 6 puntos), con mortalidad esperada con IC del 95% de 10.93-11.54 y mortalidad observada 11.2% con IC del 95% de 10.25-12.16% [9]. Poco tiempo después fue lanzada la versión logística a través de su página oficial y luego publicados los coeficientes β usados en el cálculo [10].

Sin embargo, a los pocos años de su publicación, algunos autores señalaron limitaciones en su calibración, ya que el modelo en su versión aditiva sobrestimaba la mortalidad en pacientes de bajo riesgo y la infraestimaba en pacientes de alto riesgo [11]. Luego varios autores han publicado evidencia de que el modelo, tanto en su versión aditiva como logística en general sobrestima el riesgo [12-20]. Dado estos resultados, diversos investigadores recomendaron que el modelo fuera recalibrado. Derivado de lo anterior, la versión actualizada del score llamada EuroSCORE II fue presentada en el congreso de la *European Association for Cardio-Thoracic Surgery* (EACTS) llevada a cabo en la ciudad de Lisboa, Portugal en Octubre del 2011 y luego publicada durante el primer trimestre de 2012 [4,21].

II.3 EuroSCORE II

El EuroSCORE II se originó en una base de datos de 22381 pacientes operados de corazón en 154 hospitales de 43 países de varias partes del mundo. La base de datos fue dividida al azar en dos grupos, un grupo formado por el 75% de casos y el otro grupo por el 25% restante; el primero (n=16828 pacientes) fue para desarrollar el modelo y el segundo grupo (n=5553 pacientes) fue para su validación interna. Los factores de riesgo recabados aparecen en la siguiente tabla.

Tabla 3. Factores de riesgo relacionados con el paciente en EuroSCORE II

Factores generales	
Edad	Arteriopatía extracardíaca
Sexo	Disfunción neurológica o musculoesquelética
Peso	Diálisis
Talla	Creatinina
Diabetes	BNP (péptido cerebral natriurético)
Albumina sérica	
Factores dependientes del estado cardiológico	
Clase funcional NYHA	Infarto reciente y tamaño
Clase funcional CCS	Presión pulmonar sistólica
Disfunción sistólica	Endocarditis activa
Factores dependientes de la intervención	
Prioridad de la intervención	Tiempo de circulación extracorpórea

Tiempo de pinzamiento aórtico	Tiempo de paro circulatorio e hipotermia
Perfusión cerebral selectiva	Cirugía cardíaca previa
NYHA: New York heart Association	CCS: Canadian Cardiology Society

Fuente: datos tomados de la referencia [20]

Los datos fueron obtenidos en un período de 12 semanas (Mayo-Julio 2010). El desarrollo del modelo fue con una visión parsimoniosa, término que propone que la explicación completa más simple suele ser la mejor, lo que podríamos interpretar como que los modelos con menor número de variables predictivas posibles rinden más adecuadamente [4]. Después de efectuar múltiples ensayos, se decidió integrar las siguientes 18 variables predictivas:

Tabla 4. Variables incluidas en EuroSCORE II

Factores relacionados al paciente	Factores cardíacos relacionados
1.- Edad	11.- Clase funcional de la NYHA (disnea)
2.- Sexo femenino	12.- Clase IV por angina, SCC
3.- Depuración de creatinina	13.- Función ventricular izquierda
4.- Arteriopatía extracardíaca	14.- Hipertensión arterial pulmonar
5.- Pobre movilidad por disfunción neurológica o musculoesquelética	15.- Infarto de miocardio reciente < 90 días
6.- Cirugía cardíaca previa	Factores relacionados con la cirugía
7.- Enfermedad pulmonar crónica	16.- Carácter de la cirugía
8.- Endocarditis activa	17.- Peso de la intervención
9.- Estado preoperatorio crítico	18.- Cirugía de la aorta torácica
10. Diabetes mellitus en control con insulina	

Fuente: información tomada de la referencia [20]

Como podemos observar, en relación a la lista de variables del modelo original en el EuroSCORE (tabla 2) algunas permanecieron igual (edad, género femenino, arteriopatía extracardíaca, enfermedad pulmonar crónica, endocarditis activa, estado preoperatorio crítico, infarto de miocardio reciente, cirugía de la aorta torácica), otras fueron modificadas (creatinina, pobre movilidad por disfunción neurológica, clase funcional de la NYHA, clase IV de la Sociedad de

Cardiología de Canadá, función ventricular izquierda, hipertensión arterial pulmonar, carácter de la cirugía, peso de la intervención) y otras fueron excluidas (ruptura del septum interventricular).

Por otro lado sabemos que hay variables que guardan una relación causal estadísticamente significativa con la mortalidad en estos pacientes como la disfunción hepática severa y las concentraciones séricas del péptido cerebral natriurético (BNP) y que no fueron incluidas en la construcción del modelo. La primera, porque la albumina sérica subrogatoria de la falla hepática no mostró asociación estadísticamente significativa con la mortalidad perioperatoria y la segunda porque en su determinación se usaron distintas unidades de medida y distintas tecnologías en laboratorios diferentes. Finalmente quedaron 18 variables, una continua, seis discretas categóricas ordinales o multicategóricas y once categóricas nominales dicotómicas, que aparecen en la tabla 4.

La mortalidad fue definida como:

- muerte en el mismo hospital donde tuvo lugar la operación, antes del alta hospitalaria;
- muerte en el mismo hospital o en otro hospital, pero antes del alta hospitalaria;
- muerte dentro de los 30 días de la cirugía, independientemente de la ubicación;
- muerte dentro de los 90 días de la cirugía, independientemente de la ubicación.

Sin embargo, los investigadores sólo tuvieron información del estado vital de los pacientes egresados vivos del hospital, del 56% a los 30 días y del 44.4% a los 90 días. Dado lo anterior, decidieron dejar la definición de mortalidad como la muerte ocurrida antes del alta hospitalaria [20].

En comparación con la base de datos EuroSCORE original de 1995 (entre paréntesis) la edad promedio aumentó a 64.7 (62.5), 31% fueron mujeres (28%). Más pacientes tenían clase funcional IV de la NYHA, arteriopatía extracardíaca, disfunción renal y pulmonar. La mortalidad pronosticada por el modelo fue de 3.95 y la mortalidad hospitalaria observada fue de 4.18% en el grupo de la validación interna. La discriminación mostró un área bajo la curva ROC de 0.809, la que podemos calificar como buena [9, 20].

Sin embargo, desde el momento mismo de su publicación el modelo ha sido muy cuestionado. Los puntos débiles más frecuentemente señalados han sido: el cambio en la definición de mortalidad, la elección de algunas variables y el no considerar otras condiciones conocidas como relevantes, especialmente en el paciente anciano (fragilidad, estado mental, etc.), el no haber desarrollado

modelos pronósticos para diferentes patologías cardíacas, el haber considerado como adecuado el resultado dado por la prueba de Hosmer-Lemeshow en el grupo de validación interna, cuando su nivel fue limítrofe ($p= 0.0505$). Incluso, se aconsejó no usar el modelo para informar al paciente y su familia del riesgo quirúrgico ni tampoco para monitorizar la calidad asistencial hospitalaria [21]. Desde su publicación y hasta la actualidad se han publicado diversos estudios de validación externa, unos con resultados desfavorables y otros con resultados favorables [22, 27].

En el servicio de cardiología del Hospital Regional de Alta Especialidad del Bajío (HRAEB) el modelo EuroSCORE II es utilizado frecuentemente para valorar el riesgo de muerte en los pacientes enviados a cirugía cardiovascular y con esa información solicitar el consentimiento informado. Considerando los resultados contradictorios encontrados en múltiples estudios efectuados en diversas latitudes, es procedente preguntarnos ¿es confiable utilizar el modelo EuroSCORE II para valorar el riesgo de muerte en cirugía cardíaca mayor en nuestro hospital?. Para responder a esta pregunta, es necesario llevar a cabo la validación del mencionado modelo EuroSCORE II en nuestro medio. La pregunta anterior muestra la pertinencia y relevancia de la investigación ha desarrollar.

III. Justificación

El modelo pronóstico EuroSCORE para mortalidad en cirugía cardíaca fue publicado en 1999 y su validación en ámbitos diferentes a las condiciones de su construcción y validación fue recomendada por los autores del puntaje de riesgo [8,9,10]. La validación externa fue exitosa en múltiples centros médicos de diversas latitudes. Sin embargo, la pérdida de poder de calibración, señalada en la última década [11-19] comentada en párrafos anteriores originó que los autores del modelo llevaran a cabo modificaciones en su estructura predictiva para mejorar la calibración y tratar de mantener su buena discriminación o mejor aun mejorarla por lo que a principios del 2012 publicaron la versión actualizada de la escala llamada EuroSCORE II la cual mostró mejoras en la calibración, manteniendo un muy buen desempeño en su discriminación [20]. Nuevamente, los autores invitaron a la comunidad internacional de cardiólogos y cirujanos de corazón a evaluar su desempeño en otras latitudes. Los resultados de múltiples estudios arrojan resultados contradictorios [22-27], por lo que consideramos que es conveniente que cada centro conozca el desempeño del nuevo modelo en su propia área, para así decidir usarlo o no.

En nuestro medio hemos evaluado en un estudio retrospectivo el desempeño del modelo EuroSCORE, mostrando calibración y discriminación aceptables, tanto de la versión aditiva como de la logística [29]. En la práctica cotidiana en nuestro hospital el puntaje de riesgo EuroSCORE II es usado frecuentemente, confiando en la actualización del modelo. Sin embargo, hasta el momento no se ha llevado a cabo algún estudio que evalúe la pertinencia y el desempeño del nuevo modelo EuroSCORE II en nuestro hospital. Por lo tanto, es necesario saber si podemos confiar en dicha escala o no. Conocer sus limitantes y ventajas en su utilización.

IV. Planteamiento del problema

¿Cuál es desempeño del modelo EuroSCORE II en la población de pacientes sometidos a cirugía cardíaca en el HRAEB?

¿Cuál es la calibración y el poder discriminatorio del modelo EuroSCORE II en pacientes operados de cirugía de corazón en el HRAEB?

¿Es pertinente utilizar el modelo EuroSCORE II para calcular el riesgo de mortalidad en pacientes sometidos a cirugía cardíaca en el Hospital Regional de Alta Especialidad del Bajío?

V. Hipótesis

El desempeño del modelo EuroSCORE II en el pronóstico de la mortalidad perioperatoria de los pacientes sometidos a cirugía cardíaca en el HRAEB es adecuado.

VI. Objetivos

Objetivo General o primario:

Evaluar el desempeño del modelo EuroSCORE II en la población de pacientes sometidos a cirugía de corazón en el HRAEB

Objetivos secundarios:

Valorar el desempeño del modelo EuroSCORE II mediante la evaluación de su:

- 1.- Calibración
- 2.- Discriminación
- 3.- Describir las características epidemiológicas generales de la población de estudio

VII. Material y métodos

Tipo de estudio:

Observacional, transversal, retrospectivo.

Universo de estudio:

Expedientes clínicos físicos y electrónicos de pacientes adultos (≥ 16 años) llevados a cirugía de corazón en el servicio de cirugía cardiovascular del HRAEB en el periodo del 1 de Enero del 2008 al 31 de Diciembre del 2013.

Marco muestral:

Expedientes disponibles, legibles y completos.

Muestra:

Censal disponible en hospital

Criterios de selección

Criterios de inclusión:

Expedientes clínicos físicos y electrónicos de pacientes adultos (≥ 16 años) sometidos a cirugía de corazón en el servicio de cirugía cardiovascular del HRAEB durante el período de tiempo del 1 de Enero del 2008 al 31 de Diciembre del 2013.

Criterios de no inclusión:

Expedientes incompletos en relación a la información relevante para el estudio como las variables para calcular el modelo EuroSCORE II, expedientes de pacientes fallecidos por causas no relacionadas al procedimiento quirúrgico índice de acuerdo al grupo de investigadores, información dudosa acerca de la calidad de la información de los pacientes operados de cirugía cardiovascular.

Variables:

Variables principales del estudio:

1. Valor promedio del resultado del EuroSCORE II para mortalidad esperada
2. Mortalidad peri-operatoria observada

Tabla 5. Variables para calcular el EuroSCORE II (tipo de variable y medida)

Edad (continua, promedio \pm desviación estándar)
Sexo (categórica nominal, dicotómica, %)
Enfermedad pulmonar obstructiva crónica (categórica nominal, dicotómica, %)
Arteriopatía extracardíaca (categórica nominal, dicotómica, %)
Disfunción renal (discreta ordinal, transformada a 4 variables dicotómicas excluyentes, %). Depuración de creatinina > 85 ml/min; depuración de creatinina 51-85 ml/min; depuración de creatinina < 51 ml/min; diálisis crónica
Movilidad limitada (categórica nominal, dicotómica, %)
Endocarditis activa (categórica nominal, dicotómica, %)
Cirugía cardíaca previa (categórica nominal, dicotómica, %)
Diabetes mellitus en control con insulina (categórica nominal, dicotómica, %)
Situación preoperatoria crítica (categórica nominal, %)

Función sistólica (FEVI) (discreta, ordinal o multi-categoría, transformada a 4 variables dicotómicas excluyentes, %): FEVI > 50%; FEVI 31-50%; FEVI 21-30%; FEVI < 21%
Hipertensión arterial pulmonar (HAP) (discreta ordinal o multi-categoría, transformada a 3 variables dicotómicas excluyentes, %) HAP < 31; HAP 31-55 e HAP > 55 mm Hg
Angina de reposo (clase 4 de la SCC) (categórica nominal, dicotómica, %)
Infarto de miocardio reciente (categórica nominal, dicotómica, %)
Prioridad de la intervención (categórica, ordinal o multi-categoría, transformada a 4 variables dicotómica excluyentes, %): Electiva; urgente, emergente, de rescate o salvamento
Complejidad de la intervención (categórica, ordinal o multi-categoría, transformada a 4 variables dicotómicas excluyentes, %): Cirugía coronaria; 1 procedimiento no coronario; 2 procedimientos; 3 o más procedimientos
Cirugía sobre aorta torácica (categórica nominal, dicotómica, %)
Clase funcional preoperatoria de acuerdo a la New York Heart Association (NYHA) (categórica ordinal o multi-categoría, transformada a 4 variables dicotómicas excluyentes: Clase I, Clase II, Clase III y Clase IV de acuerdo a la gravedad de la disnea.

Otras variables: enfermedades asociadas como hipertensión arterial sistémica, diabetes mellitus, diabetes mellitus en control con dieta y/o hipoglicemiantes, hipercolesterolemia, hipertrigliceridemia, tabaquismo presente, tipo de cardiopatía, tipo de operación.

Definición operacional de las variables

- 1.- Edad: años cumplidos al momento de la cirugía
- 2.- Peso: kilogramos registrados al ingreso al hospital para efectuar la cirugía índice
- 3.- Talla: metros y centímetros registrados al ingreso al hospital para efectuar la cirugía índice
- 4.- Índice de masa corporal (IMC): peso sobre la talla elevada al cuadrado, (peso/talla²) registrados al ingreso al hospital para efectuar la cirugía índice
- 5.- Sexo: género del o de la paciente (mujer u hombre)
- 6.- Depuración de creatinina usando la fórmula de Cockcroft-Gault:
 Hombres= $140 - \frac{[edad(años) \times peso (kg)]}{[72 \times creatinina plasmática (mg/dl)]}$
 Mujeres= $140 - \frac{[edad (años) \times peso (kg) \times 0.85]}{[72 \times creatinina plasmática (mg/dl)]}$

7.- Arteriopatía extracardíaca: una o más de las siguientes condiciones:

Claudicación

Oclusión o > 50% de estenosis carotídea (una o ambas carótidas).

Amputación de algún miembro inferior por enfermedad arterial

Intervención previa o planeada de la aorta abdominal, de las extremidades o de carótidas

8.- Pobre movilidad, por disfunción neurológica o musculoesquelética

9.- Cirugía cardíaca previa: una o más cirugías antes de la cirugía índice con apertura del pericardio.

10.- Enfermedad pulmonar crónica: uso crónico de esteroides o broncodilatadores para enfermedad pulmonar.

11.- Endocarditis activa: pacientes bajo tratamiento antibiótico por endocarditis infecciosa al momento de la cirugía.

12.- Estado preoperatorio crítico: cualquiera de las siguientes condiciones previas a la cirugía: taquicardia/fibrilación ventricular/ o muerte súbita abortada; masaje cardíaco o ventilación asistida al arribar al quirófano, inotrópicos, balón de contrapulso aórtico o dispositivo de asistencia ventricular al arribar al quirófano, insuficiencia renal aguda (anuria u oliguria, < 10 ml orina/hora).

13.- Diabetes en control con insulina: el o la paciente requieren insulina para el control de la diabetes mellitus.

14.- Clase funcional por disnea de la New York Heart Association

Clase I: sin limitación a ninguna actividad física

Clase II: ligera limitación a la actividad física. Confortable en reposo. Disnea al caminar dos o más cuadras o subir 2 o más niveles o pisos.

Clase III: Marcada limitación a la actividad física. Confortable en reposo. Actividad física menor a la ordinal produce fatiga. No tolera caminar una cuadra o subir un piso.

Clase IV: Disnea en el reposo.

15.- Angina de reposo o clase IV de la Sociedad de Cardiología de Canadá: angina con mínimos esfuerzos y/o aun en reposo.

16.- Función sistólica del ventrículo izquierdo (FEVI), medida por métodos no invasivos (Ecocardiografía transtorácica o por métodos invasivos, cateterismo cardíaco), expresada en %.

> 50%; 31-50%; 21-30%; ≤ 20%.

17.- Infarto de miocardio reciente: infarto de miocardio con o sin elevación del segmento ST en los últimos 90 días.

18.- Hipertensión arterial pulmonar: medición de la presión arterial pulmonar en forma directa (cateterismo) o indirecta (ecocardiograma).

- a.- No (≤ 30 mm Hg)
- b.- Moderada (31-55 mm Hg)
- c.- Severa (> 55 mm Hg)

19.- Carácter o Urgencia de la cirugía

- a.- electiva: programación habitual
- b.- urgente: paciente admitido por descompensación de su cardiopatía y que no puede ser egresado antes de efectuar el procedimiento quirúrgico requerido.
- c.- emergencia: procedimiento quirúrgico que debe realizarse antes de iniciar las labores habituales del día siguiente.
- d.- salvamento: paciente que requiere de reanimación cardiopulmonar en su camino al quirófano o antes de la inducción anestésica.

20.- Peso de la intervención:

- a.- cirugía de bypass coronario
- b.- cirugía diferente al bypass coronario, un solo procedimiento
- c.- dos procedimientos (p.e., bypass coronario+ cambio valvular aórtico)
- d.- tres procedimientos (p.e., bypass coronario+cambio valvular aórtico+cambio valvular mitral)

21.- Cirugía de aorta torácica

Procedimiento quirúrgico que involucra cualquier intervención por arriba de la válvula aortica (aorta ascendente, cayado, descendente.).

22.- EuroSCORE II: Valor del resultado de la suma de las variables predictivas del EuroSCORE II para obtener la mortalidad esperada en cada paciente del estudio, mediante la calculadora en línea:

www.euroscore.org

23.- Mortalidad peri-operatoria observada: mortalidad que se presenta en el quirófano o en el postoperatorio durante la hospitalización índice en el HRAEB.

VIII. Análisis estadístico

Las variables continuas se reportan como medias y desviación estándar si su distribución fue aproximadamente normal, o como mediana y rangos intercuartílicos en el caso de no serlo. La comparación de dos medias se hizo con t de student. Las variables categóricas se reportan como porcentajes de frecuencias y se compararon por medio de la prueba chi cuadrada. Se consideró nivel de significancia con $p \leq 0.05$. La variable respuesta (mortalidad perioperatoria), también se reporta como porcentaje.

La calibración se evaluó mediante la prueba de bondad de ajuste de Hosmer-Lemeshow (H-L). Lo que hace esta prueba es comprobar si el modelo puede predecir lo que se observa en la realidad. Evalúa la distancia entre lo esperado de acuerdo al modelo con lo observado en la realidad. La prueba consiste en dividir el recorrido de valores de la variable dependiente y en una serie de intervalos (deciles). Se cuenta en cada intervalo o decil los valores esperados y los observados para cada uno de los dos resultados posibles de la variable dependiente dicotómica (muerto/vivo). El valor esperado teórico es el calculado usando el modelo construido y el valor observado es el que se tiene en la realidad. La prueba calcula un estadístico C a partir de la diferencia entre los valores de mortalidad observados y los valores de mortalidad esperados por el modelo en los distintos intervalos del grupo en estudio. Cuanto menor sea el valor del estadístico C, mejor será la calibración; *la mortalidad predicha y la observada se acercan, se juntan* [4,5]. Un valor de $p >$ de 0.05 sugiere que el modelo ajusta bien los datos (los valores no difieren significativamente entre si). Para evaluar la calibración fue seleccionada esta prueba porque fue la prueba empleada en la validación interna del modelo EuroSCORE II y también por ser la prueba más utilizada en los estudios de validación de modelos predictivos. Obtuvimos los resultados de mortalidad esperada y mortalidad observada por deciles de riesgo, el estadístico C y su valor de p .

Para valorar la discriminación del EuroSCORE II, es decir, la capacidad del modelo de identificar los pacientes que fallecieron de los que sobrevivieron, utilizamos el área bajo la curva (AUC, *area under the curve*) del *receiver operating characteristic* o curva ROC, con un intervalo de confianza del 95%. Esta fue la misma prueba utilizada para la validación interna del EuroSCORE II.

Se utilizó hoja de cálculo excel del programa Microsoft versión 2013 para elaborar la base de datos y el programa SPSS versión 23 para efectuar el análisis descriptivo y la estadística inferencial.

IX. Consideraciones éticas

Este estudio corresponde a un diseño con Riesgo I de acuerdo a lo señalado en el Título Segundo, Artículo 17 del Reglamento de la Ley General de Salud en materia de investigación en Salud y sólo requirió de la declaración de confidencialidad de los datos personales.

No se manejó ninguna forma de energía, sustancias químicas, incluyendo fármacos, material biológico u objetos en movimiento que pudieran haber sido lesivos a los sujetos de investigación, investigadores, familiares o entorno, por lo que no conllevó problemas de bioseguridad.

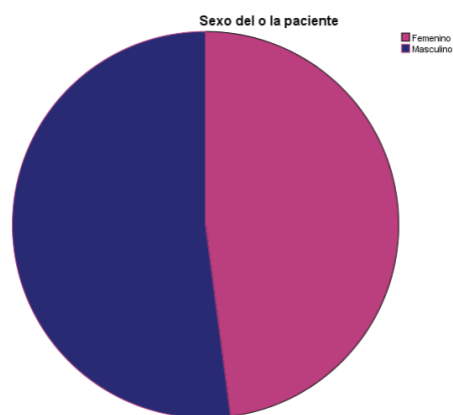
X. Resultados

Características generales de la población evaluada

Se solicitó y obtuvo del departamento de estadística del HRAEB el listado de expedientes de pacientes sometidos a cirugía de corazón del 1 de enero del 2008 al 31 de diciembre de 2013. Se identificaron 343 expedientes de pacientes sometidos a cirugía de corazón en el periodo de tiempo señalado. De estos 343 expedientes se eliminaron 5 de ellos. Cuatro por no tener la información requerida para efectuar la valoración de riesgo EuroSCORE II y uno por estar triplicado, con información diferente. Por lo tanto, quedaron 338 expedientes que cumplieron con los criterios de selección y que son la población evaluada. Las características generales de la población estudiada son las siguientes:

Estadística descriptiva. Variables categóricas

Distribución según sexo



Sexo del o la paciente

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido Femenino	162	47.9	47.9	47.9
Masculino	176	52.1	52.1	100.0
Total	338	100.0	100.0	

Fuente: Directa

Las características demográficas se resumen en la siguiente tabla:

Tabla 6. Características demográficas de la población estudiada

Población evaluada (n)	338	
Mujeres, número y porcentaje (n, %)	162	(47.9%)
Peso promedio \pm DE	67.21 \pm 14.49	
Estatura promedio \pm DE	1.60 \pm 0.09	
Índice de masa corporal (IMC) promedio \pm DE	26.07 \pm 5.05	
Clasificación IMC: (n y %)		
Insuficiencia ponderal (<18.5)	17	(05.02%)
Peso normal (18.5-24.9]	135	(39.94%)
Sobrepeso (25.0-29.9]	111	(32.84%)
Obesidad I (30.0-34.9]	61	(18.04%)
Obesidad II (35.0-39.9]	11	(03.25%)
Obesidad III (\geq 40.0)	3	(0.88%)

Fuente: Directa

DE: desviación estándar

En relación al peso y de acuerdo al IMC, sólo cerca del 40% tenían peso normal mientras que el 32.84% de la población tenían sobrepeso y el 22.17 algún grado de obesidad. La suma de sobrepeso y obesidad fue de 55.01, cifra menor a la de la población general mexicana que es de alrededor del 70%.

Las enfermedades concomitantes o asociadas fueron las siguientes:

Tabla 7. Comorbilidades de la población estudiada (n y %)

Hipertensión arterial sistémica	153	(45.3%)
Hipercolesterolemia	78	(23.1%)
Hipertrigliceridemia	105	(31.1%)
Diabetes (total)	91	(26.9%)
Diabetes insulino-dependiente	42	(12.4%)
Diabetes no insulino-dependiente	49	(14.49%)
Tabaquismo	104	(30.8%)

Fuente: Directa

La frecuencia de hipertensión arterial y de diabetes mellitus fueron mayores que las observadas en la población general mexicana (45.3% vs 30% para la HAS y 26.9% vs 14% para DM), mientras la frecuencia de tabaquismo fue muy semejante a lo observado en la población general de México.

Tabla 8. Características de la cirugía de la población estudiada

Revascularización aorto-coronaria aislada (n y porcentaje)	90	(26.6%)
Un procedimiento	189	(55.9%)
Dos procedimientos	38	(11.2%)
Tres procedimientos	21	(06.2%)
Total	338	(100.0%)
Electiva	179	(53.0%)
Urgente	155	(45.8%)
Emergente	3	(0.9%)
Salvamento	1	(0.3%)
Total	338	(100.0%)

Fuente: Directa

En relación al peso de la intervención o procedimientos efectuados, la mayor parte fueron de un procedimiento, seguida de la revascularización aorto-coronaria, generalmente cambio valvular (aórtico seguido de mitral) y el menor número fue de tres procedimientos.

El carácter de la cirugía fue tipo electiva en poco más de la mitad de los casos (53%) seguida de cirugías urgentes (45.9%). Las cirugías emergentes fueron raras (0.9%) y las de salvamento excepcionales (0.3%). Llama la atención la alta frecuencia de cirugías urgentes comparada con la frecuencia reportada en otras series.

Tabla 9. Distribución, número y porcentaje de variables del EuroSCORE II

Factores relacionados con el paciente		
Edad (media±desviación estándar)	49.87±16.61	
Mujeres (n y porcentaje)	162	(47.9%)
Arteriopatía extracardíaca	19	(5.6%)
Depuración de creatinina, Cockcroft-Gault (ml/min) (n y porcentaje)		
Mayor de 85	201	(59.46 %)
Entre 50-85	101	(29.88 %)
Menor de 50	16	(4.73 %)
En diálisis	20	(5.91 %)
Pobre movilidad	25	(7.39 %)
Cirugía cardíaca previa	47	(13.9 %)
Enfermedad pulmonar obstructiva crónica	21	(6.2 %)
Endocarditis activa	24	(7.4 %)
Estado preoperatorio crítico	49	(14.5 %)
Diabetes en control con insulina	42	(12.4 %)
Factores relacionados con la condición cardíaca		
Clase funcional de la NYHA (n y porcentaje)		
Clase I	46	(13.6 %)
Clase II	172	(50.9 %)
Clase III	95	(28.1 %)
Clase IV	25	(7.4 %)
Angina clase IV SCC	39	(11.5 %)
Función ventricular izquierda (porcentaje FEVI) (n y porcentaje)		
Mayor de 50	223	(65.97 %)
31-50	113	(33.43 %)
21-30	2	(0.59 %)
≤ 20	0	
Infarto de miocardio reciente	46	(13.64 %)
Hipertensión arterial pulmonar (PSAP, mm Hg) (n y porcentaje)		
No	115	(30.02%)
Moderada (31-54)	171	(50.59 %)
Severa (≥ 55)	53	(15.68 %)
Factores relacionados con la cirugía		
Carácter de la cirugía (n y porcentaje)		
Electiva	179	(53.0 %)
Urgente	155	(45.9 %)
Emergente	3	(0.9 %)
Salvamento	1	(0.3 %)
Peso de la intervención (n y porcentaje)		
CABG aislada	90	(26.6 %)
Un procedimiento no CABG	189	(55.9%)
Dos procedimientos	38	(11.2 %)
Tres procedimientos	21	(6.2 %)
Cirugía sobre la aorta torácica	23	(6.8 %)

Fuente: Directa

Estadística descriptiva. Variables numéricas

Edad

Los valores de distribución central y de dispersión de la variable edad aparecen en el cuadro siguiente:

Edad		
N	Válido	338
	Perdidos	0
Media		49.87
Error estándar de la media		.903
Mediana		54.00
Moda		63
Desviación estándar		16.611
Varianza		275.910
Rango		64
Mínimo		16
Máximo		80
Percentiles	25	37.00
	50	54.00
	75	62.00

Fuente: Directa

El promedio de edad fue de 49.87 con desviación estándar de 16.61. El valor mínimo fue de 16 años y el mayor de 80 años, con rango de 64 años. La edad media y su desviación estándar sugiere que la población estudiada fue relativamente joven. La distribución por grupos de edad es la siguiente:

Tabla 10. Distribución etaria

	Frecuencia	Porcentaje
Menos de 20 años	26	7.69
De 21 a 30 años	34	10.06
De 31 a 40 años	37	10.95
De 41 a 50 años	42	12.43
De 51 a 60 años	101	29.88
De 61 a 70 años	72	21.3
Más de 70 años	26	7.69
Total	338	100

Fuente: Directa

El grupo etario de mayor frecuencia fue de los 51 a los 60 años, con un casi 30%, mientras que el grupo menor fueron dos, el grupo menor de 20 años y el grupo de más de 70 años, ambos con el 7.69%.

Resultado de las otras variables numéricas del estudio:

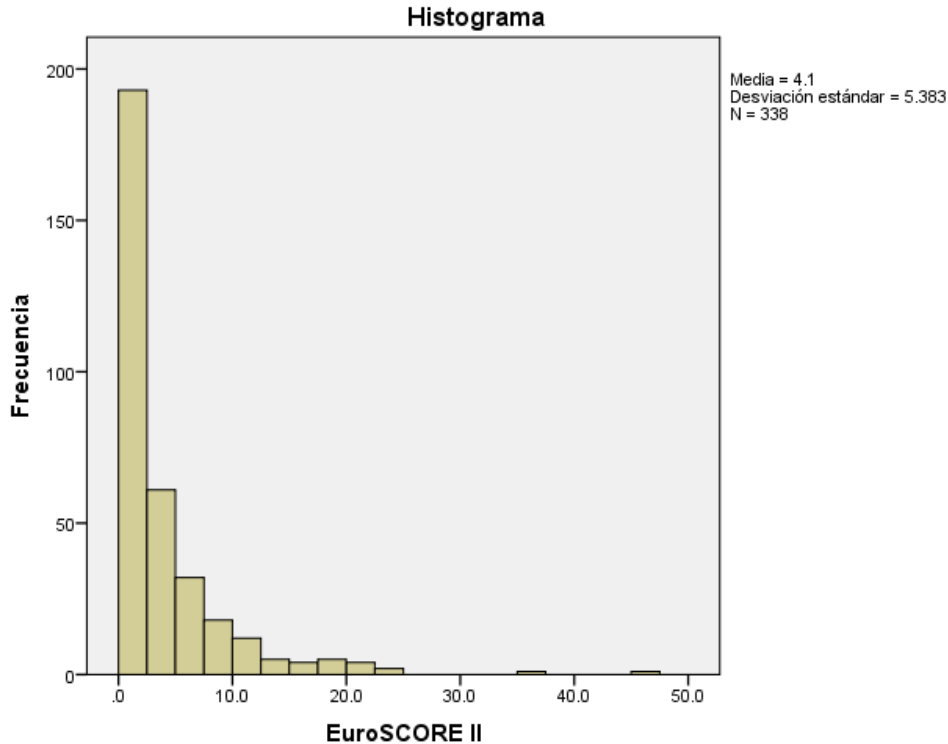
Estadísticos descriptivos

	N	Rango	Mínimo	Máximo	Media	Desviación estándar
Peso	338	83.5	35.0	118.5	67.211	14.4989
Estatura	338	.7	1.3	2.0	1.606	.0934
Índice de masa corporal	338	28	14	41	26.07	5.052
Fracción de expulsión del ventrículo izquierdo	338	70	22	92	56.10	11.952
Hipertensión arterial pulmonar	335	96	14	110	40.42	16.992
EuroSCORE II	338	45.5	.5	46.0	4.104	5.3828

EuroSCORE II

Estadísticos

N	Válido	338
	Perdidos	0
Media		4.104
Mediana		1.965
Moda		.8
Desviación estándar		5.3828
Varianza		28.974
Rango		45.5
Mínimo		.5
Máximo		46.0
Percentiles	25	1.088
	50	1.965
	75	4.958



Como podemos observar, la distribución de esta variable numérica es asimétrica, no tiene una distribución normal y por lo tanto la mediana es el mejor estadístico para ubicar el centro de la distribución.

Estadística inferencial

De los 338 pacientes operados, hubo 37 fallecimientos (10.9%), tal como lo podemos observar en tabla siguiente:

Defunciones durante la hospitalización índice

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	No	301	89.1	89.1	89.1
	Si	37	10.9	10.9	100.0
	Total	338	100.0	100.0	

Fuente: Directa

Calibración

De acuerdo a lo señalado en el apartado de análisis estadístico, evaluamos la calibración del puntaje EuroSCORE II mediante la chi cuadrada de Hosmer-Lemeshow, obteniendo el siguiente resultado:

Resumen del modelo

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	202.109 ^a	.089	.178

a. La estimación ha terminado en el número de iteración 5 porque las estimaciones de parámetro han cambiado en menos de .001.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	Gl	Sig.
1	13.222	8	.104

La Chi cuadrada de bondad de ajuste de Hosmer-Lemeshow obtenida fue de 13.222, con 8 grados de libertad, tiene una $p= 0.104$, (no significativa), lo que sugiere no diferencia estadísticamente significativa entre la mortalidad pronosticada por el EuroSCORE II y la observada en la realidad. Es decir, la línea hipotética de los valores observados y la línea hipotética de los valores esperados no se apartan significativamente. Este resultado es compatible con que el modelo EuroSCORE II tiene buena calibración en la población evaluada.

En la siguiente tabla de contingencia obtenida con la prueba de Hosmer-Lemeshow podemos ver la distribución de los valores esperados vs los valores observados en la realidad en la población, dividida en deciles.

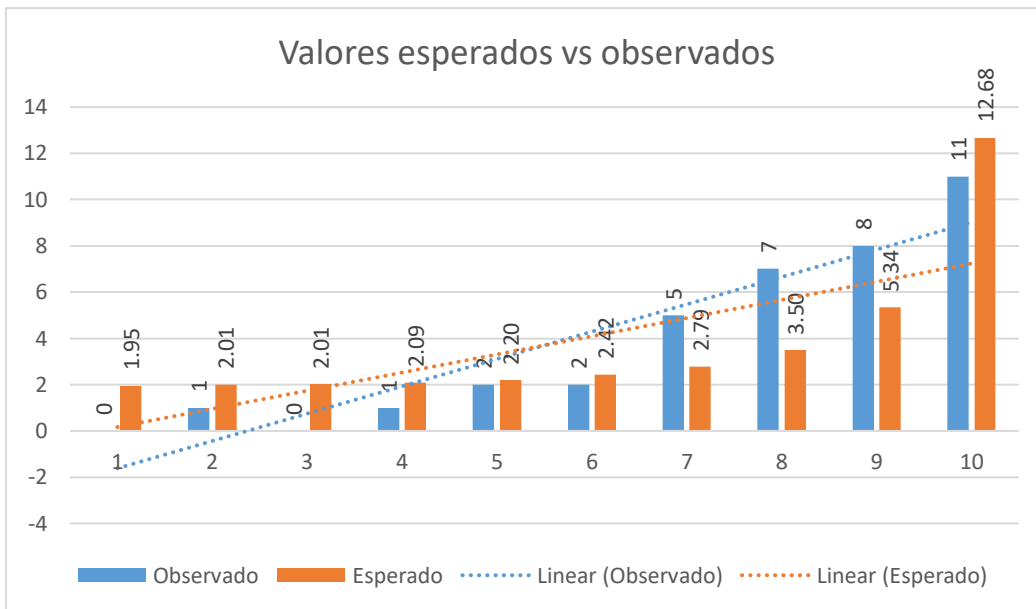
Tabla de contingencia para la prueba de Hosmer y Lemeshow

		Fallecidos durante la hospitalización índice = No		Fallecidos durante la hospitalización índice = Si		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	35	33.049	0	1.951	35
	2	34	32.988	1	2.012	35
	3	34	31.986	0	2.014	34
	4	33	31.911	1	2.089	34
	5	32	31.797	2	2.203	34
	6	32	31.576	2	2.424	34
	7	29	31.213	5	2.787	34
	8	27	30.500	7	3.500	34
	9	26	28.659	8	5.341	34
	10	19	17.323	11	12.677	30

Fuente: Directa

Cabe mencionar que en los últimos años se ha cuestionado el concepto de que un valor de la Chi cuadrada de bondad de ajustes de Hosmer-Lemeshow estadísticamente no significativa ($p > 0.05$) traduzca necesariamente buena calibración. “La estadística de Hosmer-Lemeshow compara proporciones pero es ciertamente imperfecta”. “Un resultado no significativo ($p > 0.05$) significa que no hay evidencia de una mala calibración, pero no necesariamente es compatible con una buena calibración” [21, 28].

Algunos autores proponen que la *razón de mortalidad ajustada al riesgo* (RAMR), que consiste en la razón de la mortalidad observada entre la mortalidad esperada ($RAMR = O/E$) es una medida que puede ser útil para evaluar la calibración. Una razón de 1.0 significa que el score o modelo de prueba predice la mortalidad perfectamente (fallece el mismo número de pacientes (observados) que el número de esperados). Una $RAMR > 1.0$ significa que el modelo subestima la mortalidad mientras que una $RAMR < 1.0$ significa que el modelo predice en exceso la mortalidad (la sobrestima) [29]. En el gráfico de abajo se muestra la relación de la RAMR por cada decil de riesgo con los datos de la tabla de contingencia que aparece en el cuadro de arriba.

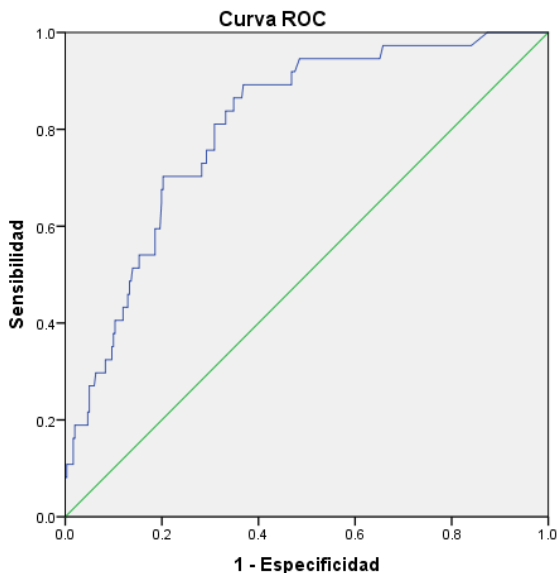


Como podemos observar, hasta el decil 4 los valores observados están por debajo de los valores esperados. En el decil 5 casi se empatan para luego a partir del decil 6 los valores observados se sitúan por arriba de los valores esperados.

Este resultado sugiere que en los primeros cuatro deciles el riesgo está sobre-estimado, es adecuado en el decil 5, pero a partir del decil 6 y hasta el decil 9 el modelo subestima la mortalidad. Finalmente en el decil 10 los valores observados y los esperados se acercan nuevamente. En este decil, la RARM es de 0.91. Sin embargo, la razón de mortalidad ajustada al riesgo (RAMR) global, es de 2.65 (10.9/4.1), compatible con el concepto de que el modelo en general subestima la mortalidad en esta población estudiada.

Discriminación

La discriminación fue valorada mediante el área bajo la curva ROC, obteniendo el resultado siguiente:



Los segmentos de diagonal se generan mediante empates.

Área bajo la curva ROC

VARIABLES DE RESULTADO DE PRUEBA: EuroSCORE II

Área	Error estándar ^a	Significación asintótica ^b	95% de intervalo de confianza asintótico	
			Límite inferior	Límite superior
.806	.034	.000	.739	.872

El resultado obtenido en la curva ROC (ABC) fue de 0.806, (IC 95% 0.739-0.872), compatible con una buena discriminación del score.

De acuerdo a los resultados obtenidos con la prueba Chi2 de bondad de ajuste podemos decir que el modelo tiene una calibración aceptable en general y el ABC ROC sugiere buena discriminación. Con estos resultados por lo tanto rechazamos la hipótesis nula que establece: La calibración y la discriminación del modelo EuroSCORE II no son adecuadas para valorar el riesgo de muerte en pacientes sometidos a cirugía cardíaca en el HRAEB y, validamos la hipótesis alterna: La calibración y la discriminación del modelo EuroSCORE II son adecuadas para valorar el riesgo de muerte en pacientes sometidos a cirugía cardíaca en el HRAEB.

**Otros resultados de interés
Género y Defunciones**

Defunciones durante la hospitalización índice

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	No	301	89.1	89.1	89.1
	Si	37	10.9	10.9	100.0
	Total	338	100.0	100.0	

Fuente: Directa

Como ya habíamos visto en párrafos anteriores y aparece en el cuadro de arriba, se observaron 37 defunciones en los 338 pacientes operados (10.9%). Cuando analizamos las defunciones de acuerdo al sexo, observamos que hay un mayor número de mujeres fallecidas (22/162, 13.58%) en comparación a los hombres fallecidos (15/176, 8.52 %), como lo podemos observar enseguida.

Tabla cruzada de Género y defunciones durante la hospitalización índice

		Fallecidos durante la hospitalización		Total
		No	Si	
Sexo del o la paciente	Femenino	140	22	162
	Masculino	161	15	176
Total		301	37	338

Aquí podemos plantearnos la hipótesis estadística siguiente:

H0: No hay diferencia en la mortalidad entre mujeres y hombres.

H1: Hay diferencia en la mortalidad entre mujeres y hombres.

Efectuamos prueba de chi², obteniendo el resultado siguiente:

Prueba de chi-cuadrado

	Valor	Gl	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	2.213 ^a	1	.137		
Corrección de continuidad ^b	1.725	1	.189		
Razón de verosimilitud	2.219	1	.136		
Prueba exacta de Fisher				.164	.094
N de casos válidos	338				

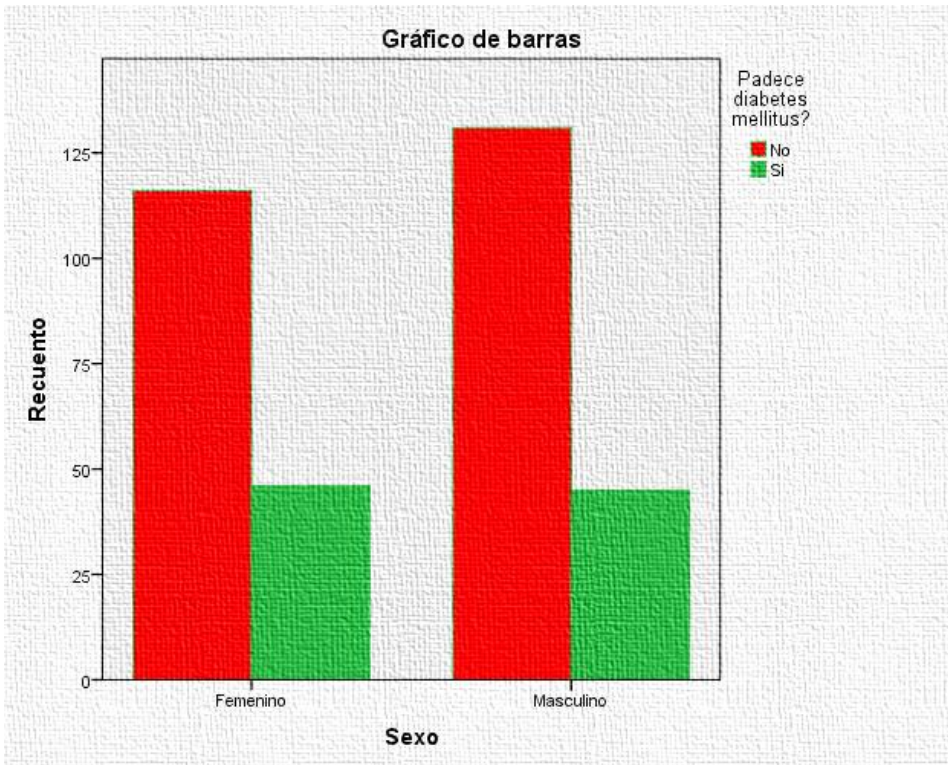
La Chi² de Pearson es de 2.213, con valor de p (0.094), estadísticamente no significativo. **Por lo tanto, No rechazamos la H₀ de no diferencias.** Podemos concluir que si bien hay una tendencia a mayor número de defunciones en las mujeres, la diferencia no es estadísticamente significativa en la población evaluada.

Cuando evaluamos si los padecimientos crónicos (diabetes mellitus, hipertensión arterial, hipercolesterolemia y tabaquismo) se distribuían de manera diferente entre hombres y mujeres, encontramos los siguientes resultados:

Género y Diabetes mellitus

Tabla cruzada Sexo y Diabetes mellitus

		Diabetes mellitus		Total
		No	Si	
Sexo	Femenino	116	46	162
	Masculino	131	45	176
Total		247	91	338



De acuerdo a las frecuencias que aparecen en el cuadro y el tamaño de frecuencias en el gráfico, pareciera no haber diferencias. Para aclararlo, efectuamos Chi² y establecimos las siguientes hipótesis:

H₀: La frecuencia de diabetes mellitus (DM) es semejante o igual tanto en mujeres como en hombres.

H₁: La frecuencia de diabetes mellitus (DM) es diferente en mujeres y en hombres.

Pruebas de chi-cuadrado

	Valor	gl	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	.343 ^a	1	.558		
Corrección de continuidad ^b	.214	1	.644		
Razón de verosimilitud	.342	1	.558		
Prueba exacta de Fisher				.624	.322
N de casos válidos	338				

Como podemos observar, el valor de p no es estadísticamente significativo. Concluimos:

En base al resultado anterior, No rechazamos la H₀ de no diferencias

Diabetes mellitus dependiente de insulina

Para el caso de la DM dependiente de insulina, observamos la frecuencia de esta en la tabla de abajo.

Tabla cruzada Sexo *Diabetes en control con insulina

		Diabetes en control con insulina		Total
		No	Si	
Sexo	Femenino	142	20	162
	Masculino	154	22	176
Total		296	42	338

Como podemos ver, la diabetes mellitus dependiente de insulina fue muy similar: 20/162 (12.34%) en mujeres vs 22/176 (12.50%) en hombres. Sin embargo, efectuamos prueba de hipótesis. Establecimos la siguiente hipótesis:

H0: La frecuencia de DM insulino-dependiente es igual tanto en mujeres como en hombres.

H1: La frecuencia de DM insulino-dependiente es diferente en mujeres y en hombres.

Pruebas de chi-cuadrado

	Valor	gl	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	3.229 ^a	1	.072		
Corrección de continuidad ^b	2.349	1	.125		
Razón de verosimilitud	2.791	1	.095		
Prueba exacta de Fisher				.107	.069
N de casos válidos	338				

Como podemos observar, el valor de p no es estadísticamente significativo. Concluimos:

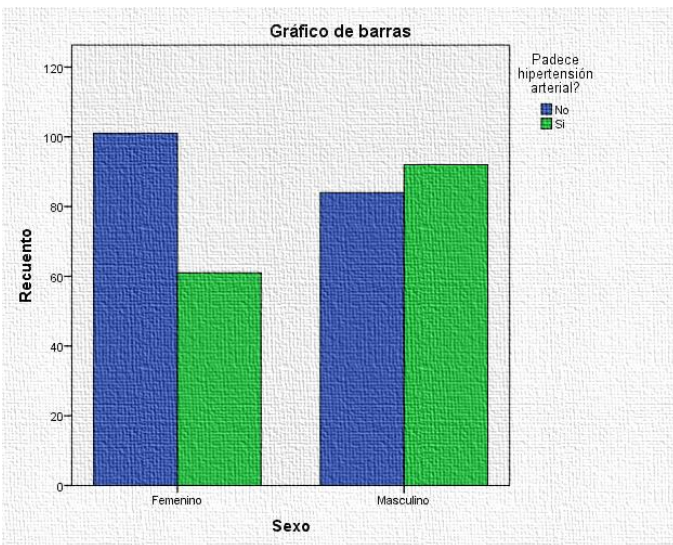
No se rechaza la H0 de no diferencias. La DM insulino-dependiente es igual de frecuente entre mujeres y hombres

Género e Hipertensión arterial

En cambio, para el caso de hipertensión arterial sistémica (HAS), observamos un mayor número de hombres en relación a las mujeres: 92/176 (52.27%) vs 61/162 (37.65%) como se observa en la tabla y en el gráfico de abajo.

Tabla cruzada Sexo y hipertensión arterial

		hipertensión arterial		Total
		No	Si	
Sexo	Femenino	101	61	162
	Masculino	84	92	176
Total		185	153	338



Para saber si esa diferencia que parece ser significativa lo es, aplicamos la prueba de Chi cuadrado. Establecimos la prueba de hipótesis:

H0: La frecuencia de HAS es igual tanto en mujeres como en hombres.

H1: La frecuencia de HAS es diferente en mujeres y en hombres. (prueba de dos colas)

Pruebas de chi-cuadrado

	Valor	gl	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	7.276 ^a	1	.007		
Corrección de continuidad ^b	6.698	1	.010		
Razón de verosimilitud	7.309	1	.007		
Prueba exacta de Fisher				.009	.005
N de casos válidos	338				

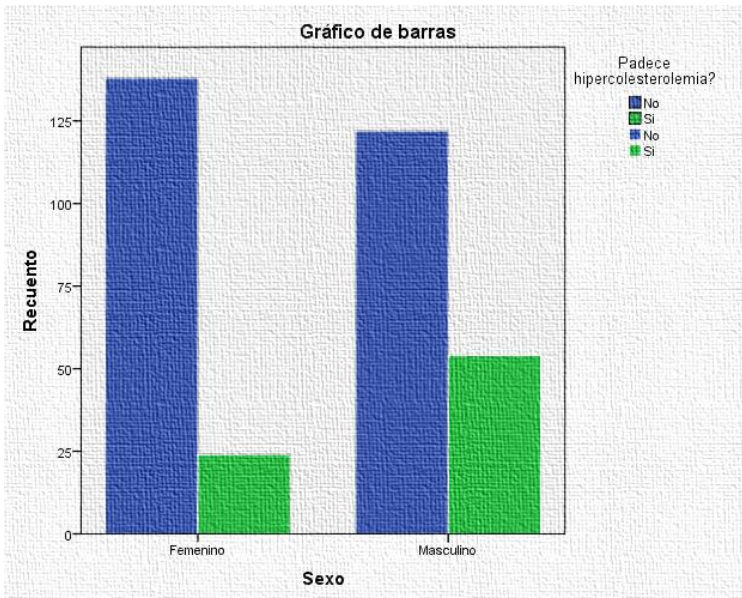
Como vemos, el valor de p es altamente significativo (p 0.005) El resultado confirma la diferencia existente en padecer HAS entre hombres y mujeres, a favor de los hombres, Por lo tanto, **rechazamos la H0 y validamos la H1.** La HAS es más frecuente en hombres que en mujeres en esta población evaluada.

Género y Colesterol sérico elevado

La elevación de las concentraciones séricas de colesterol total (≥ 200 mg/dl) también parece ser más frecuente en hombres que en mujeres, como lo podemos observar en la tabla y el gráfico de abajo.

Tabla cruzada Sexo y hipercolesterolemia

		Hipercolesterolemia		Total
		No	Si	
Sexo	Femenino	138	24	162
	Masculino	122	54	176
Total		260	78	338



Mientras en los hombres se observa que el 30.68% tienen hipercolesterolemia, en las mujeres la frecuencia es de 14.81%, la mitad. Para verificar si la diferencia observada es estadísticamente significativa efectuamos una chi cuadrada. Establecimos la siguiente hipótesis

H0: La frecuencia de colesterol alto es semejante en mujeres que en hombres

H1: La frecuencia de colesterol alto es mayor en hombres que en mujeres (una cola)

Pruebas de chi-cuadrado

	Valor	gl	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	11.964 ^a	1	.001		
Corrección de continuidad ^b	11.087	1	.001		
Razón de verosimilitud	12.246	1	.000		
Prueba exacta de Fisher				.001	.000
N de casos válidos	338				

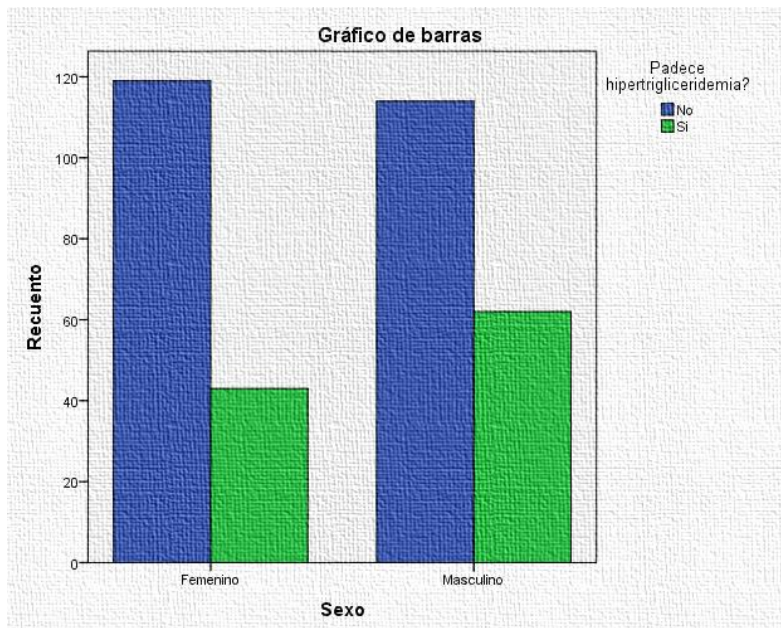
Como podemos ver, la p es altamente significativa ($p=0.000$). **Por lo tanto, rechazamos la H0 y validamos la H1: La hipercolesterolemia es más frecuente en hombres que en mujeres en esta población estudiada.**

Género y elevación de Triglicéridos séricos

La frecuencia de tener triglicéridos séricos elevados por arriba del nivel de normalidad (200 mg/dl) es la que se observa en la tabla y el gráfico que aparecen a continuación:

Tabla cruzada Sexo e hipertrigliceridemia

		Hipertrigliceridemia		Total
		No	Si	
Sexo	Femenino	119	43	162
	Masculino	114	62	176
Total		233	105	338



Como podemos observar la frecuencia de hipertrigliceridemia es mayor en hombres que en mujeres. Para probar si esa diferencia es estadísticamente significativa efectuamos prueba chi². Establecimos la hipótesis estadística siguiente:

H0: La concentración elevada de triglicéridos en sangre es semejante entre mujeres y hombres

H1: La concentración elevada de triglicéridos en sangre es diferente entre mujeres y hombres (dos colas)

Pruebas de chi-cuadrado

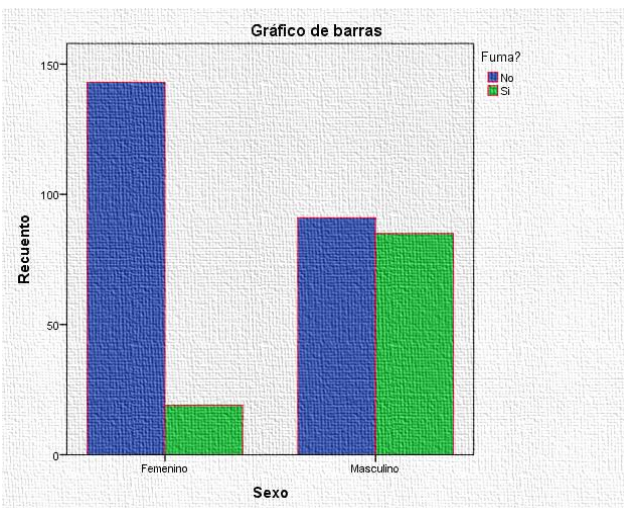
	Valor	gl	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	2.971 ^a	1	.085		
Corrección de continuidad ^b	2.579	1	.108		
Razón de verosimilitud	2.984	1	.084		
Prueba exacta de Fisher				.100	.054
N de casos válidos	338				

Dado el valor de la p 0.054, aceptamos la H0 de no diferencias en la concentración sérica de triglicéridos entre mujeres y hombres en esta población evaluada

Género y tabaquismo

Tabla cruzada Sexo *Fuma?

		Fuma/Fumó		Total
		No	Si	
Sexo	Femenino	143	19	162
	Masculino	91	85	176
Total		234	104	338



Como podemos observar a simple vista en el cuadro y gráfico de arriba, el tabaquismo parece ser más frecuente en los hombres que en las mujeres, con 48.29 % en hombres vs 11.72% en mujeres. Para verificar que esta diferencia es estadísticamente significativa, efectuamos Chi cuadrada. Establecimos la hipótesis de prueba siguiente:

H0: El tabaquismo es igual o semejante en su frecuencia entre mujeres y hombres

H1: El tabaquismo es diferente en su frecuencia entre hombres y mujeres (más frecuente en los hombres, prueba de una sola cola)

Pruebas de chi-cuadrado

	Valor	Gl	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	52.951 ^a	1	.000		
Corrección de continuidad ^b	51.248	1	.000		
Razón de verosimilitud	56.353	1	.000		
Prueba exacta de Fisher				.000	.000
N de casos válidos	338				

El resultado confirma que hay diferencia estadística y que esta es significativa, con valor de $p= 0.000$.

Dado lo anterior, rechazamos la H0 y la evidencia favorece la H1. El tabaquismo es más frecuente en los hombres que en las mujeres en esta población estudiada.

XI. Discusión

En 1999 fue publicado el análisis del perfil de riesgo de 19030 pacientes sometidos a cirugía cardíaca general en ocho países europeos. Se identificaron 17 factores asociados estadísticamente a la mortalidad hospitalaria o a 30 días después de la cirugía. La mortalidad hospitalaria general fue del 4.8% [4]. Esta base de datos sirvió para que el grupo de trabajo desarrollara el modelo pronóstico llamado EuroSCORE [8,9]. El score modelado fue llamado EuroSCORE aditivo cuyo rendimiento mostró calibración con estadístico C de 7.5 y $p=0.68$, mientras la discriminación mostró una área bajo la curva ROC de 0.76. Es decir, el modelo tuvo buena calibración y aceptable discriminación. Poco tiempo después fue lanzada la versión logística, con el valor de los coeficientes β usados en el cálculo [10].

El sistema EuroSCORE en su versión aditiva y logística tuvo mucho éxito, se utilizó en todo el mundo tanto para medir el riesgo de mortalidad peri-operatoria así como para valorar la calidad de los servicios de cirugía cardíaca, con más de 1300 citas formales en la literatura médica [4]. Sin embargo, en la última década algunos autores señalaron que si bien la discriminación seguía siendo buena, su calibración era ya deficiente [11-19]. Derivado de lo anterior, los autores del modelo EuroSCORE decidieron actualizar el score al que llamaron EuroSCORE II, que fue presentado en el congreso de la *European Association for Cardio-Thoracic Surgery (EACTS)* en el 2011 y publicado en el 2012 [4,20].

El modelo EuroSCORE II se originó en una base de datos de 22 381 pacientes operados de corazón en 154 hospitales de 43 países de varias partes del mundo. El score fue modelado con 18 variables. Diez variables relacionadas con el paciente: edad, género femenino, funcionamiento renal, arteriopatía extracardíaca, pobre movilidad (neurológica o musculo-esquelético), cirugía cardíaca previa, enfermedad pulmonar crónica, endocarditis activa, estado preoperatorio crítico y diabetes mellitus en control con insulina. Cinco variables se relacionaron con la condición cardíaca, a saber: clase funcional por disnea de la clasificación de la Sociedad de Cardiólogos de Nueva York, clase 4 de angina de la Sociedad de Cardiólogos de Canadá, función ventricular izquierda, infarto de miocardio reciente (menos de 90 días), y presencia y grado de hipertensión arterial pulmonar. Tres variables se relacionan con factores de la cirugía: carácter o urgencia, peso de la intervención y si se

involucra o no a la aorta torácica. De las 18 variables previamente señaladas, una fue continua, seis discretas categóricas ordinales o multicategóricas y once categóricas nominales dicotómicas.

En comparación con la base de datos EuroSCORE original de 1995 (entre paréntesis) la edad promedio aumentó a 64.7 (62.5), 31% fueron mujeres (28%). Más pacientes tenían clase funcional IV de la NYHA, arteriopatía extracardíaca, disfunción renal y pulmonar. La mortalidad pronosticada por el modelo fue de 3.95 y la mortalidad hospitalaria observada fue de 4.18% en el grupo de la validación interna. La discriminación mostró área bajo la curva ROC de 0.809 y la calibración con Chi² de Hosmer-Lemeshow con p “límitrofe” (p=0.0505 [21]).

Sin embargo, desde el momento mismo de su publicación el modelo ha sido muy cuestionado. Los puntos débiles más frecuentemente señalados han sido: el cambio en la definición de mortalidad, la elección de algunas variables y el no considerar otras condiciones conocidas como relevantes, especialmente en el paciente anciano (fragilidad, estado mental, etc.), el no haber desarrollado modelos pronósticos para diferentes patologías cardíacas, el haber considerado como adecuado el resultado dado por la prueba de Hosmer-Lemeshow en el grupo de validación interna, cuando su nivel fue limitrofe (p= 0.0505). Incluso, se aconsejó no usar el modelo para informar al paciente y su familia del riesgo quirúrgico ni tampoco para monitorizar la calidad asistencial hospitalaria [4, 20, 21]. Desde su aparición y hasta la actualidad se han publicado diversos estudios de validación externa, unos con resultados desfavorables y otros con resultados favorables [22-27].

El presente estudio incluyó a 338 pacientes y cuyo objetivo fue evaluar el desempeño del modelo actualizado EuroSCORE II en nuestro hospital. La población evaluada tuvo las siguientes semejanzas y diferencias con la población de donde emergió el EuroSCORE II: nuestra población fue más joven 14.7 años (49.8±16.61 vs 64.6±12.5 años), hubo más mujeres (47.9% vs 30.9%), más diabéticos en control con insulina (12.4% vs 7.6%), más pacientes con falla renal en diálisis (5.91% vs 1.1%), más pacientes con pobre movilidad (7.39 vs 3.2%), más pacientes con endocarditis activa (7.4% vs 2.2%), más pacientes en estado preoperatorio crítico (14.5% vs 4.1%) y más pacientes con cirugía urgente (45.8% vs 18.5%). El modelo EuroSCORE II en su validación interna mostró buena discriminación, con área bajo la curva ROC de 0.8095 y buena calibración, con mortalidad esperada de 3.95 vs mortalidad observada de 4.18 En nuestra serie tuvimos una área bajo la curva de 0.806 (IC

95% 0.739-0.872), sugestiva de buena discriminación. La prueba Chi² de bondad de ajustes de Hosmer-Lemeshow fue de 13.22, con p de 0.104 (no significativa), compatible con buena calibración. La mortalidad esperada fue de 4.1 y la observada de 10.9%

Sin embargo, en los últimos años se ha cuestionado el concepto de que un valor de Chi cuadrado de bondad de ajustes de Hosmer-Lemeshow estadísticamente no significativa ($p > 0.05$) traduzca necesariamente buena calibración. “La estadística de Hosmer-Lemeshow compara proporciones pero es ciertamente imperfecta” [21]. Un resultado estadísticamente no significativo ($p > 0.05$) expresa que no hay evidencia de una mala calibración pero no necesariamente traduce una buena calibración han señalado algunos investigadores en este campo [28]. Dado lo anterior, proponen que la prueba Chi cuadrado de bondad ajustes de Hosmer y Lemeshow debe ser reemplazada o complementada con la *razón de mortalidad ajustada al riesgo* (RAMR), que consiste en obtener la razón de la mortalidad observada entre la mortalidad esperada ($\text{RAMR} = \text{O/E}$). Una razón de 1.0 significa que el score o modelo de prueba predice la mortalidad perfectamente (se muere el mismo número de pacientes que los que pronosticó el modelo). Una $\text{RAMR} > 1.0$ significa que el modelo subestima la mortalidad mientras que una $\text{RAMR} < 1.0$ significa que el modelo predice en exceso la mortalidad (la sobrestima) [28]. En nuestro estudio, la RAMR fue de 2.65, compatible con una subestimación de la mortalidad por el modelo. Si bien este resultado es para el total de los deciles, cuando observamos la RAMR de decil a decil, podemos ver que en los deciles inferiores el modelo sobreestima y, en los deciles superiores infra-estima el riesgo, excepto en el decil 10, donde el RAMR es de 0.87, sugestivo de sólo discreta sobre-estimación del riesgo de mortalidad.

Otra visión que se le ha dado al RAMR es considerarla como una herramienta que nos permite valorar el rendimiento de las unidades de cirugía de corazón. Un resultado de 1.0 significa que la mortalidad observada es igual a la esperada y que el rendimiento de una unidad quirúrgica es igual al esperado de acuerdo a algún modelo predictivo utilizado, por ejemplo el EuroSCORE II. Una razón $\text{O/E} < 1$ significa que la mortalidad observada es menor a la esperada de acuerdo al pronóstico del modelo y que los resultados obtenidos en la unidad quirúrgica son mejores que los proporcionados por el modelo predictivo. En cambio una razón $\text{O/E} > 1$ sugiere deficiencias en el rendimiento de la unidad quirúrgica, con mortalidades peri-operatorias mayores a las pronosticadas por el modelo [4].

En nuestra serie, el gráfico de la mortalidad ajustada al riesgo por deciles, observamos que desde el decil 1 y hasta el decil 6, la RAMR se mantiene por debajo de 1.0, sugiriendo buen rendimiento pronóstico del modelo EuroSCORE II. Sin embargo en los deciles 7, 8 y 9 el número de fallecimientos observados son mayores que los pronosticados por el EuroSCORE II, sugiriendo que estos intervalos representan una área de oportunidad para mejorar el rendimiento quirúrgico del servicio de cirugía cardíaca del HRAEB. Finalmente el decil 10 muestra nuevamente que la mortalidad observada fue discretamente menor a la pronosticada por EuroSCORE II, con una RAMR de 0.87, lo que sugiere buen rendimiento del EuroSCORE II en pacientes del más alto riesgo.

XII. Conclusión

El modelo actualizado llamado EuroSCORE II evaluado en nuestra población mostró buen rendimiento general, con aceptable calibración global pero con sobre-estimación en los primeros deciles e infraestimación en los últimos deciles de riesgo, excepto en el decil 10, donde vuelve a sobre-estimar el riesgo. Por otra parte, la discriminación del modelo fue buena.

Nuestra población es más joven en casi 15 años comparada con la población de donde se obtuvo el modelo EuroSCORE II pero tiene mayor prevalencia de la mayoría de los factores de riesgo en comparación a la población de referencia ya señalada.

Podemos seguir usando el modelo EuroSCORE II en nuestro hospital considerando las limitaciones señaladas al inicio del presente párrafo.

Dado el tamaño de la población y los años en que fueron operados este grupo de pacientes estudiados, necesitamos aumentar el tamaño de muestra, actualizar la evaluación del rendimiento del modelo con pacientes operados en años recientes y/ o diseñar en un estudio prospectivo para así tener una evaluación más justa y adecuada del modelo EuroSCORE II en nuestro medio.

XIII. Bibliografía

- 1.- Bonaccorsi Héctor A. Cuantificación del riesgo en cirugía Cardiovascular. Parte II. En: Cuidados perioperatorios en cirugía cardiovascular. Ricardo Iglesias y Claudio Pensa. Sociedad Argentina de Cardiología. Editorial Médica Panamericana. Buenos Aires, Argentina, 2000. p 13-18.
- 2.- Cortina Romero JM, López Gude MJ, Pérez de la Sota E. Estratificación y análisis de riesgo quirúrgico. Capítulo 2. En: Riesgo y complicaciones en cirugía cardíaca. Otero, Rufilanchas, Balda. Editorial Médica Panamericana. Madrid, España. 2005.p 15-27.
- 3.- Méndez Gómez I. Análisis de datos categóricos. Especialidad en Métodos Estadísticos a Distancia. Centro de Investigación en Matemáticas (CIMAT). Unidad Aguascalientes. 2019-2020.
- 4.- García AJ. Validación de EuroSCORE II en un Centro de Medio Volumen [Doctorado]. Universidad Miguel Hernández; Alicante, España; 2015
- 5.- Lemeshow S and Hosmer D. A review of goodness of fit statistic for use in the development of logistic regression models. Am J Epidemiol. 1982;115:[1]:92-106
- 6.- Iglesias CT. Métodos en Bondad de Ajuste en Regresión Logística. [Máster Oficial en Estadística Aplicada] Universidad de Granada, España; 2013.
- 7.- Cerda J, Cifuentes L. Uso de curvas ROC en investigación Clínica. Aspectos teorico-prácticos. Rev Chil Infect 2012;29(2):138-141
- 8.- Roques F, Nashef SAM, Michel P, Gauducheau E, de Vincentiis 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 :816-823.
- 9.- Nashef SAM, Roques F, Michel P, Gauducheau 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.
- 10.- Roques F, Michel P, Goldstone A, Nashef SAM. The logistic EuroSCORE. Eur Heart J 2003;24: 881-2
- 11.- Gogbashian A, et al.EuroSCORE: a systematic review of international performance. Eur J Cardiothorac Surg 2004;25:695-700

- 12.- Bhatti F, Grayson AD, Grotte G, Fabri bm, Au J, Jobes M, Bridgewater B on behalf of North West Quality Improvement Programme in Cardiac Intervention. The logistic EuroSCORE in cardiac surgery: how well does it predict operative risk? *Heart* 2006;92:1817-1820.
- 13.- 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.
- 14.- Parolari A, Pesce LL, Trezzi M, Loardi C, Kassem 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.
- 15.- Qadir I, Perveen S, Furnaz S, Shahabuddin S, Sharif H. Risk stratification analysis of operative mortality in isolated coronary artery bypass graft patients in Pakistan: comparison between additive and logistic EuroSCORE models. *Interact Cardiovasc Thorac Surg* 2011;13:137–41.
- 16.- Lebreton G, Merle S, Inamo J, Hennequin JL, Sanchez B, Rilos Z et al. Limitations in the inter-observer reliability of EuroSCORE: what should change in EuroSCORE II? *Eur J Cardiothorac Surg* 2011;40:1304–8.
- 17.- Shih HH, Kang PL, Pan JY, Wu TH, Wu CT, Lin CY et al. Performance of European System for Cardiac Operative Risk Evaluation in Veterans General Hospital Kaohsiung cardiac surgery. *J Chin Med Assoc* 2011;74: 115–20.
- 18.- Akar AR, Kurtcephe M, Sener E, Alhan C, Durdu S, Kunt AG et al.; The working Group for the Turkish Society of Cardiovascular Surgery and Turkish Ministry of Health. Validation of the EuroSCORE risk models in Turkish adult cardiac surgical population. *Eur J Cardiothorac Surg* 2011;40:730–5.
- 19.- 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.
- 20.- Samer A.M. Nashef, François Roques, Linda D. Sharples, Johan Nilsson, Christopher Smith, Antony R. Goldstone and Ulf Lockowandt. EuroSCORE II. *Eur J Cardiothorac Surg* 2012; 41:734–745
- 21.- Sergeant P, Meuris B, Pettinari M. EuroSCORE II, *illum qui est gravitates magni observe* *Eur J CardioThorac Surg* 2012;41:729–731
- 22.- Di Deddaa U, Pelisserob G, Agnellia B, De Vincentiisc C, Castelvecchioa S, Ranuccia M. Accuracy, calibration and clinical performance of the new EuroSCORE II risk stratification system. *Eur J CardioThorac Surg* 2013;43:27–32

- 23.- Zhang GX, Wang C, Wang L, Lu FL, Li BL, Han L, et al. Validation of EuroSCORE II in Chinese patients undergoing heart valve surgery. *Heart Lung Circ.* 2013;22:606-11
- 24.- Biancari F, Vasques F, Mikkola R, Martin M, Lahtinen J, Heikkinen J. Validation of EuroSCORE in patients undergoing coronary artery bypass surgery. *Ann Thorac Surg* 2012;93:1930-5
- 25.- Borracci R, Rubio M, Celano L, Ingino C, Allende N, Ahuad R. Prospective validation of EuroSCORE II in patients undergoing cardiac surgery in Argentinean centres. *Interact CardioVasc Thorac Surg* 2014;18: 539-543.
- 26.- Ayse GK, Kurtcephe M, Hidiroglu M, Cetin L, Kucuker A, Bacuy V et al. Comparison of original EuroSCORE, EuroSCORE II and STS risk models in a Turkish cardiac surgical cohort. *Interact CardioVasc Thorac Surg* 2013;16: 625-629.
- 27.- Barili F, Pacini D, Capo A, Rasovic O, Grossi C, Alamanni F et al. Does EuroSCORE II perform better than its original versions? A multicenter validation study. *Eur Heart J* 2013;34:22-29
- 28.- Nezic D, Borzanovic M, Spasic T and Vukovic P. Calibration of the EuroSCORE II risk stratification model: Is the Hosmer-Lemeshow test acceptable any more? *Eur J CardioThorac Surg* 2013;43:206
- 29.- Teniente R, González FE, Chagolla MA, Acevedo I, Bernal EA, Medina OS, López LR, Romo RE, Landeros JE. Validation of the EuroSCORE model in patients undergoing heart surgery in Regional Hospital of High Speciality of Bajío. *Rev Mex Cardiol* 2018; 29 (3): 134-143