

Condiciones de aplicación de modelos de riesgo en cirugía cardíaca

José M. Cortina Romero

Servicio de Cirugía Cardíaca. Hospital 12 de Octubre. Madrid. España.

Pocos procedimientos terapéuticos han sido tan investigados, evaluados y criticados como la cirugía cardíaca. Especialmente la cirugía de revascularización coronaria. Entre otros muchos, se han evaluado aspectos en relación con los resultados y también con la adecuación de las indicaciones desde múltiples puntos de vista.

Las iniciativas para este tipo de valoraciones han surgido desde ámbitos diversos: cirujanos cardíacos, cardiólogos, epidemiólogos, gestores sanitarios, proveedores de servicios, prensa, etc. Es de las pocas especialidades en que se plantean auditorías de resultados cuando se disparan las señales de alarma.

Cabría preguntarse cuál es la razón de que la cirugía cardíaca haya sido objeto de un análisis tan exhaustivo en comparación con otro tipo de recursos terapéuticos. No es el objetivo de este comentario entrar en esta disquisición, pero una de las razones de esta dinámica probablemente sea la propia actitud de los cirujanos cardíacos dirigida a la evaluación de sus propias actuaciones.

En este orden de cosas, los grupos de cirugía cardíaca han introducido, considero que de forma pionera, sistemas predictivos que permiten la estimación de los riesgos quirúrgicos de los procedimientos que se realizan. La introducción de esta metodología ha generado una cultura del «riesgo» y de su estimación. Actualmente es frecuente que los cirujanos cardíacos y también los cardiólogos manejen en el proceso de toma de decisiones con sus pacientes quirúrgicos parámetros como el riesgo según EuroSCORE, Parsonnet o STS¹⁻⁵. Las guías de práctica clínica de la AHA/ACC de 2004⁶ consideran que el uso de sistemas de predicción para

realizar estimaciones preoperatorias del riesgo debe servir a médicos y enfermos para ponderar el balance riesgo-beneficio del procedimiento quirúrgico que se plantea (recomendación de clase IIa, nivel de evidencia C).

Aparentemente, no hay una recomendación equivalente cuando se trata de otro procedimiento terapéutico, como es la revascularización coronaria mediante actuaciones endoluminales. Y esto no ocurre por la ausencia de modelos predictivos adecuados y extensamente validados para actuaciones coronarias percutáneas^{7,8}. En mi opinión, sería deseable que se extendiese su uso a otros procedimientos terapéuticos, y no sólo en el campo de las enfermedades cardíacas, pero es evidente que todavía no existe la cultura necesaria para la aplicación y generalización de esta metodología.

Posiblemente sea pretencioso querer establecer aquí recomendaciones para el uso de este tipo de herramientas. Pero, si analizamos la trascendencia de las conclusiones que pueden llegar a derivarse de su aplicación, estaremos de acuerdo en que se deben seguir unas condiciones estrictas de aplicación, que garanticen la obtención de unas conclusiones válidas. En caso contrario, se puede caer en interpretaciones erróneas que, en lugar de clarificar cuestiones esenciales, añaden más confusión.

Este tipo de condiciones deben ser especialmente más estrictas cuando el uso que se da a los sistemas predictivos es el de herramienta de control de la calidad de la asistencia.

Si consideramos el ámbito de las poblaciones a las que se aplica este tipo de modelos predictivos, podemos distinguir tres estratos: estimaciones individuales, estimaciones en grupos específicos de enfermedades y estimaciones en series completas de la actividad de los grupos quirúrgicos.

Aunque el uso en los primeros dos niveles es aceptable, y más adelante se comentan estos puntos, hay que recordar que el desarrollo de estas herramientas estadísticas está dirigido a la aplicación a series de actividad completas.

El objetivo último de la construcción de este tipo de modelos siempre ha sido disponer de una herramienta que permita hacer evaluaciones y comparaciones objetivas de resultados y del *case-mix* de poblaciones ope-

VÉASE ARTÍCULO EN PÁGS. 589-94

Correspondencia: Dr. J.M. Cortina Romero.
Servicio de Cirugía Cardíaca. Hospital 12 de Octubre.
Avda. Córdoba, s/n. 28041 Madrid. España.
Correo electrónico: jcortina.hdoc@salud.madrid.org

Full English text available from: www.revespcardiol.org

radas de cirugía cardíaca. Éste es un requisito derivado de la filosofía del análisis de la calidad asistencial. Conviene recordar que la hipótesis al uso sobre la evaluación de la calidad infiere que, si hacemos que las poblaciones tratadas sean comparables, las diferencias significativas en los resultados se deberían a diferencias de calidad asistencial. Sobre esta manera de hacer comparaciones se podría objetar que las diferencias sobre aspectos estructurales que existen en los diferentes grupos no están ajustadas. Lamentablemente, en la actualidad no hay modelos que ajusten por esta importante dimensión.

La publicación del trabajo «Validación del modelo probabilístico EuroSCORE en pacientes intervenidos de injerto coronario»⁹ suscita diversos aspectos en relación con la aplicación de este tipo de herramientas. Como ya se ha comentado, el uso de *scores* o modelos predictivos se ha convertido en una rutina en muchos servicios de cirugía cardíaca. Sin embargo, posiblemente todavía existan suficientes elementos de confusión como para considerar que su uso todavía no está adecuadamente estandarizado. Con el deseo de intentar clarificar algunos aspectos puntuales sobre el uso de esta metodología, me gustaría comentar los siguientes aspectos:

- Poblaciones a las que se aplican los modelos.
- Qué modelo es el adecuado: EuroSCORE logístico frente a EuroSCORE aditivo.
- Aplicación sobre bases de datos administrativas frente a bases de datos clínicas.
- Significado de la validación de estos métodos.
- Interpretación de los resultados que se obtienen al aplicar los modelos.

Aunque las siguientes comentarios pueden ser de aplicación a cualquiera de los modelos predictivos disponibles, en general nos vamos a referir al EuroSCORE porque es el más ampliamente utilizado en nuestro entorno.

Poblaciones a las que se aplican los modelos

En la actualidad estos sistemas se están utilizando para estimación individual, estimación de riesgos de determinada enfermedad y estimación de riesgos de series completas de actividad quirúrgica.

Hay que recordar que estos modelos se construyen habitualmente mediante regresión logística identificando los factores de más significación frente a la variable dependiente, que en este caso es mortalidad hospitalaria. Aunque la aplicación de análisis multivariable a grandes poblaciones identifica muchos factores relacionados con la muerte hospitalaria, a la hora de construir estos modelos se suele utilizar un número limitado de ellos. Esto es así por varias razones, entre otras porque la capacidad de calibración y discrimina-

ción no suele mejorar a partir de un número determinado de variables y que deben cumplir la condición de ser sencillos de utilizar. Además hay otras razones puramente metodológicas. Harrel et al¹⁰ concluyen que, para modelos basados en regresión logística, los eventos (muertes) por variable que se considere incluir en el modelo deben ser por lo menos 10 veces el número de variables pronósticas que podrían incluirse. Como el evento habitual en estos sistemas es muerte, aunque las poblaciones de las que se derivan sean grandes, pocos factores cumplen este criterio.

En cuanto al uso para estimaciones individuales, dado el limitado número de variables de EuroSCORE, su uso sólo se puede considerar orientativo. Máxime si el paciente en cuestión presenta alguna variable no considerada en el modelo y que sin embargo esté claramente identificada como un factor de riesgo de muerte. No obstante, para la mayoría de los enfermos habituales en cirugía cardíaca, como los coronarios aislados, los valvulares aislados o su combinación, EuroSCORE logístico es útil para identificar la probabilidad de riesgo de un caso determinado. Probablemente, para el uso individual lo deseable sería un modelo que considere muchas variables. En la actualidad el único que incluye un elevado número de variables es el de la STS. Conviene recordar que indicar o rechazar una cirugía con base en una estimación de riesgo individual alta posiblemente sea inadecuado.

La aplicación del modelo a grupos de enfermedades concretas, como coronarios, valvulares, etc., suscita otro problema distinto de los del aspecto individual. Han surgido múltiples series que vienen a concluir en la necesidad de disponer modelos específicos para cada grupo de afecciones, preferentemente para cirugía valvular¹¹⁻¹⁴. Seguramente se pueda concluir que lo ideal es disponer de un modelo para cada tipo de enfermo, pero habría que exigir a estos modelos específicos que cumplan las debidas condiciones durante su construcción.

En todo caso, lo que debe estar claro es a qué población se está aplicando. El artículo objeto de este comentario es un ejemplo de lo contrario. El título induce a pensar que se está estudiando a una población de enfermos coronarios aislados, pero es obvio que no es así, como lo demuestra el hecho de que haya 4 casos con endocarditis y que a 100 de los pacientes se les realizó una cirugía asociada a la coronaria.

La aplicación sobre series completas de actividad en cirugía cardíaca, en principio, es el uso más razonable y el que originó la creación del EuroSCORE. El objetivo esencial de este uso es disponer de una herramienta estandarizada para el control de la calidad. La población a la que se debe aplicar sería cualquier paciente intervenido de cirugía cardíaca mediante circulación extracorpórea. Únicamente parece razonable añadir a este grupo los pacientes intervenidos de cirugía coronaria sin circulación extracorpórea, ya que el

uso en este grupo está extensamente validado^{15,16}. Cuando se utiliza con esta filosofía, se deben cumplir con rigor varias condiciones: no excluir a ningún paciente al que se le pueda aplicar, no permitir que haya pérdida de información en los factores que se computan y, sobre todo, que no haya pérdidas en el evento muerte.

Qué modelo es el adecuado para usar: EuroSCORE logístico frente a EuroSCORE aditivo

Como ya he comentado, fundamentalmente estas herramientas deben servir para evaluar la calidad¹⁷, por lo que se deben aplicar a todos los enfermos. Para este propósito, y aunque hay muchos modelos disponibles, al día de hoy creo que se debe seguir utilizando EuroSCORE. Y esto por múltiples razones: está extensamente validado en múltiples entornos^{18,19}, España participó con un 10% de la población que lo generó, es sencillo de utilizar y permite incluir a la población habitual de los enfermos que se operan de cirugía cardíaca. En la actualidad, los mismos autores tienen la intención de desarrollar un nuevo EuroSCORE siguiendo la misma metodología que en su día se utilizó.

Con la utilización de EuroSCORE hay cierto grado de confusión en cuanto a la utilización del modelo logístico frente al aditivo. Conviene recordar que ambos modelos son iguales y que el modelo aditivo es una simplificación del logístico, y no una variante, en la que se hace una asimilación y simplificación de los coeficientes beta de la ecuación de regresión logística, de modo que sea más fácil de usar. La estimación de la probabilidad de muerte por la suma simple de los *scores* es una fuente de error, sobre todo con puntuaciones altas. La asignación de una probabilidad de muerte a una puntuación determinada debería realizarse utilizando la tabla 1¹.

El interesante ejercicio de comparar EuroSCORE logístico con aditivo ya se ha realizado²⁰, pero en realidad aporta cierta confusión, y las conclusiones que se obtienen se entienden fácilmente observando la tabla. Además, dejando de lado el tema de la calibración de

los segmentos de alto riesgo, que suelen ser pocos pacientes en la experiencia global de un grupo, la capacidad discriminación es exactamente la misma con uno que con otro.

En resumen, lo ideal es la utilización del modelo logístico para evitar la confusión. Si se emplea el modelo aditivo, sobre todo para estimaciones individuales de probabilidad de muerte, se debe tener en cuenta la probabilidad de muerte que se asocia a cada segmento de puntuación.

Utilización en bases de datos administrativas o bases de datos clínicas

Otro aspecto de interés se refiere a la aplicación de modelos predictivos a bases de datos administrativas o a bases de datos clínicas. La mayoría de los modelos existentes se han desarrollado sobre las últimas. Sin embargo, cada vez es más frecuente la aplicación a bases de datos administrativas. Por el momento estas bases todavía no permiten una aplicación rigurosa de estas herramientas, por lo que las conclusiones que se obtengan deben ser cuidadosamente observadas.

Aunque los responsables de ese tipo de base de datos suelen postular un alto nivel de fiabilidad, la realidad es que, de momento, posiblemente no sea suficiente. Las razones son múltiples y entre ellas cabe destacar:

- Las bases administrativas suelen recoger la información retrospectivamente, lo que implica indefectiblemente un nivel de pérdida de datos que obliga a presunciones que pueden ser aceptables o no^{21,22}. Por supuesto, lo ideal es una recogida prospectiva de los datos.

- Su uso suele conllevar errores de clasificación de pacientes que pueden hacer que se pierda la perspectiva de qué tipo de patología se está analizando^{23,24}.

- La ausencia de *end-points* no estandarizados es otra fuente de error^{25,26}. En este caso habría que asumir la definición de mortalidad hospitalaria, que se define como la que ocurre durante la misma estancia o durante los primeros 30 días postoperatorios. No es admisible estipular que la mortalidad fuera de la estancia sea mínima sólo por creencias subjetivas.

Por supuesto, lo ideal sería que no existiese ningún tipo discrepancia según se use una base de datos u otra. Pero más deseable sería todavía que sólo existiera una herramienta de gestión de la información, que pudiera ser compartida y que, a la hora de su explotación, se utilizasen criterios estandarizados y compartidos.

En todo caso, con independencia de la herramienta que se utilice, lo que dará valor a las conclusiones que se obtengan será la auditoría externa de los datos y de su sistema de gestión.

TABLA 1. Probabilidad de muerte según EuroSCORE aditivo

| Score | Mortalidad (%) | IC del 95% |
|---------------|----------------|------------|
| 0, 1, 2 | 0,8 | 0,6-1,1 |
| 3, 4, 5 | 3 | 2,6-3,5 |
| > 5 | 11,2 | 10,2-12,2 |
| Subgrupos > 5 | | |
| 6, 7, 8 | 7,4 | 6,5-8,4 |
| 9, 10, 11 | 16,4 | 14-19,1 |
| > 11 | 36,5 | 31-42,3 |

Significado de la validación de estas herramientas

Posiblemente la palabra validación se haya utilizado incorrectamente en muchas ocasiones a la hora de publicar el uso de este tipo de sistemas. Es claramente equívoco pretender validar un modelo simplemente por el hecho de usarlo en unas circunstancias determinadas. La correcta validación de un modelo probabilístico presenta una serie de condiciones que pocas veces se dan cuando se aplican²⁷. Desde el punto de vista estadístico, la validación de un modelo pronóstico significa documentar que funciona correctamente en una población distinta de la que se derivó.

La primera condición para pretender la validación de un modelo es la del tamaño muestral. Algunos autores²⁸ consideran que se requieren al menos 100 muertes para la validación del modelo. Esto implica que, para que la mortalidad sea en torno al 5% se requiere una muestra de unos 2.000 pacientes.

Además del tamaño, se requiere la aplicación estricta del *score* a todos y cada uno de los pacientes correspondientes, sin presunciones sobre factores perdidos, y lo que es más importante sin pérdidas en el evento que sea la variable dependiente.

Finalmente, cuando un sistema ha sido tan extensamente validado como EuroSCORE, existen dos riesgos si la conclusión que se obtiene es que no lo hemos validado. Se puede concluir que nuestra metodología de aplicación es incorrecta o, lo que sería peor, que nuestro funcionamiento en términos de calidad no es bueno.

Interpretación de los resultados que se obtienen al aplicar los modelos

Si el modelo que utilizamos ha sido adecuadamente validado y lo aplicamos correctamente, la interpretación sólo puede tener que ver con la calidad de la asistencia prestada, con la salvedad anteriormente expuesta de la ausencia de corrección para diferencias de estructura.

Diferencias significativas por encima o por debajo respecto a los resultados esperados por el modelo que usemos sólo se pueden interpretar como un funcionamiento incorrecto o correcto.

No solamente es lícito, sino también obligatorio, comparar estadísticamente la mortalidad observada con el promedio de las observadas. Estos modelos se desarrollaron específicamente para esto, como así se están utilizando²⁹, y eludir realizar este tipo de comparación como proponen Lafuente et al en el artículo citado implica alejarse de la esencia de este tipo de metodología.

Desde mi punto de vista sería recomendable y conveniente que las sociedades científicas establecieran las coordenadas y recomendaciones de uso de estas

herramientas de la forma más precisa posible. Y además, que los grupos que realizan procedimientos susceptibles de este análisis incorporen esta metodología a su práctica diaria. Es evidente que el uso inapropiado o incorrecto de estas herramientas puede llevar a conclusiones erróneas, aspecto especialmente sensible cuando de lo que se trata es de evaluar la calidad.

BIBLIOGRAFÍA

- Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg.* 1999;16:9-13.
- Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;79:13-12.
- Bernstein AD, Parsonnet V. Bedside estimation of risk as an aid for decision-making in cardiac surgery. *Ann Thorac Surg.* 2000;69:823-8.
- Shroyer AL, Grover FL, Edwards FH. 1995 coronary artery bypass risk model: The Society of Thoracic Surgeons Adult Cardiac National Database. *Ann Thorac Surg.* 1998;65:879-84.
- Shroyer AL, Plomondon ME, Grover FL, Edwards FH. The 1996 coronary artery bypass risk model: the Society of Thoracic Surgeons Adult Cardiac National Database. *Ann Thorac Surg.* 1999;67:1205-8.
- Eagle KA, Guyton RA, Davidoff R, Edwards FH, Ewy GA, Gardner TJ, et al. ACC/AHA 2004 guideline update for coronary artery bypass graft surgery: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1999 Guidelines for Coronary Artery Bypass Graft Surgery). *Circulation* 2004;110:e340-437.
- Kunadian B, Dunning J, Das R, Roberts AP, Turley AJ, Morley R, et al. External validation of established risk adjustment models for procedural complications after percutaneous coronary intervention. 1. *Heart.* 2007 Nov 21 [Epub ahead of print].
- Moscucci M, O'Connor GT, Ellis SG, Malenka DJ, Sievers J, Bates ER, et al. Validation of risk adjustment models for in-hospital percutaneous transluminal coronary angioplasty mortality on an independent data set. *J Am Coll Cardiol.* 1999;34:692-7.
- Lafuente S, Trilla A, Bruni L, González R, Bertrán MJ, Pomar JL, et al. Validación del modelo probabilístico EuroSCORE en pacientes intervenidos de injerto coronario. *Rev Esp Cardiol.* 2008;61:589-94.
- Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic modelling. *Stat Med.* 1984;3:143-52.
- Ambler G, Omar RZ, Royston P, Kinsman R, Keogh BE, Taylor KM. Generic, simple risk stratification model for heart valve surgery. *Circulation.* 2005;112:224-31.
- Grossi EA, Schwartz CF, Yu PJ, Jorde UP, Croke GA, Grau JB, et al. High-risk aortic valve replacement: are the outcomes as bad as predicted? *Ann Thorac Surg.* 2008;85:102-6.
- Jin R, Grunkemeier GL, Starr A. Validation and refinement of mortality risk models for heart valve surgery. *Ann Thorac Surg.* 2005;80:471-9.
- Van Gameren M, Kappetein AP, Steyerberg EW, Venema AC, Berenschot EA, Hannan EL, et al. Do we need separate risk stratification models for hospital mortality after heart valve surgery? *Ann Thorac Surg.* 2008;85:921-30.
- Vázquez Roque FJ, Fernández TR, Pita S, Cuenca JJ, Herrera JM, Campos V, et al. Evaluación preoperatoria del riesgo en la cirugía coronaria sin circulación extracorpórea. *Rev Esp Cardiol.* 2005;58:1302-9.

16. Wu Y, Grunkemeier GL, Handy JR Jr. Coronary artery bypass grafting: are risk models developed from on-pump surgery valid for off-pump surgery? *J Thorac Cardiovasc Surg.* 2004;127:174-8.
17. Díaz de Tuesta I, Cuenca J, Fresneda PC, Calleja M, Llorens R, Aldámiz G, et al. No hay relación entre el volumen quirúrgico y la mortalidad en los servicios de cirugía cardíaca en España. *Rev Esp Cardiol.* 2008;61:276-82.
18. Nashef SA, Roques F, Hammill BG, Peterson ED, Michel P, Grover FL, et al. Validation of European System for Cardiac Operative Risk Evaluation (EuroSCORE) in North American cardiac surgery. *Eur J Cardiothorac Surg.* 2002;22:101-5.
19. Roques F, Nashef SA, Michel P, Pinna PP, David M, Baudet E, et al. Does EuroSCORE work in individual European countries? *Eur J Cardiothorac Surg.* 2000;18:27-30.
20. Michel P, Roques F, Nashef SA. Logistic or additive EuroSCORE for high-risk patients? *Eur J Cardiothorac Surg.* 2003;23:684-7.
21. Herbert MA, Prince SL, Williams JL, Magee MJ, Mack MJ. Are unaudited records from an outcomes registry database accurate? *Ann Thorac Surg.* 2004;77:1960-4.
22. Mack MJ, Herbert M, Prince S, Dewey TM, Magee MJ, Edgerton JR. Does reporting of coronary artery bypass grafting from administrative databases accurately reflect actual clinical outcomes? *J Thorac Cardiovasc Surg.* 2005;129:1309-17.
23. Glance LG, Dick AW, Osler TM, Mukamel DB. Accuracy of hospital report cards based on administrative data. *Health Serv Res.* 2006;41:1413-37.
24. Shahian DM, Silverstein T, Lovett AF, Wolf RE, Normand SL. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. *Circulation.* 2007;115:1518-27.
25. Hannan EL, Racz MJ, Jollis JG, Peterson ED. Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? *Health Serv Res.* 1997;31:659-78.
26. Parker JP, Li Z, Damberg CL, Danielsen B, Carlisle DM. Administrative versus clinical data for coronary artery bypass graft surgery report cards: the view from California. *Med Care.* 2006;44:687-95.
27. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med.* 2000;19:453-73.
28. Harrell FE. *Regression models strategies.* New York: Springer-Verlag; 2001.
29. Ribera A, Ferreira-González I, Cascant P, Pons JM, Permanyer-Miralda G. Evaluación de la mortalidad hospitalaria ajustada al riesgo de la cirugía coronaria en la sanidad pública catalana. Influencia del tipo de gestión del centro (estudio ARCA). *Rev Esp Cardiol.* 2006;59:431-40.