



Acta Gastroenterológica Latinoamericana
ISSN: 0300-9033
ISSN: 2429-1119
actasage@gmail.com
Sociedad Argentina de Gastroenterología
Argentina

¿Cómo utilizamos los puntajes de riesgo?

Lucas, Luciano Oscar

¿Cómo utilizamos los puntajes de riesgo?

Acta Gastroenterológica Latinoamericana, vol. 51, núm. 3, 2021

Sociedad Argentina de Gastroenterología, Argentina

Disponible en: <https://www.redalyc.org/articulo.oa?id=199368908001>

DOI: <https://doi.org/10.52787/kidl6065>

2021 Acta Gastroenterológica Latinoamericana. Este es un artículo de acceso abierto publicado bajo los términos de la Licencia Creative Commons Attribution (CC BY-NC-SA 4.0), la cual permite el uso, la distribución y la reproducción de forma no comercial, siempre que se cite al autor y la fuente original.

2021 Acta Gastroenterológica Latinoamericana. Este es un artículo de acceso abierto publicado bajo los términos de la Licencia Creative Commons Attribution (CC BY-NC-SA 4.0), la cual permite el uso, la distribución y la reproducción de forma no comercial, siempre que se cite al autor y la fuente original.




Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

¿Cómo utilizamos los puntajes de riesgo?

Luciano Oscar Lucas luciano.lucas@hospitalitaliano.org.ar

Hospital Italiano de Buenos Aires, Argentina

 <https://orcid.org/0000-0001-5816-644X>

Acta Gastroenterológica
Latinoamericana, vol. 51, núm. 3, 2021

Sociedad Argentina de Gastroenterología,
Argentina

Recepción: 06 Julio 2021
Aprobación: 21 Agosto 2021
Publicación: 27 Septiembre 2021

DOI: <https://doi.org/10.52787/kidl6065>

Redalyc: <https://www.redalyc.org/articulo.oa?id=199368908001>

Román Conroy comienza el capítulo "From epidemiological risk to clinical practice by way of statistics" del libro *Therapeutic Strategies in Cardiovascular Risk* con una frase de Tavia Gordon: "The power and elegance of the logistic function make it an attractive and elegant statistical instrument, but in the end we cannot push a button and hope that everything will come out all right. Because frequently it will not".¹

En la medicina actual los puntajes de riesgo tienen una amplia difusión. Los utilizamos diagnosticar y estratificar el riesgo en la internación y nos orientan a la hora de tomar conductas y de evaluar el riesgo al alta y en el largo plazo. Dado que su aplicación está muy extendida y que se usan con mucha frecuencia, es crucial que sean poco complejos, fáciles de usar y que su utilización no demande mucho tiempo. Ahora bien, una vez que encontramos un puntaje de riesgo que cumple con estos atributos fundamentales, surgen otras preguntas: ¿cumple con las leyes a las que están sujetos los puntajes?, ¿funciona?, ¿siempre y en cualquier situación? ¿Por qué este y no otro?

Los puntajes funcionan de igual manera que un método de diagnóstico común (como pueden ser la troponina para el síndrome coronario agudo o el NT proBNP para la insuficiencia cardíaca) y se les aplican las mismas leyes de sensibilidad y especificidad. A su vez, tienen algunas particularidades a la hora de ser generados y de evaluar su funcionamiento.

En forma simplificada, y sin entrar en vericuetos estadísticos complejos, los dos requisitos más relevantes que deben cumplir son: en primer lugar, que el punto final que prediga sea claro, estandarizado y fácilmente replicable. Una cosa es la muerte como punto final, y otra la disnea. Este último no parece ser el mejor punto final a la hora de generar un puntaje de riesgo, puesto que el diagnóstico se presta a subjetividad en muchos casos.

En segundo lugar, debe surgir de una muestra representativa de la población sobre la cual va a aplicarse.² A este respecto podemos utilizar dos de ellos a modo de ejemplo. El puntaje de Framingham deriva de una cohorte de 5345 personas oriundas de esa comunidad que fueron seguidas hasta por doce años, y el punto final pautado estaba compuesto de muerte atribuida a enfermedad coronaria, infarto de miocardio, angina o "insuficiencia" coronaria.³ Mientras tanto, más acá en el tiempo, el grupo del Proyecto SCORE utilizó una cohorte de 200.000 individuos de once

países europeos, seguidos hasta por trece años, y el punto final evaluado fue la muerte cardiovascular. En su generación se utilizaron también factores genéticos y medioambientales subrogantes de las diferentes regiones geográficas en la función.⁴ A simple vista, parecería que el segundo requisito fue diseñado a partir de una muestra más representativa de la población sobre la cual iba a ser aplicado y que, además, su punto final era bien concreto. Aun considerando estas diferencias trascendentes, en su editorial sobre la publicación del puntaje SCORE, Topol y col. critican el punto final utilizado en ambos puntajes. En el primer caso por ambiguo y sensible a sesgos y en el segundo por insuficiente.⁵ Esto demuestra lo difícil que puede ser encontrar los puntos finales adecuados.

Una vez generado, la forma correcta de evaluar el desempeño de los puntajes de riesgo es mediante la medición de tres características: la eficacia de discriminación, la calibración y la capacidad de reclasificación que se da ante el agregado de variables al modelo original.

La primera de las cualidades, la eficacia de discriminación, es la idoneidad de la función para separar a aquellos que poseen una alta probabilidad de presentar el punto final evaluado de aquellos que no la poseen. Esto puede evaluarse dentro de la misma población que utilizamos para generar la función, a partir de una parte de la muestra que se reserva antes de la confección del puntaje a los fines de su validación: a esto llamaremos *validez interna*. También existe una validez externa, que es la más importante y surge de aplicar la función a otras poblaciones y evaluar su capacidad discriminativa.

Para entender cuán tan trascendente es que la población sobre la cual aplicamos el puntaje sea similar a aquella a partir de la cual se generó, podemos usar de ejemplo la edad. A la hora de aplicar los puntajes de riesgo a pacientes en rangos etarios extremos –por ejemplo, mayores de 65 años–, debemos tener cuidado, porque si estas edades no están bien representadas en la muestra a partir de la cual se generó el puntaje, podría suceder que el rendimiento de la función no sea del todo bueno si lo aplicamos a una población que en su mayoría es de edad igual o mayor. A su vez, estos modelos plantean un coeficiente beta idéntico para todos los rangos etarios y puede ser que en la práctica esto no ocurra. Incluso puede suceder que otros factores afecten la función de diferente manera según la edad. Esto podría corregirse, por lo menos parcialmente, generando un factor de interacción entre la edad y los demás factores incluidos en la función.⁹

La calibración, en cambio, es una medida de cuán fidedigna es la predicción. Es decir, ¿cuántos de los que se predijo que iban a presentar el punto final realmente lo presentaron? Tanto el lugar como el momento en los que se aplica la fórmula afectarán la calibración. Por ejemplo, si la aplicamos en un lugar que presenta una mayor prevalencia del evento que el lugar en donde fue generado el puntaje, habrá una tendencia a la infraestimación. Ocurrirá lo contrario si lo aplicamos en un lugar donde la incidencia del punto final es menor: en este caso habrá una tendencia a la sobreestimación. La mejor calibración se obtiene al aplicar la fórmula en una población de características similares a aquella a partir de la cual fue

generada.^{2,6} Esto es así porque la probabilidad previa, al aplicar cualquier tipo de test o puntaje, es trascendente en el rendimiento que este tendrá. ¿Será necesario, entonces, hacer algo con la fórmula para que calibre mejor en mi población? La respuesta es que muchas veces sí. En efecto, así lo han hecho con el puntaje de riesgo de Framingham países como China⁷ y el Reino Unido.⁸

Por último, está la reclasificación, que es una medida del porcentaje de individuos que presentaron o no presentaron un evento y que fueron correctamente reclasificados a una nueva categoría luego de añadirse alguna variable de riesgo a la fórmula.² Es un concepto más nuevo que los anteriores, pero se encuentra muy en boga actualmente.

A modo de ejemplo conceptual, es cuestión de tiempo para que la genética nos muestre sustratos de riesgo para la enfermedad cardiovascular, como lo ha hecho con otras patologías. Hoy en día no encontramos ningún puntaje de riesgo que considere de manera directa variables genéticas en su fórmula. Es posible pensar que en el futuro el genotipo tendrá tanta o más relevancia que el fenotipo, en el cual nos basamos hoy para estratificar el riesgo de nuestros pacientes. Cuando la ciencia halle estas variables y se las incorpore a los modelos existentes, la cuantificación de la capacidad de reclasificación será trascendente para medir el aporte de estas variables a los diversos puntajes.

En conclusión, los puntajes de riesgo, sea cual fuere, están sujetos a las mismas leyes que los demás estudios o métodos diagnósticos. Los debemos utilizar teniendo siempre en cuenta la probabilidad previa de la población sobre la que estamos aplicándolo, puesto que el rédito que le extraeremos a la estimación depende de esa población. Incluso cuando la población posea características similares, la distribución etaria dentro de ella debe ser tomada en cuenta, puesto que, al ser un predictor de los más importantes, si el rango etario está poco representado la función no va a rendir adecuadamente. La capacidad discriminativa, la calibración y el poder de reclasificación son las tres características que debemos evaluar en un puntaje de riesgo.

Los puntajes, si bien son una herramienta, nunca podrán reemplazar el criterio médico del profesional que lo está aplicando. La frase de Tavia Gordon con la que iniciamos el texto, quizá con algunos matices y complejidades mayores, continúa siendo cierta en nuestros días.

Referencias

1. Gordon T. Editorial: Hazards in the use of the logistic function with special reference to data from prospective cardiovascular studies. *J Chronic Dis.* 1974;27(3):97-102.
2. Cooney MT, Dudina AL, Graham IM. Value and limitations of existing scores for the assessment of cardiovascular risk: a review for clinicians. *J Am Coll Cardiol.* 2009;54(14):1209-27.
3. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation.* 1998;97(18):1837-47.

4. Conroy RM, Pyörälä K, Fitzgerald AP, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J*. 2003;24(11):987-1003.
5. Topol EJ, Lauer MS. The rudimentary phase of personalised medicine: coronary risk scores. *Lancet*. 2003;362(9398):1776-7.
6. Brindle P, Beswick A, Fahey T, Ebrahim S. Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. *Heart*. 2006;92(12):1752-9.
7. Liu J, Hong Y, D'Agostino RB Sr, et al. Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study. *JAMA*. 2004;291(21):2591-9.
8. Brindle P, May M, Gill P, et al. Primary prevention of cardiovascular disease: a web-based risk score for seven British black and minority ethnic groups. *Heart*. 2006;92(11):1595-602.
9. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ*. 2008;336(7659):1475-82.

Notas

Cite este artículo como: Lucas LO. ¿Cómo utilizamos los puntajes de riesgo? *Acta Gastroenterol Latinoam*. 2021;51(3):242-4. <https://doi.org/10.52787/kidl6065>

Notas de autor

Luciano Oscar Lucas. Correo electrónico:
luciano.lucas@hospitalitaliano.org.ar