

**FORMACIÓN MÉDICA****La bola de cristal. Lectura crítica de reglas de predicción clínica***Molina Arias M.**Hospital Infantil Universitario La Paz. Madrid.***Resumen**

Una regla de predicción clínica es una herramienta compuesta por un conjunto de variables de la historia clínica, exploración física y pruebas complementarias básicas, que nos proporciona una estimación de la probabilidad de un evento, nos sugiere un diagnóstico o nos predice una respuesta concreta a un tratamiento. Para conocer el valor de sus conclusiones será importante estudiar su validez interna, la importancia clínica de sus conclusiones y si son aplicables y de utilidad en nuestro entorno clínico.

Introducción

Una regla de predicción clínica es una herramienta compuesta por un conjunto de variables de la historia clínica, exploración física y pruebas complementarias básicas, que nos proporciona una estimación de la probabilidad de un evento, nos sugiere un diagnóstico o nos predice una respuesta concreta a un tratamiento. Para conocer el valor de sus conclusiones será importante estudiar su validez interna, la importancia clínica de sus conclusiones y si son aplicables y de utilidad en nuestro entorno clínico.

¡Cómo me gustaría poder predecir el futuro! Y no solo para ganar millones en la lotería, que es lo primero que a uno se le ocurre. Hay cosas más

importantes en la vida que el dinero (o eso dicen algunos), decisiones que tomamos basadas en suposiciones que acaban por no cumplirse y que nos complican la vida hasta límites insospechados. Todos habremos pensado alguna vez aquello de “si se viviera dos veces...” No me cabe duda, si me encontrase con el genio de la lámpara uno de los tres deseos que le pediría sería una bola de cristal para ver el futuro.

Y también nos vendría bien en nuestro trabajo como médicos. En nuestro día a día nos vemos obligados a tomar decisiones sobre el diagnóstico o el pronóstico de nuestros pacientes y lo hacemos siempre sobre el pantanoso terreno de la incertidumbre, asumiendo siempre el riesgo de cometer algún error. Nosotros, sobre todo cuando vamos siendo más experimentados, estimamos de forma consciente o inconsciente la verosimilitud de nuestras suposiciones, lo que nos ayuda en la toma de decisiones diagnósticas o terapéuticas. Sin embargo, qué bueno sería también poder disponer de una bola de cristal para saber con más

exactitud el devenir del curso del paciente.

El problema, como ocurre con otros inventos que serían de gran utilidad en medicina (como la máquina del tiempo), es que nadie ha conseguido todavía fabricar una bola de cristal que funcione de verdad. Pero no nos vengamos abajo. No podremos saber con seguridad qué va a ocurrir, pero sí que podemos estimar la probabilidad de que ocurra determinado resultado.

Para ello, podemos utilizar todas aquellas variables relacionadas con el paciente que tengan una validez diagnóstica o pronóstica conocidas e integrarlas para realizar el cálculo de probabilidades. Pues bien, hacer semejante cosa sería lo mismo que diseñar y aplicar lo que se conoce como una regla de predicción clínica (RPC).

Así, si nos ponemos un poco formales, podremos definir una RPC como una herramienta compuesta por un conjunto de variables de la historia clínica, exploración física y pruebas complementarias básicas, que nos proporciona una estimación de la probabilidad de un evento, nos sugiere un diagnóstico o nos predice una respuesta concreta a un tratamiento.

La lectura crítica de un trabajo sobre una RPC comparte aspectos similares con los de los trabajos sobre pruebas diagnósticas y tiene, además, aspectos específicos relacionados con la metodología de su diseño y aplicación. Por este motivo, vamos a ver someramente los aspectos metodológicos de las RPC antes de entrar en su valoración crítica.

En el proceso de desarrollo de una RPC, lo primero que hay que hacer es **definirla**. Los cuatro elementos clave son la población de estudio, las variables que vamos a considerar como

potencialmente predictivas, el patrón de referencia (el *gold standard* de los que sabéis inglés) que clasifica si el evento que queremos predecir se produce o no y el criterio de valoración del resultado.

Hay que tener en cuenta que las variables que elijamos deben ser clínicamente relevantes, se deben poder recoger de manera precisa y, como es lógico, deben estar disponibles en el momento que queramos aplicar la RPC para la toma de decisiones. Es conveniente no caer en la tentación de meter variables a diestro y siniestro ya que, aparte de complicar la aplicación de la RPC, puede disminuir su validez. En general, se recomienda que por cada variable que se introduce en el modelo deberían haberse producido, al menos, 10 eventos de los que queremos predecir (el diseño se hace en una determinada muestra cuyos componentes presentan las variables pero solo un número determinado acabaron presentando el evento a predecir).

También me gustaría resaltar la importancia del patrón de referencia. Tiene que haber una prueba diagnóstica o un conjunto de criterios bien definidos que nos permitan definir de manera clara el evento que queremos predecir con la RPC.

Por último, es conveniente que los que recojan las variables durante esta fase de definición desconozcan los resultados del patrón de referencia, y viceversa. La ausencia de cegamiento disminuye la validez de la GPC.

La siguiente fase es la **fase de derivación** o de diseño, propiamente dicha. Aquí es donde se aplican los métodos estadísticos que permitirán incluir las variables con capacidad predictiva y excluir las que no nos vayan a aportar nada. No vamos a entrar en la estadística, solo decir que los

métodos empleados con más frecuencia son los basados en la regresión logística, aunque pueden usarse análisis discriminantes, de supervivencia e, incluso, otros mucho más exóticos basados en riesgos discriminantes o redes neurales, al alcance solo de unos pocos virtuosos.

En los modelos de regresión logística, el evento será la variable dependiente dicotómica (ocurre o no ocurre) y las otras variables serán las predictivas o independientes. Así, cada coeficiente que multiplique cada variable predictiva será el antilogaritmo natural de la odds ratio ajustada. Por si alguien no lo ha entendido, la odds ratio ajustada para cada variable predictiva se calculará elevando el número “e” al valor del coeficiente de esa variable en el modelo de regresión.

Lo habitual es que a partir del peso de cada variable se le asigne una determinada puntuación en una escala, de forma que la suma total de puntos de todas las variables predictivas permitirá clasificar al paciente en un rango concreto de predicción de producción del evento. Existen también otros métodos más complejos empleando las ecuaciones de regresión, pero al final siempre se obtiene lo mismo: una estimación individualizada de la probabilidad del evento en un paciente concreto.

Con este proceso se produce la **categorización** de los pacientes en grupos homogéneos de probabilidad, pero aún nos falta saber si esta categorización se ajusta a la realidad o, lo que es lo mismo, cuál es la capacidad de discriminación de la RPC.

La validez global o **capacidad de discriminación** de la RPC se hará contrastando sus resultados con los del patrón de referencia, empleando para ello técnicas similares a las utilizadas

para valorar la capacidad de las pruebas diagnósticas: sensibilidad, especificidad, valores predictivos y cocientes de probabilidad. Además, en los casos en los que la RPC proporcione una estimación cuantitativa, podremos recurrir al uso de las curvas ROC, ya que el área bajo la curva representará la validez global de la RPC.

El último paso de la fase de diseño será la **calibración** de la RPC, que no es más que comprobar su buen comportamiento en todo el intervalo de resultados posibles.

Algunos autores de RPC se quedan aquí, pero olvidan dos pasos fundamentales de la elaboración: la validación y el cálculo del impacto clínico de la regla.

La **validación** consiste en probar la RPC en muestras diferentes a la utilizada para su diseño. Nos podemos llevar una sorpresa y comprobar que una regla que funciona bien en una determinada muestra no lo hace en otra. Por eso hay que probarla, no solo en pacientes similares (validación limitada), sino también en entornos clínicos diferentes (validación amplia), lo que incrementará la validez externa de la RPC.

La última fase es la de comprobar su rendimiento clínico. Aquí es donde muchas RPC se van a pique después de haber pasado por todas las fases anteriores (quizás sea por eso que muchas veces se evita esta última comprobación). Para valorar el **impacto clínico** tendremos que aplicar la RPC en nuestros pacientes y ver cómo cambian medidas de resultados clínicos como supervivencia, complicaciones, costes, etc. La forma ideal para analizar el impacto clínico de una RPC es hacer un ensayo clínico con dos grupos de pacientes manejados con y sin la regla.

Para aquellos abnegados que todavía sigan leyendo, ahora que conocemos qué es y cómo se diseña una RPC, vamos a ver cómo se hace la lectura crítica de estos trabajos. Y para ello, como es habitual, vamos a recurrir a nuestros tres pilares: validez, importancia y aplicabilidad. Para no olvidarnos de nada, seguiremos las preguntas que se enumeran en la parrilla para estudios de RPC de la herramienta CASPe.

En cuanto a la **VALIDEZ**, comenzaremos primero con unas preguntas de eliminación. Si la respuesta es negativa puede que haya llegado la hora de esperar hasta que alguien, al fin, invente una bola de cristal que funcione.

¿La regla responde a una pregunta bien definida? Deben estar claramente definidos la población, el evento a predecir, las variables predictivas y el criterio de valoración de los resultados. Si no se hace así o estos componentes no se ajustan a nuestro escenario clínico, la regla no nos servirá. Las variables predictivas deben ser clínicamente importantes, fiables y estar bien definidas por adelantado.

¿La población a estudio de la que se derivó la regla, incluyó un espectro adecuado de pacientes? Hay que comprobar que el método de selección de los pacientes es el adecuado y que la muestra es representativa. Además, debe incluir pacientes de todo el espectro de la enfermedad. Como ocurría con las pruebas diagnósticas, los eventos pueden ser más fáciles de predecir en determinados grupos, por lo que debe haber representantes de todos ellos. Por último, hay que ver si **la muestra se validó en un grupo diferente de pacientes**. Como ya hemos comentado, no basta con que la regla funcione en el grupo de pacientes en el que se ha derivado, sino que hay

que probarla en otros grupos que sean similares o distintos a aquellos con los que se generó.

Si la respuesta a estas tres preguntas ha sido afirmativa, podemos pasar a las tres preguntas de matiz. **¿Hubo una evaluación ciega del desenlace y de las variables predictoras?** Ya lo hemos comentado, es importante que la persona que recoge las variables predictivas no conozca el resultado del patrón de referencia, y viceversa. La recogida de la información debe ser prospectiva e independiente. Lo siguiente que debemos preguntarnos es si **se midieron las variables predictoras y el desenlace en todos los pacientes**. Si el desenlace o las variables no se miden en todos los pacientes puede comprometerse la validez de la RPC. En cualquier caso, los autores deberán explicar las exclusiones, si es que las hay. Por último, **¿se describen los métodos de derivación y validación de la regla?** Ya sabemos que es fundamental que los resultados de la regla sean validados en una población diferente de la utilizada para el diseño.

Si las respuestas a las preguntas anteriores nos indican que el estudio es válido, pasaremos a responder las preguntas sobre la **IMPORTANCIA** de los resultados. La primera es si **se puede calcular el rendimiento de la RPC**. Los resultados deben presentarse con su sensibilidad, especificidad, cocientes de probabilidades, curvas ROC, etc., en función del resultado proporcionado por la regla (escalas de puntuación, fórmulas de regresión, etc.). Todos estos indicadores nos servirán para calcular las probabilidades de ocurrencia del evento en entornos con prevalencias diferentes. Esto es similar a lo que hacíamos con los estudios de pruebas diagnósticas, así que os invito a que repaséis lo escrito sobre el tema para no repetirnos demasiado. La

segunda pregunta es **¿cuál es la precisión de los resultados?** Aquí tampoco nos vamos a extender más: recordad nuestros venerados intervalos de confianza, que serán los que nos informarán de la precisión de los resultados de la regla.

Para finalizar, pasaremos a considerar la **APLICABILIDAD** de los resultados a nuestro medio, para lo cual trataremos de contestar a tres preguntas. **¿Serán satisfactorios en el ámbito del escenario la reproducibilidad de la RPC y su interpretación?** Habrá que pensar en las similitudes y diferencias entre el ámbito en el que se desarrolla la RPC y nuestro entorno clínico. En este sentido, será de ayuda el hecho de que la regla haya sido validada en varias muestras de pacientes de entornos diferentes, lo que aumentará su validez externa. **¿Es aceptable la prueba en este caso?** Pensaremos si la regla es fácil de aplicar en nuestro medio y si tiene sentido hacerlo desde el punto de vista clínico en nuestro entorno. Finalmente, **¿modificarán los resultados la conducta clínica, los resultados en salud o los costes?** Si, desde nuestro punto de vista, los resultados de la RPC no van a cambiar nada, la regla será inútil y una pérdida de tiempo. Aquí será importante nuestra opinión, pero también deberemos buscar si existen estudios que valoren el impacto de la regla sobre costes o sobre resultados en salud.

Y hasta aquí todo lo que quería deciros sobre lectura crítica de estudios sobre RPC. De todas formas, antes de terminar me gustaría hablaros un poco sobre una lista de verificación que, como no, existe también para la valoración de este tipo de estudios: la lista de verificación CHARMS (*CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies*). No me

diréis que el nombre, aunque un poco rebuscado, no es encantador.

Esta lista está pensada para valorar los estudios primarios de una revisión sistemática sobre RPC. Trata de responder a unas preguntas de diseño general y a la valoración de 11 dominios para extraer la información suficiente para realizar la valoración crítica. Las dos grandes partes que se valoran son el riesgo de sesgo de los estudios y la aplicabilidad de los mismos. El riesgo de sesgo se refiere a los defectos del diseño o validación que pueden dar lugar a que el modelo sea poco discriminativo, excesivamente optimista, etc. La aplicabilidad, por su parte, hace referencia al grado en que los estudios primarios son concordantes con la pregunta que motiva la revisión sistemática, por lo que nos informa de si se puede aplicar la regla a la población diana. Esta lista está bien y ayuda a valorar y comprender los aspectos metodológicos de este tipo de estudios, pero, en mi humilde opinión, es más sencillo realizar una valoración crítica sistemática ayudándonos de la herramienta CASPe.

Y aquí, por fin, lo dejamos por hoy. No hemos hablado nada, para no alargarnos demasiado, de qué hacer con el resultado de la regla. Lo fundamental, ya lo sabemos, es que podemos calcular la probabilidad de producirse el evento en pacientes individuales de entornos con prevalencias diferentes. Pero esa es otra historia...

Bibliografía

- Ochoa Sangrador C. Diseño y evaluación de Reglas de Predicción Clínica. Evid Pediatr. 2009; 5: 28. ([HTML](#)) ([PDF](#))
- Ochoa Sangrador C. Valoración crítica de estudios sobre reglas de predicción clínica. Evid Pediatr. 2009; 5: 52. ([HTML](#)) ([PDF](#))

- Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med.2014;11:e1001744. ([HTML](#))

Correspondencia al autor

Manuel Molina Arias
mma1961@gmail.com
Servicio de Gastroenterología.
Hospital Infantil Universitario La Paz. Madrid.
España.

Aceptado para el blog en noviembre de 2018.