

Lectura crítica de revisiones sistemáticas de estudios de pronóstico o riesgo

Miguel Maldonado Fernández ■ Borja Manuel Fernández Félix
Juan Bautista Cabello López

OBJETIVOS DEL CAPÍTULO

- Definir los distintos tipos de estudios pronósticos.
- Revisar el formato de pregunta clínica empleado específicamente en este tipo de revisiones.
- Comentar los pasos que dan los autores de estas revisiones, haciendo hincapié en los puntos más importantes para la lectura crítica de este tipo de revisiones.

Introducción

Hacer un pronóstico consiste en conocer el futuro. Los estudios de pronóstico buscan averiguar qué le sucederá a un paciente afectado por una determinada circunstancia como una enfermedad, un factor de riesgo o un tratamiento. Este tipo de estudios son especialmente importantes para los pacientes y sus familiares. No obstante, los estudios de pronóstico son especialmente complejos y difíciles de llevar a cabo. La lectura crítica de revisiones sistemáticas de estudios pronósticos posee unas peculiaridades respecto a otro tipo de revisiones sistemáticas. Se deberá tener en cuenta el tipo de estudio de pronóstico analizado. La pregunta clínica tiene un formato PICO-TA. Se emplean herramientas específicas para la extracción de datos de los estudios individuales (CHARMS-PF) y para el estudio del riesgo de sesgo (QUIPS). La certidumbre en la evidencia (calidad de la evidencia) depende del riesgo de sesgo en cada estudio individual y, además, del riesgo de sesgo de la propia revisión por factores como la imprecisión, el sesgo de publicación o la existencia de evidencia indirecta, entre otros. GRADE es una herramienta que mide la certeza en la evidencia. La heterogeneidad entre estudios pronósticos es frecuente, por lo que se recomienda usar un modelo de efectos aleatorios para el metaanálisis y presentar intervalos de predicción para la estimación del efecto. Las revisiones de factores pronóstico son a menudo complejas, pues suelen presentar dificultades tales como sesgo de publicación o reporte selectivo, diferencias en la elección de los puntos de corte o distintos factores de ajuste.

Escenario

Formas parte de la comisión de guías y protocolos de tu hospital y en las últimas reuniones se ha planteado como objetivo actualizar los protocolos generales de reanimación cardiopulmonar (RCP) que llevan algunos años sin actualizar. En realidad piensan que será un proceso complejo con múltiples estratos que incluye desde la revisión de la infraestructura institucional, los protocolos (pre-, intra- y posparada), la actualización del sistema de recogida de datos en coherencia con los

estándares internacionales para registros de investigación sobre RCP (GWTG-R registry, UK National Cardiac Arrest. The Utstein register: DOI: 10.1161/CIR.0000000000000710) y también la promoción de campañas para generar reflexiones y cambios culturales al respecto en el hospital y en el área de salud.

En todo caso, el grupo promotor de la comisión y los documentalistas han hecho una búsqueda muy amplia y parece que hay muchos frentes que revisar. Entre los documentos que destacan está una clásica recomendación de expertos de la American Heart Association (DOI: 10.1161/CIR.0b013e31828b2770), y también aparecen bastantes estudios de factores pronósticos y una interesante revisión sistemática que resume y sintetiza todos los estudios de ese tipo.

Fernando SM, Tran A, Wei Cheng, et al. Pre-arrest and intra-arrest prognostic factors associated with survival after in-hospital cardiac arrest: systematic review and meta-analysis. *BMJ*. 2019;367:l6373.

Como saben de tu interés y de tus habilidades en la lectura crítica de estudios, te encargan revisar en profundidad ese artículo y comentarlo en la próxima sesión de la comisión.

De modo que te tocará opinar sobre:

1. ¿Qué factores influyen en el pronóstico de las paradas intrahospitalarias?
2. ¿Algunos de esos factores o grupos de factores deberían ser considerados en la elaboración de la nueva estrategia?

Lee el artículo y contesta a esas preguntas.

¿Qué es pronosticar?

Hacer un pronóstico consiste en conocer el futuro. Por lo menos, desde un punto de vista etimológico. A los efectos de este capítulo, los estudios pronósticos son aquellos que buscan averiguar qué le sucederá a un paciente afectado por una determinada circunstancia (una enfermedad, un factor de riesgo, un tratamiento). Los estudios pronósticos tienen interés tanto para profesionales de la salud como para pacientes y sus familiares, políticos y encargados de tomar decisiones sobre salud (1).

El pronóstico forma parte de la tríada de la práctica clínica: diagnóstico-tratamiento-pronóstico. En la época hipocrática el pronóstico era el elemento más importante de los tres (se conocía poco de los elementos diagnósticos de las enfermedades, y el tratamiento en muchos casos se limitaba al conocido *primum non nocere*). En el siglo XX cobró importancia el diagnóstico (las opciones de tratamiento aún estaban bastante limitadas). Hoy día, los avances en el conocimiento de los mecanismos de las enfermedades permiten conocer más íntimamente los mecanismos de las enfermedades en un paciente concreto. Esto hace posible establecer un pronóstico más afinado para una determinada persona, con unas características particulares. Además, han surgido con fuerza técnicas para extraer y analizar cantidades ingentes de datos de salud (*big data*).

Los términos referentes a los estudios pronósticos eran tradicionalmente confusos. Para solucionar este problema se creó la Estrategia en Investigación Pronóstica, en inglés PROGnosis REsearch Strategy, conocida por su acrónimo PROGRESS (2). En esta Estrategia se propone clasificar los estudios pronósticos en cuatro tipos distintos:

PROGRESS tipo I: estudios de pronóstico global. En estos estudios se analizan los resultados reales en muestras de pacientes con una determinada enfermedad o situación de salud de interés. Se llaman «globales» porque el resultado es una medida global, como los valores medios de una medida de la enfermedad. Por ejemplo, «puntuación en el MINIMENTAL test a los 12 meses» o «porcentaje que sigue sin poder trabajar a los 12 meses».

PROGRESS tipo II: estudios de factores pronósticos (3). Estudian qué características (o factores) se asocian con cambios en resultados globales para los individuos del estudio.

PROGRESS tipo III: modelos pronósticos (4). Evaluación de modelos matemáticos de predicción de riesgo que incorporan múltiples factores pronósticos.

PROGRESS tipo IV: predictores del efecto de un tratamiento (5). Estudian las características que predicen si un individuo responderá o no a un determinado tratamiento.

En este capítulo nos ceñiremos a las revisiones sistemáticas de estudios pronóstico tipo II (factores pronósticos). Un factor pronóstico puede definirse como una característica o variable del paciente, que está asociada a una determinada probabilidad de sufrir un resultado relevante (por ejemplo, sufrir un infarto). Para la estructura general de una revisión sistemática remitimos a los lectores al capítulo 12 («Lectura crítica de revisiones sistemáticas sobre estudios de prevención o tratamiento»). Muy resumidamente, los pasos de una revisión sistemática son (6):

- Protocolo.
- Búsqueda de los estudios individuales.
- Evaluación del riesgo de sesgo en cada estudio individual seleccionado.
- Extracción de los datos de cada estudio individual.
- Síntesis de la «evidencia» (si es posible).
- Diseminación de los resultados.

Puntos clave para la lectura de una revisión sistemática de estudios pronósticos

PREGUNTA PICO-TA

Para definir adecuadamente la pregunta que busca responder la revisión sistemática es necesario establecer el tipo de estudios que nos interesan según la clasificación PROGRESS. Recordemos que, en este capítulo, hablaremos específicamente del PROGRESS tipo II. Los estudios de cohortes prospectivas son los idóneos para llevar a cabo estas revisiones. No obstante, en ocasiones no se dispone de cohortes prospectivas sino de otro tipo de estudios, como cohortes históricas.

En las revisiones sistemáticas de estudios de tratamiento se utiliza el formato PICO (Paciente, Intervención, Control y *Outcome* o resultado) para construir la pregunta clínica. Para las revisiones de estudios pronóstico se adapta la pregunta PICO al formato PICO-TS, en español PICO-TA, que es el acrónimo de:

Población: población general en la que se estudiará el factor pronóstico.

Índice: factor pronóstico que se está analizando.

Comparador(es)/control(es): define dos conceptos: el comparador, otro factor de riesgo con el que se desea comparar el índice pronóstico bajo revisión; o el confusor (cuando el propósito no es comparativo) que es un factor de ajuste considerado en la estimación del efecto del factor pronóstico bajo revisión.

Outcome: resultado o evento que se está intentando predecir (por ejemplo, mortalidad por infarto agudo de miocardio).

Tiempo (*Timing*): cuándo se mide el factor pronóstico y en qué lapso de tiempo se predecirá el *outcome* o desenlace.

Ámbito (*Setting*): el escenario donde se utilizará el factor pronóstico.

ESTRATEGIA DE BÚSQUEDA

La búsqueda de los estudios individuales es más compleja en las revisiones pronósticas por el hecho de que los estudios no suelen estar etiquetados como «pronósticos» y, por lo tanto, existe el riesgo de que no se detecten mediante una estrategia de búsqueda convencional. Otra dificultad añadida es que no existen filtros metodológicos de búsqueda que hayan sido validados. Se han desarrollado y validado filtros metodológicos para la identificación de estudios de modelos pronóstico (7), que han mostrado relativa capacidad para identificar estudios de factores pronóstico.

Por lo tanto, por miedo a perder estudios primarios importantes, suele realizarse una estrategia más amplia, con el inconveniente de que se obtienen muchos resultados que no son relevantes. Cuando la pregunta de revisión se centra en un factor pronóstico, desenlace o población específica,

TABLA 15.1 ■ Dominios analizados en CHARMS-PF

Área o dominio	Comentarios
Origen de los datos	<ul style="list-style-type: none"> • (Ensayo clínico, estudio de cohortes, casos y controles, etc.)
Participantes	<ul style="list-style-type: none"> • Elegibilidad de los participantes y método de selección • Descripción de los participantes • Detalles de los tratamientos recibidos, si fuese relevante • Fechas de los estudios
Desenlaces (<i>outcomes</i>) que serán medidos	<ul style="list-style-type: none"> • Definición del desenlace y método para su medición • ¿Se han usado la misma definición y método de medición en todos los participantes? • ¿Desenlace único o combinado? • ¿Hubo enmascaramiento para el desenlace? • Tiempo hasta la aparición del desenlace
Predictores candidatos	<ul style="list-style-type: none"> • Número y tipo de predictores (por ejemplo: características demográficas, historia del paciente, exploración física, nuevas pruebas diagnósticas, características de la enfermedad) • Definición del predictor candidato y método para su medición • Momento de la medición del factor predictor (o pronóstico) • ¿Se enmascararon los factores para el resultado? • ¿Se enmascararon entre sí? • ¿Cómo se tratan los factores en el modelo de predicción? (variable continua, categórica, transformación lineal, transformación no lineal)
Tamaño de la muestra	<ul style="list-style-type: none"> • ¿Se calculó el tamaño de la muestra? ¿Cómo? • Número de participantes y de eventos • Número de desenlaces/eventos por cada factor pronóstico
Datos que faltan	<ul style="list-style-type: none"> • Número de participantes en los que falta algún valor (incluyendo factores pronósticos y desenlaces) • Número de participantes en los que falta algún dato para cada uno de los factores pronósticos • Datos de atrición. En estudios de supervivencia, número de observaciones censuradas • Cómo se manejaron los datos que faltan (imputación de datos faltantes, análisis de casos completos, etc.)
Análisis	<ul style="list-style-type: none"> • Tipo de modelo (logístico, lineal, Cox, etc.) • Cómo se comprobaron las asunciones del modelo • Método empleado para la selección de factores candidatos en el modelo multivariante • Método de selección de factores durante el modelado multivariante (selección retrógrada o anterógrada) y criterios para la selección (valor de la p; criterio de información de Akaike) • Métodos para el manejo de factores continuos (dicotomización, categorización, lineal, no lineal), incluyendo los puntos de corte elegidos y su justificación
Resultados	<ul style="list-style-type: none"> • Estimaciones del efecto pronóstico, crudas y ajustadas, junto con sus intervalos de confianza correspondientes • Para cada estimación ajustada, indicar los factores de ajuste
Interpretación	<ul style="list-style-type: none"> • Interpretación de los resultados presentados • Comparación con otros estudios

Modificado de Riley 2019 (6).

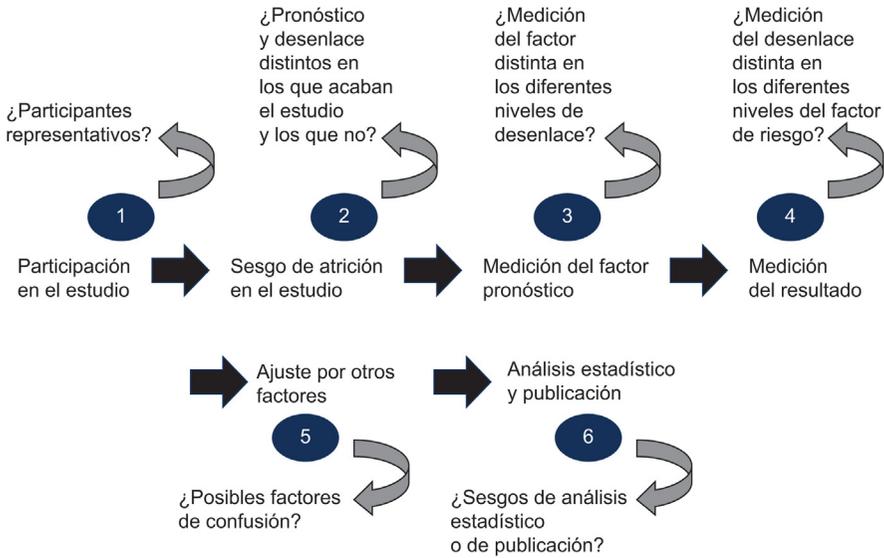


Figura 15.1 Dominios analizados por QUIPS. (Modificado de Hayden 2013 [9]).

añadir estos términos en la estrategia de búsqueda reducirá considerablemente el número de artículos identificados.

Sesgo de publicación y reporte selectivo

Uno de los problemas más importantes que nos encontramos en las revisiones sistemáticas de estudios de factores pronóstico es el sesgo de publicación. Bien es sabido que las revistas científicas son más proclives a publicar estudios que presentan resultados, digamos, estadísticamente «significativos». Pero no menos importante es que los investigadores no envíen a publicar aquellos estudios en los que no se encontraron los resultados esperados, y cuando estos se encontraron no es poco frecuente que sea tras buscar y escarbar sobremanera en los datos, dando lugar al reporte selectivo. Todo ello pone de manifiesto la importancia de los protocolos en la investigación del pronóstico.

PROTOCOLO

Al leer una revisión sistemática de estudios pronósticos se debería considerar si los autores han cumplido con sus objetivos iniciales. Si no lo han hecho, se podría estar incurriendo en algún tipo de sesgo. El lector debería poder consultar dicha información en el protocolo. Este documento es un plan detallado de lo que se va a hacer y cómo se va a llevar a cabo en la revisión sistemática. El protocolo incluye el razonamiento y la justificación de la revisión (¿por qué hay que hacer esta revisión?); los objetivos; los criterios de elegibilidad de los estudios; el método de extracción de datos; la evaluación crítica; los métodos estadísticos para sintetizar la «evidencia» (el resultado global de la revisión); y la redacción (*report*) clara y completa de los resultados. Se puede buscar el protocolo en PROSPERO, un registro específico que depende de la Universidad de York (<https://www.crd.york.ac.uk/prospere/>), o publicado en ciertas revistas científicas. Se trata, en cualquier caso, de ofrecer al lector de la revisión la opción de contrastar los objetivos iniciales de los autores con lo que finalmente se presentó.

CÓMO SE HA EXTRAÍDO LA INFORMACIÓN DE CADA ESTUDIO

La herramienta CHARMS (CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) es una lista con los ítems que deben extraerse de los estudios individuales que se incluyen en una revisión sistemática de estudios pronósticos (8). Aunque CHARMS está diseñada para las revisiones de modelos pronósticos (es decir, los PROGRESS tipo III), existe una adaptación para las revisiones de estudios de factores pronósticos, que recibe el imaginativo nombre de CHARMS-PF (6). Los ítems incluidos vienen recogidos en la [tabla 15.1](#).

RIESGO DE SESGO

El siguiente paso en la revisión es comprobar cuán creíble es la información que hemos encontrado, es decir, cuál es el riesgo de sesgo de cada uno de los estudios. Es probable encontrarnos en algunas publicaciones «riesgo de sesgo». Para medir el riesgo de sesgo existe la herramienta QUIPS, acrónimo de QUality In Prognosis Studies (9), que evalúa los siguientes «dominios» reportados según CHRAM-PF ([fig. 15.1](#)):

- **Participación en el estudio.** En este ítem se comprueba si la relación entre el factor pronóstico y el desenlace puede ser diferente en los participantes en el estudio que estamos analizando, comparado con la «población elegible» que no está en ese estudio. Por eso se describe la fuente de la muestra extraída, las características basales de esa muestra, cómo se ha obtenido y durante cuánto tiempo, y cuáles son los criterios de inclusión y exclusión. En resumen: los participantes del estudio ¿son «similares» a la población donde usaremos los factores pronósticos?
- **Sesgo de atrición en el estudio.** Analizaremos si la relación entre factor pronóstico y desenlace es probablemente muy distinta entre los que han concluido el estudio y aquellos que no han llegado al final (o «se han perdido» por el camino).
- **Medición del factor pronóstico.** Investigaremos si la medición del factor pronóstico se ha llevado a cabo de manera diferente en los distintos niveles del desenlace (*outcome*) estudiado. Si se midiese de modo más exhaustivo el factor pronóstico en los que han desarrollado el desenlace comparado con los que no lo han desarrollado, podríamos encontrar que el factor se relaciona con el desenlace pero que este hallazgo no sea cierto.
- **Medición del resultado.** La medición del resultado o desenlace (*outcome*), ¿se ha realizado de modo distinto en los diferentes niveles del factor pronóstico (es decir, más en los que han estado expuestos al factor que entre los que no han estado expuestos, por ejemplo)?
- **Ajuste por otros factores.** Comprobaremos si la relación entre el factor pronóstico y el desenlace es probable que esté afectada por otra variable o factor que esté relacionado con el factor pronóstico estudiado y con el desenlace. Es decir, estudiaremos si es posible que existan factores de confusión.
- **Análisis estadístico y publicación.** Consideraremos si el resultado (del estudio individual que estamos analizando) sea espurio y realmente se deba a sesgos en el análisis estadístico o en la publicación de resultados.

¿SE HA HECHO UN METAANÁLISIS?

El metaanálisis no siempre es el producto final de la revisión sistemática. Un metaanálisis solo se debe llevar a cabo cuando los estudios identificados sean suficientemente robustos y comparables, de modo que los resultados derivados de este tengan una interpretación y un impacto directo en los cuidados de salud. El metaanálisis requiere de, al menos, dos estudios que estimen el mismo parámetro. Cuando se agregan estudios de baja calidad, la evidencia también será de baja calidad.

Aunque nos vamos a centrar en el metaanálisis de datos agregados, es decir, de estimaciones extraídas desde los estudios identificados en la búsqueda, una alternativa es el metaanálisis de datos

individuales de pacientes, el cual precisa de la información individual de los pacientes incluidos en los estudios primarios.

Para el metaanálisis de datos agregados son varios los obstáculos que se suelen encontrar, y que dificultan la interpretación de los resultados. Algunas de las dificultades más comunes en los estudios de factores pronóstico son: *primero*, diferentes tipos de estimaciones, en este tipo de estudios es frecuente encontrar la estimación del tamaño del efecto en términos de riesgo relativo (RR), *odds ratio* (OR) o *hazard ratio* (HR) cuando la variable de resultado es dicotómica, o la diferencia media cuando la variable es cuantitativa; *segundo*, estimaciones sin errores estándar, cuando en los métodos de metaanálisis estándar se emplean para ponderar el peso de cada estudio; *tercero*, estimaciones en diferentes tiempos de predicción: en ocasiones, los tiempos de predicción de un mismo desenlace o el momento de medición del factor pronóstico bajo revisión difieren entre estudios; *cuarto*, diferentes métodos o instrumentos de medida tanto para el desenlace como para el factor pronóstico; *quinto*, mezcla de estudios en los cuales la estimación del efecto fue ajustado en unos y crudo (o sin ajuste) en otros, y aunque todos los estudios reporten una estimación del efecto ajustada el conjunto de variables de ajuste (confusores) frecuentemente divergirá entre ellos. En este sentido, suele ayudar predefinir un conjunto básico de factores de ajuste que represente un ajuste mínimo necesario para la inclusión del estudio, por ejemplo que el efecto del factor pronóstico este ajustado, al menos, por género y edad; *sexto*, el manejo de factores pronóstico de tipo continuo: algunos estudios podrían considerar una relación lineal, otros ajustar tendencias no lineales, y otros establecer diferentes puntos de corte para categorizar o dicotomizar el factor pronóstico bajo revisión. Cuando el punto de corte elegido es seleccionado porque minimiza el *p*-valor asociado con el efecto pronóstico de interés, tenderá a sesgar los resultados hacia un mayor efecto pronóstico. Cuando se categoriza o dicotomiza el factor pronóstico o desenlace de interés la decisión de los puntos de corte debe establecerse *a priori*. Muchos de estos aspectos dirigen a sustancial heterogeneidad, causando que la estimación del efecto pronóstico varíe entre estudios.

Por otra parte, en el caso de realizar el metaanálisis, dada la heterogeneidad inherente de los estudios de factores pronóstico, es recomendado el uso de métodos de efectos aleatorios para considerar la heterogeneidad no explicada entre estudios.

Un metaanálisis de efectos aleatorios combina las estimaciones del efecto del factor pronóstico bajo revisión entre los estudios, obteniendo un efecto promedio (μ) y la desviación estándar a través de los estudios (τ). Si Y_i y $\text{var}(Y_i)$ denotan la estimación del efecto y su varianza en el estudio i , en términos generales un modelo de metaanálisis de efectos aleatorios se puede especificar como:

$$Y_i \sim N(\mu, \text{var}(Y_i) + \tau^2)$$

Esta simpática fórmula indica que, según el modelo de efectos aleatorios, se tienen en cuenta dos fuentes de variabilidad: la propia *dentro* de cada estudio y la variabilidad *entre* estudios.

Común a otros tipos de revisiones, existen diferentes métodos para estimar el modelo. El modelo de efectos fijos utiliza habitualmente el método de Mantel-Haenszel, mientras que para el modelo de efectos aleatorios el más frecuentemente utilizado por los investigadores es el método de DerSimonian y Laird (10), usual en metaanálisis de ensayos clínicos y de especial utilidad en el caso de los estudios de factores pronósticos por la heterogeneidad arriba señalada. En el contexto de heterogeneidad de los estudios de factores pronóstico se recomienda el método de Hartung-Knapp (11) —y, cuando el número de estudios que se van a combinar es pequeño, el método de Hartung-Knapp-Sidik-Jonkman (12)—, dado que ha demostrado ser más robusto.

Para llevar a cabo el metaanálisis se recomienda emplear la escala original solo cuando el estadístico estimado es la diferencia media (desenlace de tipo cuantitativo); cuando el estadístico que se desea agregar es un RR, OR o HR, la escala apropiada para el metaanálisis es el logaritmo neperiano. En este caso, el estadístico promedio y sus intervalos de confianza son estimados en escala logarítmica y, posteriormente, se deben transformar de nuevo a la escala original.

HETEROGENEIDAD

Cuando la heterogeneidad entre los estudios identificados en la revisión sistemática es substancial, la estimación promedio resultante del metaanálisis es difícil de trasladar a la práctica clínica. En tales situaciones el hallazgo principal de la revisión es la propia heterogeneidad identificada entre los estudios y la necesidad de investigar las posibles causas. La variabilidad entre los estudios puede ser mostrada mediante un *forest plot*, preferiblemente sin el resultado de la estimación promedio del efecto pronóstico del factor bajo revisión.

La heterogeneidad, como ya se ha mencionado en el capítulo 12, se puede cuantificar mediante el estadístico I^2 , el cual mide el porcentaje de la variabilidad total debida a las diferencias entre estudios, y cuyo rango de valores oscila entre 0 y 100%. Valores próximos a 0% indican poca heterogeneidad, y a medida que aumenta, se incrementa la sospecha de heterogeneidad.

Si el metaanálisis se realiza a pesar de la presencia de heterogeneidad, es recomendable presentar el intervalo de predicción de la estimación del efecto pronóstico. Dicho intervalo de predicción indica el potencial valor del verdadero efecto pronóstico del factor en una nueva población a partir de los resultados de la revisión. Técnicas bayesianas también pueden ser empleadas para obtener inferencias predictivas. Por ejemplo, tras el metaanálisis se podría obtener la probabilidad de que el verdadero efecto pronóstico del factor sea superior a un valor dado (p. ej., un HR > 1,3 para un factor binario, que indica un incremento del riesgo de al menos un 30%).

Como en revisiones sistemáticas de intervenciones, análisis de subgrupos y metarregresión pueden ser empleados para explorar y examinar las potenciales causas de heterogeneidad.

ANÁLISIS DE SENSIBILIDAD

En ocasiones es preciso hacer un análisis a parte de algún subgrupo de estudios, que por algún motivo nos interesan de forma especial. Por ejemplo, podemos querer hacer un análisis de los estudios con poco riesgo de sesgo, excluyendo los que tienen un riesgo de sesgo elevado. O podemos querer analizar estudios realizados exclusivamente en ancianos o en personas con o sin una determinada característica, para comprobar qué resultado arroja el metaanálisis en ese caso.

GRADE (CERTIDUMBRE DE LA EVIDENCIA)

GRADE es el acrónimo de Grading of Recommendations Assessment, Development and Evaluation (13). Es una herramienta que mide, por un lado, la fuerza de la evidencia científica y, por otro, la fuerza de la recomendación basada en esa evidencia. La calidad de la evidencia o certidumbre depende de dos factores. Por una parte, el riesgo de sesgo en cada uno de los estudios incluidos en la revisión. Por otro, el riesgo de sesgo de la propia revisión, debido a factores como la imprecisión, los sesgos de publicación, la evidencia indirecta, etc. GRADE analiza de forma individualizada cada desenlace y cada factor pronóstico.

Conclusión

La lectura crítica de revisiones sistemáticas de estudios pronósticos posee unas peculiaridades respecto a otro tipo de revisiones sistemáticas. Se deberá tener en cuenta el tipo de estudio de pronóstico analizado. La pregunta clínica tiene un formato PICO-TA. La herramienta CHARMS (CHECKlist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) es una lista con los ítems que deben extraerse de los estudios individuales que se incluyen en una revisión sistemática de estudios pronósticos. La herramienta QUIPS (QUALity In Prognosis Studies) se emplea para evaluar el riesgo de sesgo en estas revisiones. GRADE es una herramienta que mide de modo independiente la certidumbre en la evidencia y la fuerza de la evidencia. Las revisiones de estudios pronósticos tipo II enlazan con otros estudios tipo III y tipo IV (modelos

pronósticos y estudios predictores a tratamiento, y con estudios de medicina personalizada, que influirán de forma determinante en el desenlace vital de nuestros pacientes.

Artículo

Fernando SM, Tran A, Wei Cheng, et al. Pre-arrest and intra-arrest prognostic factors associated with survival after in-hospital cardiac arrest: systematic review and meta-analysis. *BMJ*. 2019;367:l6373. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6891802/>.

Plantilla CASPe contestada para este artículo concreto

En el [cuadro 15.1](#) se muestra la plantilla CASPe contestada para este artículo concreto.

CUADRO 15.1 ■ Evaluación crítica del artículo propuesto (plantillas CASPe)

A) ¿Los resultados de la revisión son válidos?

Preguntas «de eliminación»

1. ¿Se hizo la revisión sobre un tema claramente definido?

PISTA: un tema debe ser definido en términos de (PICO-TA):

- *La Población de estudio.*
- *Los Índices pronósticos.*
- *Comparadores (si procede).*
- *Los desenlaces considerados (Outcomes).*
- *Tiempo (factor y desenlace).*
- *Ámbito.*

Sí ✓

No sé

No

No se describe específicamente la pregunta en formato PICO-TA (en inglés, PICO-TS).

No obstante, del texto se recoge:

- *Pacientes:* pacientes ingresados que han sufrido un paro cardíaco. Se incluyen los estudios con al menos un 80% de «adultos», considerados como al menos de 16 años.
- *Índice pronóstico:* no se menciona ninguno específico. Factores preparada y factores intraparada.
- *Comparador/control:* recuérdese que el comparador hace referencia a otro factor pronóstico con el que se compara el factor a estudio; mientras que el control es otro factor que se utiliza en el ajuste de la predicción con el factor de estudio. En esta revisión no existe un *comparador*, es decir, se evalúa el efecto pronóstico de múltiples factores sin un objetivo comparativo entre ellos. En cuanto a factores de *control* (o ajuste), en el manuscrito no se hace referencia a ellos. Sin embargo, en el protocolo se ha predefinido que los estudios de factores pronósticos incluidos en la revisión deben haber ajustado las estimaciones del efecto al menos por edad y sexo.
- *Outcome:* mortalidad intrahospitalaria a los 28 o 30 días.
- *Timing:* se evalúan por separado los factores medidos previos al paro cardíaco y los factores medidos durante el paro cardíaco. El tiempo de desenlace se especifica en el desenlace (*outcome*) de interés.
- *Ámbito (Setting):* ámbito hospitalario.

CONCLUSIÓN: sí.

<p>2. ¿Buscaron los autores el tipo de artículos adecuado?</p>	Sí ✓	No sé	No
<p><i>PISTA: el mejor «tipo de estudio» es el que:</i></p> <ul style="list-style-type: none"> • Se dirige a la pregunta objeto de la revisión y tiene el diseño apropiado. • Tipo de estudio de pronóstico. 	<p>Los autores hacen referencia a la inclusión de estudios observacionales con diseño retrospectivo y prospectivo, ensayos clínicos aleatorizados y ensayos cuasi aleatorizados. Los estudios con un diseño prospectivo son considerados los mejores para responder a una pregunta de factores pronóstico. Estos permiten una mejor definición de los criterios de inclusión. La recogida de información, basal y durante el seguimiento, es más completa y estandarizada, tanto en forma como en la definición de factores y desenlaces bajo estudio. Esa diversidad contribuye a la heterogeneidad.</p>		
CONCLUSIÓN: sí.			
Preguntas detalladas			
<p>3. ¿Crees que estaban incluidos los estudios importantes y pertinentes?</p>	Sí ✓	No sé	No
<p><i>PISTAS</i></p> <p><i>DE LA BÚSQUEDA:</i></p> <ul style="list-style-type: none"> • ¿Qué bases de datos bibliográficas se han usado? ¿Qué estrategia de búsqueda? • Seguimiento de las referencias. • Contacto personal con autores. • Búsqueda de estudios no publicados. • Idiomas distintos del inglés. 	<p>Se llevó a cabo la búsqueda en Medline, PubMed, Embase, Scopus, Web of Science, y la Cochrane Database of Systematic Reviews. Se utilizó la herramienta <i>Related Articles</i> («artículos relacionados») de PubMed para ampliar la búsqueda. La estrategia de búsqueda, que se presenta en el material suplementario, fue diseñada por una bibliotecaria experimentada en ciencias de la salud. Esta combina términos específicos de la población bajo revisión, tales como <i>cardiac arrest</i>, y relacionados con la investigación en pronóstico, como <i>prognostic o risk</i>.</p>		
<p><i>PISTAS</i></p> <p><i>DE LA SELECCIÓN:</i></p> <ul style="list-style-type: none"> • Criterios de inclusión/exclusión. • Selección estudios. ½ • Extracción de datos ½ (usaron CHARMS-PF). 	<p>Los autores de la revisión contactaron con los autores de correspondencia de los estudios primarios, cuando estos no reportaron las estimaciones de los <i>odds ratios</i>, ajustados o crudos, o los datos necesarios a partir de los cuales poder calcularlos. Aunque en la estrategia de búsqueda no hay limitaciones idiomáticas, solo se incluyen artículos en inglés. Aunque no consideramos que se trate de un porcentaje relevante, sería interesante conocer cuántos estudios fueron descartados por estar escritos en otro idioma, en nuestra opinión presumiblemente pocos.</p>		
CONCLUSIÓN: sí.			
<p>4. ¿Crees que los autores de la revisión han hecho suficiente esfuerzo para valorar la calidad de los estudios incluidos?</p>	Sí ✓	No sé	No
<p><i>PISTA 1: QUIPS «riesgo de sesgo» depende de seis dominios: población, atrición, medición de factores pronósticos, medición de desenlaces, confusión y análisis estadístico correcto.</i></p>	<p>Los autores han valorado la calidad de los estudios mediante la herramienta QUIPS (QUality In Prognosis Studies), la herramienta recomendada para evaluar el riesgo de sesgo en este tipo de revisión sistemática.</p> <p>El riesgo de sesgo se evalúa usando 31 preguntas divididas en seis dominios (población, atrición, medición de los factores pronósticos, medición del desenlace, factores de ajuste y análisis estadístico), y para dominio los estudios se clasifican en bajo, moderado y alto riesgo de sesgo.</p>		
<p>Se presentan los resultados en la tabla 7 del material adicional. Véase como, en la mayoría de los dominios, el riesgo de sesgo es bajo. En otros, como el sesgo de atrición en Ballew, es moderado.</p>			
CONCLUSIÓN: sí.			

(Continúa)

5. Si los resultados de los diferentes estudios han sido mezclados para obtener un resultado «combinado», ¿era razonable hacer eso?

PISTA: la heterogeneidad (I^2) puede ser:

- Clínica.
- Metodológica.
- Estadística.

Sí ✓

No sé

No

La heterogeneidad fue evaluada usando el estadístico I^2 , el test χ^2 para la homogeneidad y, visualmente, mediante los *forest plots*.

El estadístico I^2 era muy alto para varios de los metaanálisis realizados.

Por ejemplo, en la evaluación de los factores pronóstico preparada: sexo masculino $I^2 = 66\%$ o edad > 70 $I^2 = 69\%$; y de los factores intraparada: diagnóstico de síndrome coronario agudo $I^2 = 99\%$. Los autores indican que la heterogeneidad es debida a que en el metaanálisis se combinan estudios de grandes registros nacionales con una variabilidad pequeña de la estimación del efecto y estudios pequeños con mayor variabilidad, y justifican su uso basado en el solapamiento de las estimaciones puntuales y los correspondientes intervalos de confianza al 95%. Sin embargo, y ante la evidente presencia de heterogeneidad entre estudios, hubiera sido de interés para el lector que en el metaanálisis se hubieran presentado los intervalos de predicción, que informan del potencial valor verdadero del efecto en un nuevo estudio.

Un análisis de sensibilidad, excluyendo del metaanálisis los estudios pequeños de gran variabilidad, podría interesar al lector, de modo que permita examinar esas potenciales causas de heterogeneidad.

CONCLUSIÓN: sí.

B) ¿Cuáles son los resultados?

6. ¿Cuál es el resultado global de la revisión?

PISTA: considera:

- Valora para los desenlaces positivos y también los negativos.
- ¿Cuáles son los resultados para cada desenlace?
- ¿Cómo están expresados los resultados RR, HR, etc.?)
- ¿Muestran gráficos forest plots?

Los resultados para el análisis principal que evalúa la asociación entre los factores pronósticos preparada e intraparada con la *odds* de supervivencia, se presentan en la tabla 2 y las figuras 2-3 (*forest plots* para los factores pronóstico preparada) y 4-5 (*forest plots* para los factores pronóstico intraparada).

Para cada factor pronóstico estudiado se indican los estudios incluidos en el metaanálisis, el tamaño del efecto en términos de *odds ratios* y sus correspondientes intervalos de confianza al 95%, la heterogeneidad mediante el estadístico I^2 y el grado de certeza mediante la evaluación GRADE (tabla 15.2).

Factores preparada:

El sexo masculino (OR = 0,84 [0,73 a 0,95] con grado de certeza moderado), la edad avanzada (OR = 0,50 [0,40 a 0,62] para edad mayor a 60 años y OR = 0,42 [0,18 a 0,99] para edad mayor a 70 años, ambos con grado de certeza bajo), la existencia de una neoplasia concomitante (OR = 0,57 [0,45 a 0,71] con grado de certeza alto) y la existencia de enfermedad renal crónica (OR = 0,56 [0,40 a 0,78] con grado de certeza alto) se relacionaron con un peor pronóstico. Otras comorbilidades (insuficiencia cardíaca congestiva, enfermedad pulmonar obstructiva crónica y diabetes mellitus) y el diagnóstico de sepsis a la admisión se asociaron a un peor pronóstico en estudios individuales (sin metaanálisis).

Factores intraparada:

Los paros cardíacos ante testigos (OR = 2,71 [2,17 a 3,38]), la monitorización (OR = 2,23 [1,41 a 3,52]), los paros cardíacos diurnos (con plantilla al completo) (OR = 1,41 [1,20 a 1,66]), la fibrilación ventricular (OR = 3,68 [2,68 a 5,05]) y la taquicardia ventricular (OR = 3,76 [2,95 a 4,78]) se relacionaron con un mejor pronóstico. Todos ellos con alto grado de certeza. Mientras que la asistolia (OR = 0,42 [0,32 a 0,56] con grado de certeza alto), la intubación traqueal (OR = 0,54 [0,42 a 0,70] con grado de certeza moderado) y la duración prolongada de las maniobras de resucitación cardiopulmonar (OR = 0,12 [0,07 a 0,19] con grado de certeza alto) se relacionaron con un peor pronóstico.

7. Para el conjunto de los estudios (en cada desenlace concreto)

- ¿Cuál es la precisión de los resultados?
- ¿Son consistentes los resultados de los estudios para cada desenlace?
- ¿Es indirecta la evidencia en algún desenlace?

En la revisión sistemática de ejemplo, dado que son varios los factores pronósticos bajo estudio, las siguientes cuestiones deben discutirse para cada factor concreto.

- **¿Cuál es la precisión de los resultados?**
Para evaluar de forma crítica la precisión de los resultados debemos fijarnos en los intervalos de confianza de la estimación puntual del efecto. Por ejemplo, para el factor pronóstico *historia de malignidad* (preparada) la estimación puntual del *pooled odds ratio* es 0,57, con un intervalo de confianza al 95% relativamente ajustado (preciso), entre 0,45 y 0,71. Sin embargo, para el factor *edad ≥ 70* (preparada), la estimación puntual del *pooled odds ratio* es 0,42, y su intervalo de confianza mucho más holgado (impreciso), entre 0,18 y 0,99. Ver tabla 2.
- **¿Son consistentes los resultados de los estudios para cada desenlace?**
La consistencia de los resultados depende de la heterogeneidad entre los estudios. Esta se puede valorar a partir de los valores del estadístico I^2 , los test estadísticos de heterogeneidad u observando el grado de solapamiento entre los intervalos de confianza de los estudios identificados. En el caso de disponer de ellos, también podríamos ayudarnos de los intervalos de predicción. El valor del estadístico I^2 en todos los factores pronóstico metaanalizados excede del 50%, indicando un importante grado de heterogeneidad entre los estudios; además, las potenciales causas de heterogeneidad no han sido exploradas mediante análisis de subgrupos. Por tanto, la consistencia de los resultados es moderada.
- **¿Es indirecta la evidencia en algún desenlace?**
Cuando la definición del desenlace, el factor pronóstico o la población bajo revisión diverge entre los estudios incluidos, podría ser un síntoma de evidencia indirecta.

C) ¿Son los resultados aplicables en tu medio?

8. ¿Se pueden aplicar los resultados en tu medio?

Sí ✓	No sé	No
------	-------	----

- PISTA: considera si:*
- *Los pacientes cubiertos por la revisión pueden ser suficientemente diferentes de los de tu área.*
 - *Tu medio es muy diferente a los del estudio.*

Los pacientes y el ámbito donde se ha elaborado la revisión sistemática son, en principio, similares a los que encontraríamos en otros hospitales de nuestro entorno.
CONCLUSIÓN: parece que sí, aunque desconocemos elementos como infraestructura, protocolos o entrenamiento del personal de los centros.

9. ¿Se han considerado todos los resultados necesarios para tomar una decisión?

Sí	No sé ✓	No
----	---------	----

- ¿Qué te gustaría saber además de esto?

Sería interesante comprobar el papel de la telemedicina para detección de paradas cardíacas, que parece que desempeñará un papel de importancia creciente. También convendría saber cuál es el nivel de entrenamiento y formación del personal hospitalario en esta cuestión. Otra variable que se debe tener en cuenta es el tipo de planta hospitalaria y sus características (médica/quirúrgica, etc.).
CONCLUSIÓN: no sé.

(Continúa)

10. ¿Crees que hay alguna medida que tomar en tu caso?	Sí ✓	No sé	No
<ul style="list-style-type: none"> Aunque no esté planteado explícitamente en la revisión, ¿qué opinas? 	<p>El manejo de las paradas cardíacas es complejo y depende de varios factores, que incluyen las características estructurales del propio hospital, las características y el entrenamiento del personal y los propios protocolos de manejo de las paradas.</p> <p>Conocer los factores pronósticos relacionados con la odds de supervivencia es útil para afinar el diseño de protocolos y guías de actuación. Además, a partir de los resultados de la revisión tenemos un interesante punto de partida para el desarrollo de un modelo pronóstico que nos permita estimar de manera individualizada la probabilidad de supervivencia a una parada cardíaca según las características del paciente, así como la probabilidad de parada en un paciente cuando ingresa (pre-arrest), con la finalidad de prevenir o jerarquizar el riesgo.</p> <p>CONCLUSIÓN: sí.</p>		

TABLA 15.2 ■ Tabla resumen de resultados GRADE por factores

Factor estudiado y n.º de estudios	Riesgo de sesgo	Imprecisión	Inconsistencia*	Evidencia indirecta	Certeza en la evidencia**	Estimador IC 95%
«Parada» presenciada por terceros (4 estudios)	No importante	No importante	No importante $I^2 = 68\%$	No importante	Certeza GRADE alta	2,71 (2,17-3,38)
Paciente monitorizado (6 estudios)	No importante	No importante	No importante $I^2 = 97\%$	No importante	Certeza GRADE alta	2,23 (1,41-3,52)
Parada en horario diurno (5 estudios)	No importante	No importante	No importante $I^2 = 94\%$	No importante	Certeza GRADE alta	1,41 (1,20-1,66)
Ritmo inicial susceptible de desfibrilación (12 estudios)	No importante	No importante	No importante $I^2 = 96\%$	No importante	Certeza GRADE alta	5,28 (3,78-7,39)
Intubación durante la parada (5 estudios)	No importante	No importante	No importante $I^2 = 73\%$	Importante***	Certeza GRADE moderada	0,54 (0,42-0,70)
Duración de la resucitación > 15 min (2 estudios)	Importante	No importante	No importante $I^2 = 27\%$	No importante	Certeza GRADE alta	0,12 (0,07-0,19)

Explicaciones: *A pesar de altos valores de I^2 hay alto grado de solapamiento entre las estimaciones puntuales y los intervalos de confianza. **La mayor parte del peso en la estimación del efecto agrupado proviene de estudios de bajo riesgo de sesgo, excepto para el factor «duración de la resucitación», que proviene de estudios de moderado riesgo de sesgo. ***Tiempos de intubación variables y no están claras otras variables de confusión que contribuyen a si el paciente está o no intubado.

La tabla (tomada del artículo) presenta algunas discrepancias respecto a la presentada por los autores de la revisión en el material adicional (tabla suplementaria 9). Esto sugiere que se ha llevado a cabo un análisis de sensibilidad o que por alguna otra razón se ha descartado algún estudio en el análisis final presentado en el manuscrito. En cualquier caso, en ambas tablas la certeza en la evidencia de las estimaciones no cambia (es alta), y aunque los OR difieren ligeramente (como es lógico con distinto número de estudios incluidos en el análisis), la dirección del efecto es consistente en ambas tablas. Por lo tanto, no existen discrepancias esenciales. No obstante, se ha escrito a los autores y editores para la justificación o corrección de estas inconsistencias.

Bibliografía

1. Riley RD, van der Windt DA, Croft P, Moons KGM. Prognosis Research in Health Care: Concepts, Methods, and Impact. Oxford: Oxford University Press; 2019.
2. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. PROGRESS Group. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595.
3. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. PROGRESS Group. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med* 2013;10(2):e1001380.
4. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. PROGRESS Group. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10(2):e1001381.
5. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. PROGRESS Group. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ* 2013;346:e5793.
6. Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ* 2019;364:k4597.
7. Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeftang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS One* 2012;7(2):e32844.
8. Moons KG, de Groot JA, 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(10):e1001744.
9. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.
10. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-88.
11. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med* 2001;20(24):3875-89.
12. Röver C, Knapp G, Friede T. Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC* 2015;15:99.
13. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.

Cómo citar este capítulo:

Maldonado M, Fernández-Félix BM, Cabello JB. Lectura crítica de revisiones sistemáticas de estudios de pronóstico o riesgo. En: Cabello Juan B, editor. *Lectura crítica de la evidencia clínica*, 2.ª ed. Barcelona: Elsevier; 2022. p. 182-195.