



ELSEVIER

RADIOLOGÍA

[www.elsevier.es/rx](http://www.elsevier.es/rx)



## ARTÍCULO ESPECIAL

# La aplicabilidad de los resultados



CrossMark

I. Marín-León

Medicina Interna, Hospital Universitario Virgen del Rocío, Sevilla, CIBERESP-IBIS, Fundación Enebro, Sevilla, España

Recibido el 12 de marzo de 2015; aceptado el 30 de julio de 2015

Disponible en Internet el 9 de octubre de 2015

### PALABRAS CLAVE

Medicina basada en evidencia;  
Desenlaces;  
Aplicabilidad;  
Literatura médica

**Resumen** La lectura crítica de la literatura médica tiene por objeto último la utilización de los avances científicos en la práctica asistencial o innovadora. Ello conlleva la evaluación de la aplicabilidad de los resultados de los estudios publicados, comenzando por un claro entendimiento de los resultados.

Cuando los estudios no ofrecen suficientes garantías de rigor en su desarrollo y análisis no se dan las condiciones necesarias para su aplicación, pero que sean fiables los resultados no es suficiente para tratar de utilizar sus conclusiones. En el presente artículo se explica cómo el desarrollo en condiciones experimentales o artificiales de los estudios suele alejarlos de las condiciones reales en las que se pretenden aplicar sus conclusiones. Para valorar dicha aplicabilidad se propone evaluar un conjunto de atributos, que permita discernir al lector la probabilidad de que los beneficios y riesgos que describen los estudios se den con la menor incertidumbre en el entorno asistencial en el que se pretenden poner en práctica.

© 2015 SERAM. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

### KEYWORDS

Evidence-based medicine;  
Outcomes;  
Applicability;  
Medical literature

### The applicability of results

**Abstract** The ultimate aim of the critical reading of medical literature is to use the scientific advances in clinical practice or for innovation. This requires an evaluation of the applicability of the results of the studies that have been published, which begins with a clear understanding of these results.

When the studies do not provide sufficient guarantees of rigor in design and analysis, the conditions necessary for the applicability of the results are not met; however, the fact that the results are reliable is not enough to make it worth trying to use their conclusions. This article explains how carrying out studies in experimental or artificial conditions often moves them away from the real conditions in which they claim to apply their conclusions. To evaluate

Correo electrónico: [ignacio.marin.sspa@juntadeandalucia.es](mailto:ignacio.marin.sspa@juntadeandalucia.es)

<http://dx.doi.org/10.1016/j.rx.2015.07.007>

0033-8338/© 2015 SERAM. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

this applicability, the article proposes evaluating a set of items that will enable the reader to determine the likelihood that the benefits and risks reported in the studies will yield the least uncertainty in the clinical arena where they aim to be applied.

© 2015 SERAM. Published by Elsevier España, S.L.U. All rights reserved.

## Introducción

Cuando se habla de la aplicabilidad de los resultados mostrados en una publicación nos estamos situando en la perspectiva del clínico que realiza lectura crítica de literatura científica con el fin de mantenerse al día en sus conocimientos para el mejor desempeño de su práctica clínica. Dada la abrumadora oferta de estudios publicados, interesa al clínico ser muy selectivo y optimizar su recurso tiempo para estudiar solamente aquellos artículos que puedan ser más aplicables en su situación clínica, es decir, aquellos que le hagan mejor profesional asistencial. Es por ello que parece pertinente la inclusión de un artículo sobre este tema en la monografía de lectura crítica, pues facilitará conocer aspectos sobre qué trabajos publicados leer y cómo leerlos para poder aplicarlos en la asistencia.

La aplicabilidad de un artículo requiere la premisa de que el mismo sea digno de confianza, es decir, que los resultados se consideren válidos y precisos. Lo contrario, tratar de aplicar en la vida real de la asistencia sanitaria las conclusiones de un artículo sobre el que se tienen dudas en relación con cuál es la magnitud real del efecto que describe (precisión), y si dicho efecto es real y no está falseado por limitaciones en el modo como se han obtenido los resultados (exactitud), sería una temeridad.

Por tanto, solo una vez constatada la validez de los resultados de un estudio pueden plantearse las cuestiones de utilidad, su aplicación a la práctica clínica<sup>1</sup>.

Al hablar de aplicabilidad el lector se está planteando si el estudio es generalizable, si es relevante el efecto que describe y, por último, si es utilizable en las circunstancias en las cuales ejerce su práctica asistencial. Un estudio es generalizable cuando es esperable que sus resultados sucedan igualmente más allá del lugar donde se ha realizado y con otros pacientes. Es decir, que las condiciones impuestas en la experimentación no son tan restrictivas, que sería una rareza que se dieran los resultados en otros entornos. Al hablar de relevancia del estudio nos referimos a que los desenlaces que describe son de tal magnitud que su aplicación puede merecer la pena para los pacientes y las instituciones, ya que el balance riesgo-costes/beneficios que describe es claramente favorable al beneficio. Por utilidad de los resultados estamos resaltando que la intervención que describe el estudio es razonablemente utilizable en el entorno donde se realiza la práctica clínica del lector; por ejemplo, que la incidencia del problema que aborda permita predecir una gran utilidad para el tipo de pacientes o los recursos que habitualmente encuentra el lector del estudio en la clínica. La aplicabilidad es, por tanto, un juicio personal que debe realizar el clínico

lector según las circunstancias en las que realiza su práctica habitual.

Tal juicio tiene dos partes bien diferenciadas. La primera, la correcta interpretación de los resultados a la vista de los desenlaces que interesan al lector; la segunda, la aplicación de esos resultados relevantes al entorno donde el lector realiza su práctica clínica.

El objetivo de este artículo es revisar los atributos que se han de evaluar en un estudio para saber si sus resultados son generalizables, si la relevancia de los resultados que presenta el estudio merecen la pena para el paciente medio del lector del mismo, y si son aplicables en el entorno del lector del estudio (su utilidad).

## ¿Pueden ser útiles los resultados?

Para poder valorar con rigor un estudio se precisa de un pensamiento crítico, de la actitud escéptica que la medicina basada en la evidencia pregonó para la medicina científica. Esa es la que se ejemplifica muy bien en las reflexiones de Hill en 1965, cuando reclamaba que más allá de la evidencia de los resultados de una asociación estadística válida, se precisa de un juicio científico que elabore la utilidad o la causalidad, sin limitarse solo a la estadística<sup>2</sup>. Se podrá entender mejor con el siguiente ejemplo.

Imaginemos que se publica un estudio con metodología impecable sobre la utilidad de la mamografía seriada para prolongar el pronóstico vital de las mujeres, y encuentra que el cribado con senografía empeora la expectativa de vida en 5 años con una  $p < 0,01$ . Al leerlo, algunos pensarán que hay que descartar los cribados de cáncer de mama mediante mamografía, ya que solo hay una probabilidad del 1% de que el hallazgo del estudio se haya producido por azar. Dicho planteamiento sería incorrecto; el valor de la  $p$  no significa esto, todo lo que hace es estimar de modo conjunto los datos, asumiendo la específica hipótesis nula del estudio<sup>3</sup>. Pero no puede asumirse acríticamente para realizar afirmaciones que trasciendan la realidad subyacente al estudio, y que requieren más elementos de información.

La interpretación de tal estudio hipotético incluiría varias posibilidades; por supuesto, entre ellas, que el efecto que detecta se produzca (es decir, que la mamografía seriada se relaciona con un peor pronóstico vital). Ello podría tener otra explicación. Imaginemos que el estudio se hizo en un país con bajos recursos tecnológicos, donde solo se realizan mamografías a las mujeres con lesiones de mama previamente detectadas clínicamente, y por tanto la muestra está sesgada por selección de pacientes más graves; el diseño del estudio y en concreto tal cuestión de muestreo es lo que

marca la interpretación y aplicabilidad del estudio, no el valor de la *p*.

Como norma general, cuanto menos plausible sea una hipótesis, mayor es la posibilidad de que un hallazgo muy novedoso sea una falsa alarma, independientemente del valor de la *p*. En nuestro caso hipotético, la probabilidad de que sea un hallazgo falso es de menos del 99%, aunque la posibilidad de que se produzca un resultado similar si se replica el estudio es solo del 11%<sup>3</sup>.

Por este motivo es por el que muchos metodólogos señalan que el valor de la *p* puede facilitar un pensamiento perezoso, al hacer caer en la seductiva certeza de la significación por encima de la dimensión real del efecto, obviando la relevancia clínica o biológica del mismo. Se trata de pasar de preguntarnos ¿hay algún efecto? a ¿cuánto efecto hay? En consonancia, algunos editores de revistas están planteando que se evite el valor de *p* y se solicite a los autores un estimador del efecto y su intervalo de confianza, la magnitud y la importancia relativa de un efecto.

Se trata de incorporar en el razonamiento científico del lector la gran protección que supone admitir todo lo plausible, pero todo, incluido la falsedad del hallazgo. Tal actitud crítica de escepticismo científico se alimenta con tres cuestiones que un médico científico ha de preguntarse tras leer un estudio: ¿Cuál es la evidencia? ¿Qué debo creer? ¿Qué puedo hacer con ello? Ningún método estadístico puede responder las tres preguntas porque los números (los hechos) son donde la medicina científica comienza la discusión, no donde la acaba<sup>1</sup>.

Un estudio puede aportar evidencia de baja calidad para la práctica clínica por tres vías: primero, su diseño y ejecución pueden ser excelentes, pero la muestra no tiene nada que ver con los pacientes que el lector ve en la vida real (es el caso de la transposición de estudios en animales a los humanos); segundo, la ejecución y la muestra son las correctas, pero los desenlaces que mide son poco importantes para los pacientes (es el caso de los estudios con desenlaces subrogados, como detección de nódulos menores de 1 cm frente a la supervivencia a los 10 años tras mamografías<sup>4</sup>), y tercero, la ejecución, la muestra y los desenlaces son los adecuados, pero el diseño es débil (como intentar demostrar el efecto de una intervención mediante un estudio observacional, en vez de con un ensayo clínico controlado [ECC]).

A la hora de aceptar la fiabilidad de los resultados también hay que prestar atención a algunas situaciones que se dan en los estudios, y que afectan a la aplicabilidad de los mismos<sup>5</sup>. Especial precaución hay que tener con la interpretación de estudios en cuyo diseño y desarrollo se den algunas situaciones que pueden afectar a la fiabilidad de los resultados, como son los estudios finalizados antes de tiempo<sup>6</sup>, los que utilizan desenlaces subrogados<sup>7</sup> y los que usan desenlaces compuestos. Las dos primeras situaciones se han comentado en otros artículos de esta monografía, pero a efectos de la aplicabilidad de los resultados, interesa detenernos en los problemas que plantean los artículos que usan desenlaces compuestos.

Algunos ensayos clínicos están utilizando desenlaces compuestos, que capturan cualquiera de un conjunto de posibles desenlaces como resultado primario del ensayo (muerte, infartos e ingresos, por ejemplo) y así es posible disminuir los tamaños muestrales y acortar la duración

**Tabla 1** Criterios para valorar la utilización de desenlaces combinados

- ¿Cuáles son las opciones clínicas?
- ¿Son los componentes del desenlace compuestos de similar importancia para los pacientes?
- ¿Ocurren los diferentes desenlaces con la misma frecuencia?
- ¿Tiene una razonable seguridad de que los diferentes componentes comparten una reducción del RR similar?
- ¿La biología subyacente a los componentes del desenlace es suficientemente similar como para esperar que la RRR sea similar?
- ¿Es el estimador de la RRR similar entre los componentes y con intervalos de confianza suficientemente estrecho?

RR: riesgo relativo; RRR: reducción relativa del riesgo.

de los estudios. La justificación (suficientes eventos totales, con pocos eventos en cada uno de los desenlaces del compuesto, y la presunción de que la intervención actuará de una manera similar para producir las diferentes categorías de eventos) supone el reto más importante a la hora de aplicar los resultados de los ECC a la realidad del lector.

Para ayudar en la evaluación de la aplicabilidad de los estudios con desenlaces compuestos se aconseja seguir el cuestionario recogido en la **tabla 1**<sup>8</sup>.

Si se puede contestar afirmativamente a las preguntas planteadas se puede estar seguro de que el efecto detectado para los desenlaces combinados puede ser tenido en cuenta para la toma de decisiones. Si la respuesta a alguna de las preguntas es negativa, para tomar decisiones el lector debe mirar por separado el efecto en cada uno de los componentes del desenlace compuesto.

Nos podemos encontrar con estudios en los que: 1) los componentes del desenlace compuesto son de similar relevancia clínica (infarto miocárdico o ictus); 2) los desenlaces más relevantes ocurren con frecuencias similares a los menos importantes; y 3) los distintos desenlaces están ligados por una fuerte racionalidad biológica, la reducción relativa del riesgo (RRR) es similar y los intervalos de confianza (IC) excluyen el azar en cada uno de los componentes del desenlace. Si se dan todas esas circunstancias, los clínicos pueden confiar en un desenlace combinado como resultados de un estudio<sup>9</sup>. Pero no suele ser siempre así, numerosos estudios presentan un panorama en el extremo contrario: 1) los componentes del resultado combinado tienen importancia clínica muy distinta; 2) los desenlaces más relevantes ocurren con mucha menos frecuencia que los menos importantes, y 3) la racionalidad biológica es débil, las RRR difieren ampliamente y los IC de los desenlaces importantes incluyen la posibilidad de daño. En tales casos, para tomar decisiones clínicas por parte del lector debe confiar en los resultados de la RRR y el IC de los desenlaces individuales que interesen al lector.

Las situaciones intermedias entre ambos extremos exigen interpretar la aplicabilidad de los resultados del estudio.

Así, la utilidad que se da al efecto comunicado en el estudio dependerá de que el desenlace que más interesa al lector ocurra más frecuentemente que los que menos interesan al paciente. Por ello, los clínicos han de escrutar con atención los resultados para cada uno de los desenlaces que más interesen a sus pacientes<sup>8</sup>.

**Tabla 2** Criterios para evaluar la aplicabilidad de los resultados

- ¿Son los pacientes del estudio y su asistencia similares a los de mi centro?
- ¿Ha sido la duración del estudio lo suficientemente larga?
- ¿Puedo utilizar los resultados en la asistencia de los pacientes de mi centro?
- ¿Los beneficios superan los riesgos y el coste?

No son infrecuentes las situaciones en las que la precisión del resultado la aporta la alta frecuencia de desenlaces menos interesantes para los pacientes, tal como se confirma al separar el análisis por cada desenlace del compuesto.

### Definir la diferencia mínimamente importante

La interpretación de resultados con vistas a su aplicabilidad requiere algún tipo de referencia, que pueda ser utilizada como estándar independiente de la relevancia de los resultados. Para ello se han diseñado los estándares de diferencia mínimamente importante (DMI), que se define para un desenlace como aquel valor menor del mismo que el paciente aprecia como una mejoría (o perjuicio) en la calidad del desenlace<sup>10</sup>. Imaginemos que ante determinada pérdida de agudeza visual de base, con la intervención de la catarata de un ojo el paciente recupera un 0,5 de agudeza visual, pero con la intervención de los dos ojos el paciente solo recupera el 0,6 y con este 10% más el paciente no percibe ningún beneficio añadido al que ya tenía con la intervención de un solo ojo. En este caso, la DMI sería un 0,2 de agudeza visual, por lo que cualquier resultado comunicado en una publicación que sea menos de ese, en principio no aporta beneficio detectable por el paciente, y por tanto aplicarle una intervención para obtener tan escaso beneficio puede no superar los riesgos de la misma<sup>11</sup>. Y el clínico lector no la aplicaría.

### Aplicar los resultados al paciente individual

Para facilitar el juicio de la aplicabilidad del estudio que se esté considerando se ha propuesto un proceso sistemático mediante las cuatro cuestiones recogidas en la **tabla 2**<sup>12</sup> que valoran la extrapolación al entorno concreto del lector. Dichas cuestiones resumen los criterios de validez externa y aplicabilidad que pueden ayudar a mantener un compromiso entre la generalización imprudente y el rechazo en la aplicación de los resultados de los estudios. De este modo se dispone de una ayuda que pueda dar confianza en la decisión que se tome sobre la aplicabilidad fuera del entorno donde se ha realizado el estudio. Obviamente, depende del lugar, la población del estudio y el tiempo de seguimiento de esa población, tres factores que condicionan la generalización, pero también hay otros aspectos, como los conflictos de intereses, que pueden lastrar la credibilidad de los resultados<sup>13</sup>.

Aunque los ECC suelen ser muy rigurosos en garantizar su validez interna (la fiabilidad de la comparación entre los grupos de tratamiento y el control), suelen poner menos empeño en garantizar la validez externa, la comparabilidad de la muestra con los pacientes habituales,

fundamentalmente por condicionantes en el muestreo. El motivo es que los ECC se enfocan a demostrar la eficacia del fármaco, más que en la efectividad en la vida real.

Si la vida real fuera un ensayo clínico, estos constituirían una evidencia definitiva, pero la vida real es mucho más compleja y descontrolada que las limitaciones impuestas en un ECC<sup>14</sup>. Por ello, los clínicos deben ser diestros en discernir si aplican los resultados de los estudios a los pacientes individuales. Veamos en detalle los cuatro pasos recogidos en la **tabla 2**.

### ¿Son los pacientes del estudio y su asistencia similares a los de mi centro?

La clave de esta pregunta es valorar si en el estudio en evaluación han sido excluidos factores biológicos que pudieran modificar la respuesta al tratamiento, factores que presentan los pacientes del entorno del clínico lector, entre los cuales los más típicos son la edad, el sexo y la comorbilidad<sup>9</sup>.

Sin duda, los criterios de inclusión y exclusión son una información que aportan los estudios para facilitar la aplicabilidad; del mismo modo, los análisis de subgrupos son otra ayuda para valorar el efecto en subpoblaciones específicas en las que pudiera estar interesado el lector.

Entre los criterios de exclusión hay que cotejar si por algún factor biológico pudiera ser que los pacientes habituales no hubieran cumplido los criterios de elegibilidad (o que hace a tales pacientes diferentes de los del estudio que se va a aplicar). Dichos factores diferenciadores se pueden resumir en la regla nemotécnica PESES (Patología, Edad, Comorbilidad, Etnia, Sexo)<sup>9</sup>. Por «patología» se refiere no solo a la gravedad o el estado evolutivo de la enfermedad en cuestión, también lo pueden ser el patrón de resistencia de unos gérmenes, o que la efectividad de la vacunación antigripal propuesta depende de la sensibilidad de la cepa de ese año.

Como norma general hay que tender a aplicar los resultados en la asistencia habitual, a no ser que haya fuertes evidencias para alguno de estos factores biológicos en el sentido de que se pueda ocasionar más daño que beneficio, por ejemplo en las personas a partir de determinada edad. Para ello ayuda considerar los análisis de subgrupos o algún estudio epidemiológico que sugiera lo contrario al estudio en evaluación.

Si desde el punto de vista biológico no hay nada que conduzca a no aplicar el resultado, a continuación ha de repararse en si existen aspectos sociales o personales que puedan modificar la efectividad o la seguridad de la aplicación a los pacientes de la asistencia habitual, por darse esta en condiciones muy diferentes. Entre ellos destacan condicionantes socioeconómicos que pudieran modificar la aplicabilidad del tratamiento al entorno social en el que se toma la decisión, como son el acceso al fármaco, o la cumplimentación de los requerimientos de la intervención del estudio. La cumplimentación es un aspecto clave, pues en la vida real es siempre menor que en los estudios, por cuestiones de actitud de los pacientes o de recursos para facilitarla<sup>11</sup>.

En resumen, saber si el fármaco o la intervención funcionará para todos los pacientes en condiciones reales de la asistencia.

## ¿Ha sido la duración del estudio lo suficientemente larga?

Es importante valorar si los estudios epidemiológicos o los ensayos clínicos han dado ocasión a la muestra del estudio, a detectar el riesgo o desarrollar los eventos adversos de la intervención, ya que hay una diferencia fundamental entre los estudios y la vida real, y es que, mientras los primeros se realizan en un tiempo recortado, los pacientes requieren asistencia prolongada<sup>6,14</sup>.

## ¿Puedo utilizar los resultados en la asistencia de los pacientes de mi centro?

Tras la consideración del entorno social del paciente hay que analizar el contexto institucional donde los profesionales pretenden aplicar las intervenciones del estudio. Es decir, si la disponibilidad de equipos, técnicas, destrezas y capacidad organizativa permiten cumplir los requerimientos del tratamiento.

En las intervenciones que requieran especiales conocimientos o habilidades de los médicos (lo cual no es infrecuente en los estudios radiológicos), la variabilidad del efecto en la vida real puede ser muy importante, ya que las condiciones en que se ejecuta el estudio que demostró el efecto suelen ser muy diferentes (apoyo, entrenamiento, presión asistencial, medios) de las de las intervenciones del médico en su ámbito asistencial. Ello puede comprometer seriamente la aplicabilidad<sup>9</sup>.

Por tanto, se trata de garantizar que la aplicación al paciente será igual de experta, selectiva y segura como en el estudio, evitando que la aplicabilidad sea dudosa.

## ¿Los beneficios superan los riesgos y costes?

Una vez el lector se queda satisfecho con que ni los factores biológicos ni los sociales u organizativos comprometen la efectividad y seguridad de la propuesta del estudio en el medio donde se pretende aplicar, hay que pasar a evaluar el posible efecto y la seguridad en el paciente individual. Para ello, ayuda mucho la conversión de la magnitud del efecto calculándolo en términos de número necesario a tratar (NNT); esto permite hacerse una idea de hasta qué punto es probable que un paciente concreto se beneficie<sup>12</sup>.

Se trata de que cuando el clínico dispone de un resultado que considera válido (IC estrecho y sin incluir el 1) y convincente (efecto de gran magnitud), debe entenderlo con claridad para saber cómo utilizarlo en su práctica clínica. Esto es lo que se pretende con los NNT<sup>12</sup> al convertir los riesgos relativos en riesgos absolutos para un periodo relevante, ya que las medidas de riesgo relativo no ayudan mucho a discernir el impacto sobre la práctica; por contra, al comunicar los resultados en términos absolutos, su comprensión es más intuitiva. Es un método para presentar los resultados de una manera que permite calcular la importancia clínica del desenlace, y se define como el número de pacientes que debe exponerse para obtener el desenlace, así como el daño por cada unidad de efecto obtenido<sup>15</sup>.

Por ejemplo, en un estudio que muestre un riesgo absoluto (RA) de morir por cáncer tras 2 años con una

**Tabla 3** Impacto de la reducción relativa del riesgo según el riesgo basal al que se aplique

Riesgo basal (%)	RRR (%)	Riesgo tras intervención (%)	RAR (%)	NNT
20	25	15	5	20
16	25	12	4	25
12	25	9	3	33
8	25	6	2	50
4	25	3	1	100

NNT: número necesario tratar:  $(1/RAR) * 100$ ; RAR: reducción absoluta del riesgo; RRR: reducción relativa del riesgo.

intervención del 25% en el grupo A y del 10% en el grupo B, donde se calcula que la reducción relativa del riesgo (RRR) de morir es del 60% ( $[25-10]/25 * 100$ ), mientras que la reducción absoluta de riesgo (RAR) es del 15% (25%-10%), ¿cuántos pacientes han de ser intervenidos como B en vez de como A para salvar una vida? Pues 15 de cada 100 que intervengamos como al grupo B, es decir, un  $NNT = 100/15 = 6,66$ , que se interpreta como que hay que intervenir durante 2 años sobre siete pacientes para salvar una vida.

Cuando en un estudio se publica que hay una reducción del riesgo de un determinado porcentaje, habitualmente se está hablando de la reducción relativa (o el aumento relativo). En nuestro ejemplo, el 60%, que es más impactante para el lector que el riesgo absoluto, pero tiene una peor comprensión intuitiva que la reducción absoluta (o aumento absoluto) del riesgo, el 15% en el ejemplo<sup>15</sup>.

Además, la RAR, y por tanto el NNT, se modifica en relación con el riesgo basal del que se parte. Este es un aspecto primordial para entender el impacto del resultado según los riesgos del entorno del lector (la prevalencia del problema) y el tiempo para conseguir el impacto. El NNT es inversamente proporcional al riesgo basal de la población; así, si se parte de un riesgo de la mitad, el NNT será el doble, y si el riesgo basal de los pacientes es el doble, el NNT será la mitad. Veamos un ejemplo (tabla 3) de las diferencias de efectos absolutos, con un mismo efecto relativo, según sea el riesgo basal del que se parte.

La misma proporción de descenso del riesgo relativo del 25% se traduce para unos pacientes en una reducción absoluta del 5% y para otros pacientes del 1%, que conduce a una repercusión práctica muy distinta, tal como se ve en la última columna, la de tener que tratar a 20 o a 100 pacientes para que uno de ellos reciba el beneficio de la intervención<sup>9</sup>.

A la hora de interpretar los resultados para su aplicación en la práctica habitual, cuando el efecto es muy grande hay que encontrar razones muy poderosas para privar al paciente de tal beneficio; no obstante, cuando la magnitud del beneficio no es tan potente, se puede ponderar la decisión con todos los factores comentados, incluido el coste, juicio que puede llevar a no aplicar las conclusiones de un estudio.

Cuando se sospecha una aplicabilidad limitada por la baja magnitud del efecto, estimada mediante el NNT, la actuación del clínico dependerá de si en su entorno es mejorable tan bajo efecto; por ejemplo, unas diferencias por factores biológicos se pueden corregir cambiando la dosis, unas diferencias de aplicabilidad por falta de cumplimiento o de habilidades se puede corregir con formación y

entrenamiento<sup>12</sup>. Lo que es siempre imprescindible es apoyar los cálculos adaptados al paciente particular, mediante el NNT, para que sea intuitiva la decisión en términos del balance beneficio, daños y coste<sup>9</sup>. Es una forma de considerar el efecto teniendo en cuenta el riesgo concreto de los pacientes.

## Otros aspectos para la aplicación de los resultados de un estudio

### La calidad de la evidencia que aporta un estudio

La toma de decisiones requiere calibrar la certeza en que se basan las mismas, por ello es importante cuantificar la calidad de la evidencia en que se apoyan dichas decisiones. La conclusión sobre una asociación estadística puede realizarse con garantías con diversos tipos de estudios, aunque hay unos diseños más robustos que otros<sup>1</sup>. Así, a partir de un estudio observacional analítico diseñado *ex profeso* para testar una hipótesis, puede confirmarse con seguridad una relación causal, una vez que se hayan excluido el azar, los sesgos y las variables de confusión como alternativas explicativas de la asociación, mientras que para estudios con resultados de efectos moderados o pequeños es necesaria la evidencia de un ECC<sup>1</sup>. Pero ¿qué ocurre cuando los ECC disponibles tienen importantes errores metodológicos, o no hay ninguno disponible que responda la pregunta que nos planteamos en la práctica? Por ello surgió el sistema GRADE (*Grading of Recommendations Assessment Development and Evaluation*)<sup>16</sup>, que reconsidera la fuerza de la evidencia que aportan los distintos tipos de diseños, a fin de discernir sobre la certeza en la que se apoya nuestro conocimiento sobre el tema de interés en estudio. Se trata de un sistema sencillo, que utiliza escalas estructuradas para calificar la calidad de la evidencia que aportan los estudios, así como el grado de recomendación que puede sustentar dicha evidencia. Así, un estudio observacional bien diseñado y ejecutado puede alcanzar el nivel más alto de evidencia, y apoyar con mayor certeza una recomendación que un ECC con mala calidad metodológica<sup>16</sup>.

### La relación de causalidad

Toda asociación estadística de magnitud elevada no es evidencia de causalidad (etiológica o terapéutica). La causalidad es un juicio que se elabora a partir de una indubitable evidencia. Los criterios que conjuntamente apoyan un juicio positivo de causalidad incluyen fuerza de la asociación, plausibilidad biológica, consistencia, relación temporal y relación dosis-respuesta<sup>17</sup>.

Los criterios de Hill sobre causalidad se consideran las bases que permiten pasar de la asociación entre una exposición y un desenlace a una causalidad, pero requieren que previamente haya quedado sin lugar a duda que la relación entre la exposición A y el efecto B no se deba al azar.

En detalle, los criterios que permiten apoyar la causalidad de una relación son los nueve recogidos en la **tabla 4**, que se comentan brevemente a continuación:

**Tabla 4** Criterios de causalidad, tomados de Hill<sup>17</sup>

Fuerza de la asociación
Consistencia
Especificidad
Temporalidad
Gradiente biológico
Plausibilidad biológica
Coherencia
Experimentable
Analogía

1. **Fuerza de la asociación:** por ejemplo, que los grandes fumadores tienen veinte veces más probabilidad de tener cáncer de pulmón que los no fumadores.
2. **Consistencia:** la asociación se encuentra en similar grado en diferentes lugares, con diferentes poblaciones y períodos de tiempo.
3. **Especificidad:** la asociación se produce entre dicha exposición y dicho desenlace, el desenlace no se da con otra exposición, al menos con la misma fuerza que con la primera exposición.
4. **Temporalidad:** la relación temporal claramente identifica que la exposición se da antes de producirse los desenlaces.
5. **Gradiente biológico:** a más exposición, mayor magnitud del desenlace, es decir, algún tipo de gradiente dosis-respuesta.
6. **Plausibilidad biológica:** la relación es explicable en el contexto de los conocimientos generales de la ciencia biológica.
7. **Coherencia:** la asociación no es contradictoria con los conocimientos generales sobre la patología.
8. **Experimentable:** que pueda ponerse a prueba la relación de causalidad.
9. **Analogía:** que relaciones similares se hayan observado en otros juicios de causalidad.

La presencia de los nueve criterios en sí misma no lleva linealmente a la decisión de causalidad, al igual que ningún criterio es imprescindible. Así, cuanto mayor sea la fuerza de la asociación entre la exposición y el efecto, implica que las explicaciones alternativas de dicha relación, incluida el azar, son menos probables<sup>2</sup>. Una explicación biológica aceptada también incrementa la opción de causalidad, mientras que las relaciones que suponen una explicación biológica novedosa, como fue el caso de *Helicobacter pylori* y la úlcera péptica, tardan más en ser aceptadas, aunque sean ciertas. En este caso jugaba a favor de la causalidad la analogía con otras enfermedades digestivas causadas por infecciones, como la enfermedad de Whipple. Por su parte, la aceptación de causalidad es mayor si los resultados se replican consistentemente en estudios independientes, en varios lugares y diferentes poblaciones, mientras que la secuencia temporal de primero la exposición y luego el efecto a veces es difícil de establecer, ya que muchas variables de estilos de vida no son fáciles de ordenar cronológicamente.

La decisión sobre causalidad debe ser un juicio que exige considerar todos estos criterios y plantearse el razonamiento de refutación. ¿Hay alguna otra forma lógica de explicarse la asociación detectada en el estudio?<sup>17</sup>

## Conclusión

En este artículo hemos revisado la aplicabilidad de los estudios para tomar decisiones en la práctica habitual, utilizar una exploración o un tratamiento, o bien tomar una decisión de causalidad. Para ello se ha seguido un sistema que complementa la lectura crítica sobre la calidad con la que están desarrollados los estudios. Tras confirmarse la fiabilidad en los resultados que aportan, el sistema expuesto ofrece una serie de cuestiones que permiten hacer un juicio sobre la magnitud del efecto y la probabilidad de que la intervención evaluada actúe del mismo modo en el entorno y sobre los pacientes de la práctica habitual del lector.

## Responsabilidades éticas

**Protección de personas y animales.** Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

**Confidencialidad de los datos.** Los autores declaran que en este artículo no aparecen datos de pacientes.

**Derecho a la privacidad y consentimiento informado.** Los autores declaran que en este artículo no aparecen datos de pacientes.

## Conflictos de intereses

Los autores declaran no tener ningún conflicto de intereses.

## Bibliografía

1. Sackett DL, Wennberg JE. Choosing the best research design for each question. *BMJ*. 1997;315:1636.
2. Hill AB The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295-300. Disponible en: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1898525/pdf/procrsmed00196-0010.pdf>.
3. Nuzzo R. Statistical errors. *Nature*. 2014;506:150-2.
4. Oxford Centre for Evidence-Based Medicine Web site [Internet]. Explanation of the 2011 Oxford Centre for Evidence-Based Medicine Levels of Evidence (Background Document). [accesi 24/02/2015]. Disponible en: <http://www.cebm.net/index.aspx?o=5653>.
5. Harding C, Pompei F, Burmistrov D, Welch HG, Abebe R, Wilson R. Breast cancer screening, incidence, and mortality across US counties. *JAMA Intern Med*. 2015 Jul 6. doi:10.1001/jamainternmed.2015.3043.
6. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294:2203-9.
7. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2:e124.
8. Guyatt GH, Montori VM, Ferreira-González I, Busse JW, Schünemann H, Jaeschke R, Permanyer-Miralda G. Composite endpoints. En: Guyatt G, Rennie D, editores. *Users' guides to the medical literature. A manual for evidence-based clinical practice*. 2<sup>a</sup> ed. Chicago, IL: AMA Press; 2008. p. 238-48.
9. Dans AL, Dans LF, Guyatt G. Applying results to individual patients. En: Guyatt G, Rennie D, editores. *Users' guides to the medical literature. A manual for evidence-based clinical practice*. 2<sup>a</sup> ed. Chicago, IL: AMA Press; 2008. p. 273-89.
10. Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics*. 1999;15: 141-55.
11. Perea-Milla E, Vidal S, Briones E, Aguirre U, Baré M, Fernández de Larrea N, et al. Development and validation of clinical scores for visual outcomes after cataract surgery. *Ophthalmology*. 2011;118:9-16.
12. Randolph A, Bucher H, Richardson WS, Wells G, Tugwell P, Guyatt G. Prognosis. En: Guyatt G, Rennie D, editores. *Users' guides to the medical literature. A manual for evidence-based clinical practice*. Chicago, IL: AMA Press; 2002. p. 141-53.
13. Strengthening the Reporting of Observational Studies in Epidemiology [Internet] [acceso 24/2/2015]. Disponible en: [www.strobe-statement.org](http://www.strobe-statement.org).
14. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet*. 2002;359:248-52.
15. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. *Medicina basada en la evidencia*. Madrid: Churchill Livingstone España; 1997.
16. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al., GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*. 2008;336:1106-10.
17. Doll R, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ*. 1994;309:901-11.