

EVIDENCIAS EN PEDIATRÍA

Toma de decisiones clínicas basadas en las mejores pruebas científicas
www.evidenciasenpediatria.es

Fundamentos de Medicina Basada en la Evidencia

Evaluación de la importancia de los resultados de estudios clínicos. Importancia clínica frente a significación estadística

Ochoa Sangrador C
Servicio de Pediatría. Hospital Virgen de la Concha. Zamora (España).

Correspondencia: Carlos Ochoa Sangrador, cochoas@meditex.es

Palabras clave en inglés: treatment outcome; risk; evidence-based practice; statistical methods and procedures.

Palabras clave en español: resultados del tratamiento; riesgo, práctica clínica basada en la evidencia; métodos y procedimientos estadísticos.

Fecha de recepción: 14 de junio de 2010 • **Fecha de aceptación:** 14 de junio de 2010

Fecha de publicación en Internet: 24 de junio de 2010

Evid Pediatr. 2010;6:40.

CÓMO CITAR ESTE ARTÍCULO

Ochoa Sangrador C. Evaluación de la importancia de los resultados de estudios clínicos. Importancia clínica frente a significación estadística. Evid Pediatr. 2010;6:40.

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín por medio del ETOC en <http://www.evidenciasenpediatria.es>

Este artículo está disponible en: <http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2010;2;40>
©2005-10 • ISSN: 1885-7388

Evaluación de la importancia de los resultados de estudios clínicos. Importancia clínica frente a significación estadística

Ochoa Sangrador C

Servicio de Pediatría. Hospital Virgen de la Concha. Zamora (España).

Correspondencia: Carlos Ochoa Sangrador, cochoas@meditex.es

INTRODUCCIÓN

Cuando un clínico tiene que tomar decisiones sobre la eficacia de un tratamiento o el nivel de riesgo asociado a un factor de exposición y busca información en estudios publicados, tiende a juzgar la importancia de los resultados en función de su significación estadística. Sin embargo, este abordaje resulta incorrecto. La significación estadística no informa de la dimensión o importancia de los resultados, tan sólo de la probabilidad de que dichos resultados sean atribuibles al azar. Si el tamaño del efecto encontrado en un estudio resulta insignificante desde el punto de vista clínico, no importa su nivel de significación, ya que su aplicabilidad será cuestionable. De hecho, cualquier diferencia, por pequeña que sea, puede alcanzar significación estadística si el tamaño muestral del estudio es suficientemente grande.

Mientras que los criterios para juzgar la significación estadística de los resultados de estudios clínicos cuentan con suficiente consenso, no existen estándares reconocidos a la hora de establecer su importancia clínica. Además, las publicaciones biomédicas no facilitan su interpretación, ya que en ellas no se prima la presentación de resultados en forma de parámetros sencillos de fácil interpretación, sino que existe cierto culto por la utilización de estimadores o estadísticos complejos sin sentido clínico. Asimismo, muchos estudios siguen presentando los resultados con su nivel de significación (p , probabilidad de error tipo I en un contraste de hipótesis) sin los correspondientes intervalos de confianza. A diferencia del nivel de significación, que por debajo de un umbral no ofrece información de interés ($p < 0,05$), los intervalos de confianza ofrecen en sí mismos información sobre la importancia y precisión de los resultados.

Es excepcional que en el planteamiento de los estudios se hagan consideraciones sobre la magnitud del efecto a la que se atribuye importancia clínica. Sin embargo, esta información resulta fundamental, especialmente para el diseño del estudio, por lo que debe establecerse *a priori*. Aunque muchos autores mencionan la diferencia considerada en el contraste de hipótesis para el cálculo del tamaño muestral, pocas veces se hace una justificación clínica de la magnitud de dicha diferencia, surgiendo la sospecha de que haya podido ser fijada o modificada *a posteriori*, en función de los resultados. Aunque más adelante

revisaremos con detalle este aspecto, podemos decir que una diferencia clínicamente importante debería reflejarse en cambios sobre la toma de decisiones diagnósticas o terapéuticas, por parte del clínico, e idealmente en un beneficio objetivo sobre el paciente.

Es frecuente que, a la hora de resaltar la importancia de los resultados de un estudio, se emplee el término “significativo” (por ejemplo: “encontramos un significativo descenso de la duración de los síntomas”), cuando lo “significativo” no es el tamaño del efecto encontrado (por ejemplo: descenso en la duración de un síntoma de gripe de 0,5 días), sino la probabilidad de que no se deba al azar (su significación estadística). Debemos ser precisos a la hora de presentar los resultados científicos diferenciando claramente lo que es “clínicamente importante” y lo que es “estadísticamente significativo”. Para evitar confusión, parece recomendable limitar el uso del vocablo “significativo” a la indicación del nivel de significación estadístico de un contraste de hipótesis.

A la hora de interpretar la importancia clínica de los resultados de un estudio, deben tenerse en cuenta algunos factores: la variable de resultado elegida (su relación directa con el escenario clínico¹), su escala de medición (discreta, ordinal o continua), su precisión y reproducibilidad, su aplicabilidad (medidas objetivas y primarias frente a medidas subjetivas, subrogadas o compuestas) y su ámbito de interés (clínico, paciente, gestor, industria farmacéutica, autoridad sanitaria). El grado de relación de la variable de resultado con el escenario clínico será fundamental; por ejemplo, no es lo mismo comparar la eficacia de dos tratamientos en términos de mortalidad que en una escala subjetiva de valoración de síntomas. También el ámbito de interés influirá en la valoración; por ejemplo, un clínico puede asignar mayor valor a la curación, mientras que un paciente primará su calidad de vida.

En el proceso de valoración de la importancia clínica de los resultados de un estudio debemos realizar, en primer lugar, una valoración cuantitativa de los mismos. Si la magnitud del efecto es suficiente, hemos de hacer, en un segundo paso, una valoración cualitativa (rendimiento clínico). En tercer lugar, tendremos que hacer una valoración comparativa con los resultados de otros estudios o con otras variables de resultado. Superados los pasos anteriores, realizaremos una valoración de la

TABLA 1. Proceso de valoración de la importancia clínica de los resultados de un estudio

| |
|--|
| ■ Valoración cuantitativa de los resultados: estimación de medidas de efecto e impacto, evaluación de equivalencia, superioridad o no inferioridad. |
| ■ Valoración cualitativa: consideración de la repercusión sobre la toma de decisiones diagnósticas o terapéuticas y del beneficio sobre el paciente. |
| ■ Valoración comparativa: comparación con los resultados de otros estudios con las mismas o distintas medidas de efecto. |
| ■ Valoración beneficio-riesgo-coste: ponderación de beneficios con los riesgos y costes. |

relación entre beneficios, riesgos y costes (tabla 1). A continuación desarrollaremos algunos fundamentos metodológicos de interés a la hora de realizar estas valoraciones.

TAMAÑO DEL EFECTO

Los resultados de los estudios epidemiológicos deben ser expresados en forma de medidas de salud o enfermedad, con las que indicamos frecuencias, diferencias, asociación, riesgo o impacto. La forma de presentar los resultados va a depender del tipo de estudio realizado, pero sobre todo de las características de la variable o variables de interés.

A efectos de presentación y análisis podemos distinguir entre variables continuas (variables cuantitativas medidas en escalas de razones o intervalos), variables ordinales (variables cuantitativas o cualitativas medidas en escalas ordinales) y discretas (variables cualitativas medidas en escalas nominales, dicotómicas o politómicas). En función del tipo o tipos de variables implicados podremos recurrir a distintas medidas epidemiológicas.

En epidemiología, el escenario más simple lo constituye el estudio de 2 variables discretas dicotómicas. Este escenario corresponde al supuesto habitual de estudio de la asociación entre presencia-absencia de un determinado factor de exposición y presencia-absencia de enfermedad. La forma más simple de analizar estos resultados es mediante una comparación de porcentajes (por ejemplo: porcentaje de enfermos en expuestos y en no expuestos).

Otro escenario habitual es el que evalúa la asociación entre una variable discreta y otra continua. Este escenario corresponde a estudios que evalúan la repercusión de un factor de exposición (por ejemplo: tratamiento frente a placebo) sobre un efecto cuantificable en un rango continuo de valores (por ejemplo: presión arterial); en estos estudios la presentación de resultados se basará en las diferencias de medidas de tendencia (media, mediana) y dispersión (desviación típica, rango intercuartílico) de la variable continua entre grupos.

Para completar el espectro de relación entre variables, quedarían las medidas de asociación entre variables continuas (por ejemplo: asociación entre niveles de bilirrubina transcutánea y bilirrubina sérica) que se exploran con coeficientes de correlación (Pearson, Spearman, intraclass). El coeficiente de correlación presenta valores entre -1 y +1 (correlaciones

inversa y directa máximas), siendo 0 el valor de nula correlación.

Se han descrito ciertas medidas estandarizadas o ajustadas que permiten clasificar el tamaño del efecto para diferencias de medias o porcentajes². A este respecto, la magnitud de una correlación puede calcularse a partir del cuadrado del coeficiente de correlación (coeficiente de determinación) que indica la proporción de variación (varianza) de una variable que se explica por la variación de la otra variable. La utilidad de estas valoraciones resulta limitada, ya que no pueden sustituir la valoración cualitativa del rendimiento clínico del efecto encontrado. Asimismo en los estudios epidemiológicos es necesario realizar estimaciones de frecuencia, riesgo, eficacia o validez que ofrecen información cuantitativa y cualitativa con gran interés clínico, que no se contemplan en las medidas anteriormente expuestas.

MEDIDAS DE FRECUENCIA Y RIESGO

Incidencia y prevalencia son las expresiones de frecuencia de enfermedad más utilizadas en la literatura médica. Es importante distinguir entre ambas. La **incidencia** es el número de nuevos casos que han ocurrido durante un intervalo de tiempo, dividido por el tamaño de la población en riesgo al comienzo del intervalo. Esta información es habitualmente obtenida de estudios de cohortes y expresada en forma de tasas. La **prevalencia** es el número de individuos con una enfermedad o característica en un determinado punto en el tiempo, dividido por la población en riesgo en ese momento. Se calcula habitualmente a partir de estudios transversales y se expresa como una proporción.

Mientras la incidencia y prevalencia expresan frecuencia de enfermedad en grupos concretos de pacientes, otras medidas permiten comparar las diferencias de frecuencia entre dos grupos distintos o estimar el incremento de riesgo asociado a una exposición concreta (tablas 2 y 3). Las principales medidas de riesgo son el riesgo relativo y la *odds ratio* (razón de ventajas).

El **riesgo relativo** (RR) se calcula dividiendo la incidencia en el grupo de sujetos expuestos a un determinado factor de riesgo o protección entre la incidencia en el grupo no expuesto. Sólo puede ser calculado en estudios de seguimiento y mide la fuerza de la asociación entre exposición y enfermedad. Adopta valores entre 0 e infinito, menores de 1 para factores de protección y mayores para factores de riesgo; un RR de 1 supone el valor nulo (el riesgo en los 2 grupos es igual); cuanto más se

TABLA 2. Análisis de los estudios de cohortes. Medidas de riesgo e impacto

| ESTUDIOS CON SEGUIMIENTO ACUMULATIVO | | | | |
|--------------------------------------|----------------------|---------------------|-------|---|
| | Factor de exposición | | | |
| | Cohorte expuesta | Cohorte no expuesta | | |
| Enfermos | a | b | N_1 | |
| No enfermos | c | d | N_0 | |
| | M_1 | M_0 | | T |

Incidencia acumulada en expuestos: $IA_1 = a / M_1$
 Incidencia acumulada en no expuestos: $IA_0 = b / M_0$

$$\text{Riesgo relativo RR} = \frac{IA_1 \text{ expuestos}}{IA_0 \text{ no expuestos}} = \frac{a/M_1}{b/M_0}$$

| ESTUDIOS CON DATOS PERSONAS X TIEMPO | | | | |
|--------------------------------------|----------------------|---------------------|--|---|
| | Factor de exposición | | | |
| | Cohorte expuesta | Cohorte no expuesta | | |
| Casos nuevos (enfermos) | a_1 | a_0 | | |
| Total personas x tiempo | L_1 | L_0 | | T |

Personas x tiempo = personas seguidas por tiempo de seguimiento de cada una.

Densidad de incidencia (DI)

$$\text{Razón de densidades de incidencia RDI} = \frac{DI_1 \text{ expuestos}}{DI_0 \text{ no expuestos}} = \frac{a_1/L_1}{a_0/L_0}$$

MEDIDAS DE IMPACTO

Diferencia de riesgos $DR = IA_1 - IA_0 = \frac{a}{M_1} - \frac{b}{M_0}$

Proporción atribuible $PA = \frac{IA_1 - IA_0}{IA_1} = I - \frac{IA_0}{IA_1} = I - \frac{I}{RR} = \frac{RR - I}{RR}$

PA poblacional $PAP = PA \times FE_{\text{CASOS}} = \frac{RR - I}{RR} \times \frac{a}{N_1}$

FE_{CASOS} Fracción de exposición en los casos (proporción de casos expuestos)

aleje el valor de I, por arriba o por abajo, mayor será la fuerza de la asociación.

Cuando el seguimiento realizado a los sujetos incluidos en un estudio es heterogéneo, en vez de considerar la incidencia acumulada para el cálculo de riesgos, se recurre a la densidad de incidencia, en la que cada sujeto es considerado en función del tiempo que es seguido. Para calcular la densidad de incidencia se realiza un cociente en el que en el numerador se sitúa el número de casos nuevos y en el denominador el número de sujetos seguidos por el número de unidades de tiempo (habitualmente años) que es seguido cada uno de ellos. En este caso, el riesgo relativo es estimado a partir de la **razón de densidades de incidencia (RDI)** entre grupos.

En los estudios sin seguimiento longitudinal (estudios de casos y controles), como no es posible calcular la incidencia, no puede calcularse el riesgo relativo. Por ello, para estimar el riesgo se calcula la *odds ratio* (OR), que compara la *odds* de exposición (probabilidad de estar expuesto a un factor de riesgo dividida entre su complementario) en el grupo con enfermedad y la *odds* de exposición en el grupo control sin enfermedad, y que se calcula dividiendo ambas *odds*. La interpretación de la OR es similar a la del RR; I supone el valor nulo, valores menores de I indican disminución del riesgo y mayores aumentos del riesgo. Hay que tener en cuenta que sólo cuando la enfermedad estudiada es muy poco frecuente, el RR y la OR ofrecen valores similares.

TABLA 3. Análisis de estudios de casos y controles. *Odds ratio*

| Factor exposición (o riesgo) | Efecto (enfermedad) | | |
|------------------------------|---------------------|-----------|-------|
| | Casos | Controles | |
| Exposición presente | a | b | N_1 |
| Exposición ausente | c | d | N_0 |
| | M_1 | M_0 | T |

Odds de enfermedad en expuestos = a/b

Odds de enfermedad en no expuestos = c/d

$$\text{Odds ratio} = \frac{a/c}{b/d} = \frac{a \times d}{b \times c}$$

Estimación de RR a partir de OR y R_0 (riesgo basal o incidencia en grupo no expuesto) para cálculo de medidas de impacto

$$\text{RR} = \frac{\text{OR}}{(1-R_0) + (R_0 \times \text{OR})}$$

MEDIDAS DE IMPACTO

Aunque con las medidas anteriores podemos estimar el riesgo que genera un factor de exposición sobre un efecto o enfermedad, esas medidas no nos informan del impacto que dicha exposición puede originar en el conjunto de casos existentes en una población. Esta información podemos extraerla de otras medidas, como la diferencia de riesgos o la proporción atribuible (tabla 2).

Ambas medidas son estimadores del efecto absoluto que ejerce la exposición sobre la incidencia de un suceso en el grupo expuesto o en la población total. Se usan para evaluar la importancia clínica o sanitaria de una exposición y nos informan del porcentaje de incidencia que se reduciría si se eliminara el

referido factor de exposición. Son, por consiguiente, muy útiles, tanto en la clínica como en salud pública, para cuantificar el posible impacto de diferentes medidas de intervención.

La **diferencia de riesgos (DR)** se calcula restando de la incidencia en el grupo expuesto al factor de riesgo, la incidencia en el grupo no expuesto. Adopta valores entre 0 y 1 (entre 0 y 100 si se expresa en tantos por ciento), constituyendo 0 el valor nulo de ausencia de diferencias. La DR ofrece información independiente del riesgo relativo y puede variar, entre distintos grupos de pacientes, en función del riesgo propio de cada grupo. Así, podemos encontrarnos que factores que muestran un riesgo relativo muy alto apenas presenten modificaciones en la diferencia de riesgos porque el riesgo en la población (al margen de la contribución de dicho factor) sea muy bajo.

TABLA 4. Análisis de los ensayos clínicos

$$\text{Tasa de respuesta grupo de intervención} \quad P_i = \frac{\text{N}^\circ \text{ eventos grupo intervención}}{\text{total sujetos grupo intervención}}$$

$$\text{Tasa de respuesta grupo control} \quad P_c = \frac{\text{N}^\circ \text{ eventos grupo control}}{\text{total sujetos grupo control}}$$

$$\text{Reducción relativa del riesgo} \quad \text{RRR} = \frac{P_c - P_i}{P_c}$$

$$\text{Reducción absoluta del riesgo} \quad \text{RAR} = P_c - P_i$$

$$\text{NNT} \quad \text{NNT} = \frac{1}{\text{RAR}}$$

$$\text{NNT a partir de RR} \quad \text{NNT} = \frac{1}{P_c (1 - \text{RR})}$$

$$\text{NNT a partir de OR}^* \quad \text{NNT} = \frac{1 - P_c (1 - \text{OR})}{(1 - P_c) [P_c (1 - \text{OR})]}$$

*Válido para estudios prospectivos y $P_c < 0,15$; aproximado para $P_c < 0,5$. NNT: número necesario a tratar; OR: odds ratio; RR: riesgo relativo.

La **proporción atribuible (PA)**, también conocida como riesgo atribuible, fracción etiológica, fracción atribuible o población de riesgo atribuible, se define como la proporción de casos nuevos de la enfermedad, en el grupo de sujetos expuestos, que son atribuibles al factor de riesgo de interés. Se calcula dividiendo la diferencia de riesgos, anteriormente calculada, entre la incidencia en el grupo expuesto. Una extensión de esta medida es la proporción atribuible poblacional, que extiende la proporción de casos nuevos a toda la población, esto es, al conjunto de sujetos tanto expuestos como no expuestos.

Los resultados de los ensayos clínicos suelen reflejar el efecto beneficioso de intervenciones terapéuticas que reducen el riesgo en el grupo expuesto. Por ello, la diferencia de riesgos en este supuesto, conocida como **reducción absoluta del riesgo (RAR)**, se calcula en sentido contrario, restando del riesgo en el grupo control el riesgo en el grupo de intervención (tabla 4). Otra medida de impacto aplicable a estos estudios, de gran interés clínico, es el **número necesario a tratar (NNT)**, que corresponde al inverso de la RAR ($1/RAR$), y que nos informa del número de pacientes que deberían ser tratados con la intervención terapéutica para que un caso se viera beneficiado, evitando un evento desfavorable. Si aplicamos el NNT a efectos adversos asociados al tratamiento (invirtiendo la dirección del riesgo), podemos calcular el número necesario a dañar (NND) con el que ponderaríamos la relación beneficio-riesgo. En la tabla 4 podemos ver cómo calcular un NNT a partir de resultados expresados como RR u OR, forma habitual de presentar los datos agrupados en revisiones sistemáticas.

La utilidad de una prueba diagnóstica depende fundamentalmente de su validez y de su fiabilidad, pero también de su rendimiento clínico y de su coste. Si una prueba es suficientemente precisa y válida, a partir de los indicadores de validez podremos estimar su rendimiento diagnóstico, concretado en una modificación de la predicción inicial del diagnóstico (**probabilidad preprueba**) a la predicción final (**probabilidad postprueba**). Esta variación predictiva será la que nos permitirá juzgar la importancia clínica y rentabilidad de la prueba diagnóstica.

IMPORTANCIA CLÍNICA

Como se ha comentado anteriormente, resulta poco habitual encontrar en los estudios publicados comentarios sobre la magnitud del efecto que se considera clínicamente importante. Aunque esta valoración tiene un cierto componente subjetivo y no existe consenso sobre los criterios a seguir para su establecimiento, resulta fundamental a la hora de diseñar un estudio e interpretar sus resultados.

Diferentes autores han tratado de definir el concepto de “diferencia clínicamente importante”, mencionándolo la mayoría de ellos como “significación clínica”, pudiendo intuirse en la heterogeneidad de las definiciones algunos de sus componentes principales.

Hollon y Flick³ consideran que “la unidad mínima de significación clínica debería ser definida en términos del más pequeño de los cambios válidos, que sea de interés para algunas, pero no necesariamente todas las partes implicadas” en un escenario clínico.

Lindgren *et al*⁴ indican que “cuando dos métodos de tratamiento son comparados, sería clínicamente significativa la diferencia más pequeña, con respecto a una variable de resultado importante, que justificara una decisión de modificar un tratamiento”.

LeFort⁵ refiere que sería “la magnitud de cambio que supone una diferencia real para la vida de los pacientes, con un efecto duradero, aceptable para los usuarios, coste-efectiva y fácil de aplicar.

Hujoel *et al*⁶ sugieren una definición operativa de “significación clínica” como “la diferencia estadísticamente significativa, en una variable de resultado clínicamente importante, en un ensayo clínico fase III definitivo”.

Kingman⁷ establece que la significación estadística debería ser considerada una condición necesaria para la “significación clínica” y que ambas significaciones, estadística y clínica, deberían coincidir. Para cumplir este requerimiento, sería necesario que un consenso de expertos reconocidos definiera la “significación clínica”.

Killooy⁸ considera que la “significación clínica” es el resultado de una evaluación subjetiva por parte de un clínico y que antes de establecer un resultado como “clínicamente significativo” debería haber alcanzado significación estadística.

Por último, Greenstein⁹ indica que “significación clínica implica un cambio que puede alterar cómo un clínico tratará a un paciente, lo que variará en función de la situación”. En este sentido, considera que la relevancia clínica será valorada de forma diferente por parte de los clínicos, los investigadores, los pacientes, la industria farmacéutica o las agencias financiadoras, ya que pueden poner atención en diferentes tipos de resultados (por ejemplo: tamaño del efecto, coste, tiempo necesario para el tratamiento, facilidad de aplicación, duración de los resultados y grado de aceptación por los usuarios).

Podemos resumir estas definiciones recogiendo los principales componentes de lo que sería una “diferencia clínicamente importante”:

- Que debe medirse en una variable de resultado válida y con sentido clínico.
- Que debe ser valorada al menos por alguna de las partes implicadas en el proceso asistencial.
- Que justifica una modificación en el manejo del paciente.
- Que supone un beneficio para el paciente.
- Que es duradera, fácil de aplicar, aceptable (por los pacientes) y coste-efectiva.
- Que ha alcanzado significación estadística en un estudio válido (ausencia de sesgos).

- Que debe ser definida por consenso de expertos.
- Que debe ser adaptada a las circunstancias (diferente según la situación).
- Que depende de la perspectiva desde la que se juzgue: clínicos, investigadores, pacientes, industria farmacéutica o agencias financiadoras.

Aunque muchos de estos factores tienen un cierto componente subjetivo, su análisis detallado permite que los distintos agentes implicados realicen un juicio sistemático. Otras consideraciones a realizar en la valoración de la importancia clínica pueden ser tomadas de los criterios de causalidad. Un resultado clínicamente importante tendrá que ser coherente con el estado actual del conocimiento o tener algún fundamento fisiopatológico que lo respalde (plausibilidad biológica). Asimismo, se valorará la fuerza de la asociación; cuanto mayor sea el tamaño del efecto, más verosímil será. Además se tendrá en cuenta si existe algún tipo de gradiente biológico (a mayor exposición, mayor efecto), su concordancia externa (con otros estudios o medidas relacionadas) y su secuencia temporal o direccionalidad. Este último criterio resulta fundamental, ya que la exposición debe anteceder al efecto. Sólo los diseños epidemiológicos de direccionalidad anterógrada (ensayos clínicos y estudios de cohortes) permiten controlar dicha relación temporal.

Merece la pena revisar la influencia que tiene la perspectiva desde la que se evalúa la importancia clínica. El clínico valorará fundamentalmente el tamaño del efecto medido en variables objetivas y primarias, la facilidad de aplicación, la demora y la duración del efecto, aunque también tendrá en cuenta sus efectos secundarios y su coste. El paciente priorizará la resolución de los síntomas y su calidad de vida. El investigador dará importancia a diferencias mínimas que sean significativas, aunque sean basadas en medidas subrogadas o compuestas, que le permitan obtener resultados cuantificables, en los que poder apoyar su financiación. La industria farmacéutica tratará de justificar la autorización y financiación de sus productos sobre diferencias estadísticamente significativas, aunque se hayan obtenido en análisis de subgrupos o *a posteriori*, al margen de la importancia clínica de los resultados. Las agencias reguladoras se centrarán en la comprobación de la existencia de un efecto no debido al azar y en la seguridad de los productos. Las agencias proveedoras de asistencia (públicas o privadas) buscarán la financiación de productos que supongan un impacto clínico a un coste asumible y tratarán de ahorrar costes promoviendo prácticas preventivas. Por último, las autoridades de salud pública (el estado) pondrán la atención en opciones que sean accesibles a toda la población en un entorno de financiación competitivo, procurando un resultado sobre la salud pública y la productividad social.

PRECISIÓN DE LOS RESULTADOS CIENTÍFICOS

En cualquier interpretación de la importancia clínica de unos resultados, ha de tenerse en cuenta que éstos proceden de estudios en los que siempre hay un cierto grado de error. Existen dos posibles fuentes de error: el error aleatorio (falta de precisión) y el error sistemático (sesgo). No entraremos a ana-

lizar los fundamentos de los errores sistemáticos, cuya existencia cuestionaría la validez de cualquier resultado, porque excede el objetivo de esta revisión. Sin embargo, resulta trascendental tener en cuenta el error aleatorio, lo que se traduce en el grado de precisión de los resultados de los que vamos a juzgar su importancia.

En un estudio epidemiológico, la manera principal de reducir el error aleatorio consiste en aumentar el tamaño de la muestra. De esta manera, minimizamos el riesgo de obtener resultados distorsionados por azar; ya que la repetición de la medición tenderá a producir resultados distintos pero cercanos al valor verdadero del parámetro a medir. Asumiendo que la repetición de un experimento con un mismo número de pacientes puede dar resultados diferentes por mero azar, la inferencia estadística nos permite cuantificar el rango de error, a partir de medidas de dispersión de los resultados obtenidos y del tamaño muestral (error estándar). A menor dispersión de los resultados y mayor tamaño muestral, tendremos menor rango de error.

Sea cual sea el parámetro elegido para expresar los resultados de un estudio, el rango de error de nuestra estimación puede expresarse como un **intervalo de confianza** (IC), situado entre un valor por abajo y otro por arriba del resultado obtenido. Cuanto más grado de confianza queramos atribuir a nuestra estimación de error, más amplio será el rango de dicho intervalo. Habitualmente empleamos el intervalo de confianza al 95%, cuya interpretación es: “si repetimos 100 veces un experimento en las mismas condiciones y con igual número de sujetos, el verdadero valor del parámetro poblacional que queremos estimar, se encontrará incluido dentro de 95 intervalos de confianza de los 100 construidos a partir de los resultados de cada experimento”.

La presentación de la incertidumbre sobre el tamaño del efecto obtenido en un estudio en forma de intervalos de confianza es el método más recomendable. A diferencia del nivel de significación obtenido en un contraste de hipótesis (p), que sólo expresa la probabilidad de que la diferencia encontrada sea debida al azar (error tipo I o de falso positivo), el IC refleja la incertidumbre del resultado en las propias unidades de medida del efecto a evaluar, lo que facilita la interpretación de su relevancia clínica. Por otra parte, el IC permite juzgar la significación estadística del resultado, ya que sólo será significativo cuando no esté comprendido en su rango el valor nulo de ausencia de efecto (por ejemplo, 0 para diferencias de medias o riesgos, 1 para medidas de riesgo).

Frente al planteamiento frecuentista clásico de cálculo e interpretación de la precisión de los resultados científicos, existe un planteamiento alternativo, denominado “bayesiano”, en el que al cálculo de probabilidades se incorpora una estimación de la verosimilitud *a priori* de los diferentes resultados posibles de un estudio. Esta información *a priori* se obtiene del conocimiento previo procedente de otros estudios, fundamentos fisiopatológicos o de la experiencia. En el caso de que la información *a priori* no permitiera jerarquizar entre los posibles resultados, el cálculo de la precisión y, en consecuencia, de los intervalos de con-

fianza coincidiría con el del planteamiento frecuentista. Por el contrario, si se tiene cierto grado de certeza *a priori* que apoye alguno de los resultados, éste se traducirá en una modificación de los límites de los intervalos de confianza y el nivel de significación. Puede ampliarse información sobre los fundamentos e implicaciones del planteamiento bayesiano en otros textos¹⁰⁻¹².

vancia requiere su comparación con la diferencia que hemos considerado anteriormente como clínicamente importante. Si los resultados se refieren a dos opciones entre las que tenemos que elegir, esa comparación permitirá que juzguemos su equivalencia, superioridad o no inferioridad. Veamos con unos ejemplos gráficos el fundamento de esta valoración.

EQUIVALENCIA, SUPERIORIDAD Y NO INFERIORIDAD

Una vez establecida la magnitud del efecto y su grado de precisión (intervalos de confianza), la interpretación de su rele-

En la figura 1 podemos ver los IC de una serie de estudios en los que se estimaron riesgos relativos. Comparando los intervalos con el valor nulo (en este caso, RR = 1), podemos ver que sólo los estudios C, D y E son estadísticamente significativos. Sin embargo, si los comparamos con la diferencia clínicamente

FIGURA 1. Intervalos de confianza de cinco estudios con estimación de riesgo relativo. Comparación con la diferencia clínicamente importante

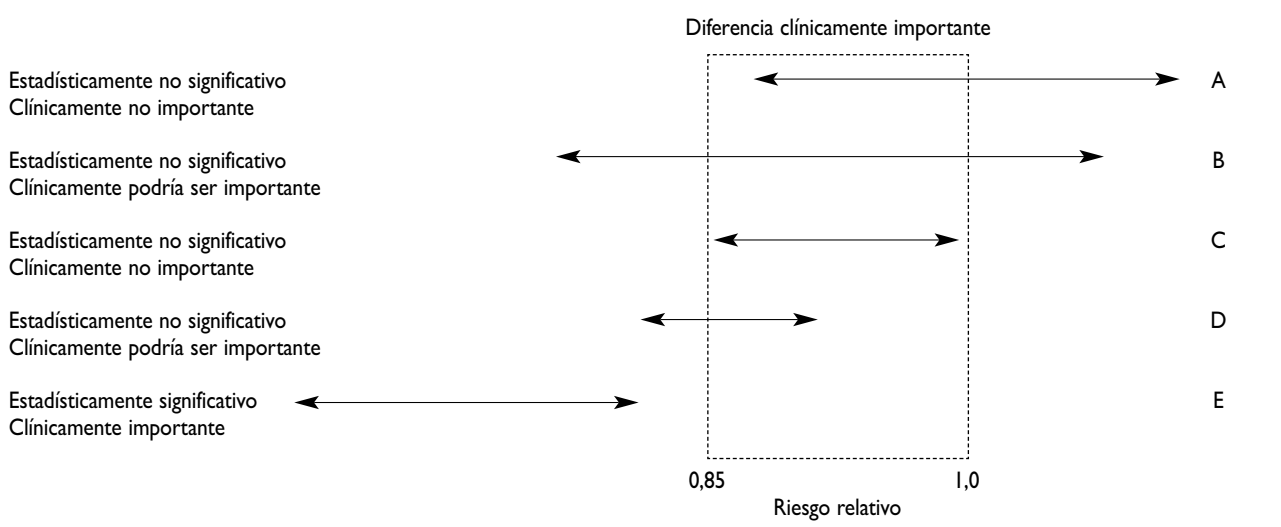
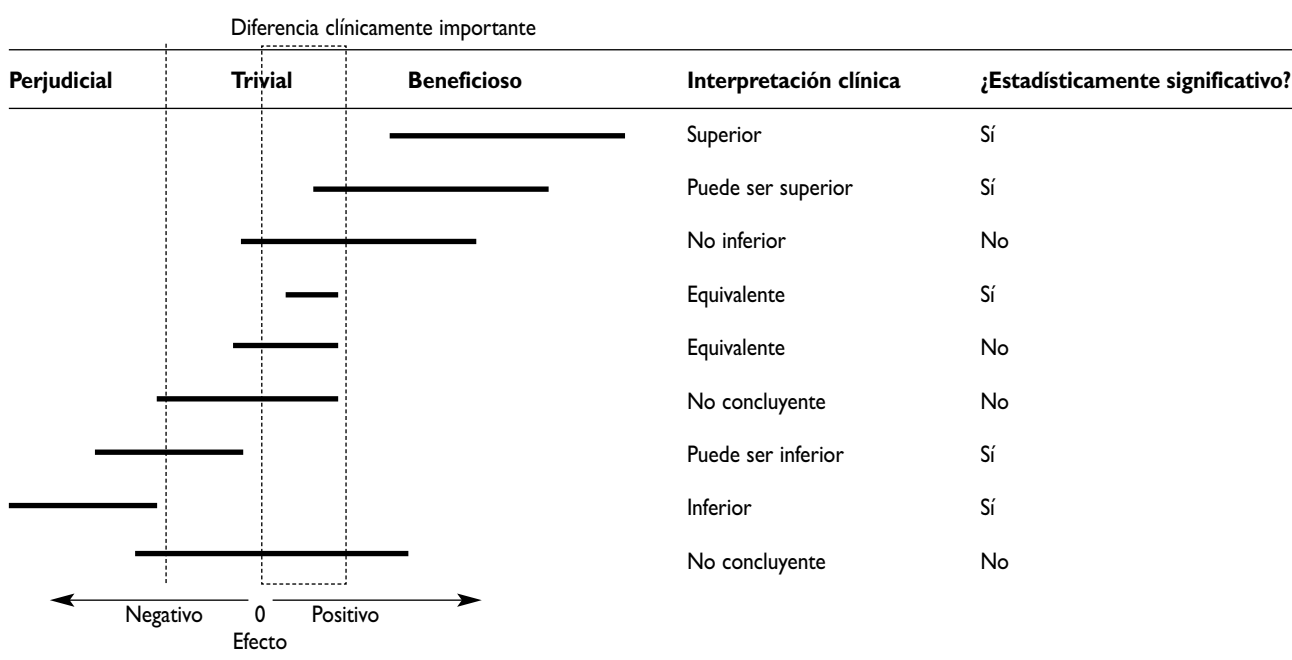


FIGURA 2. Interpretación de los intervalos de confianza de nueve estudios (datos figurados que podrían corresponder a diferencias de medias o proporciones). Valoración de significación estadística, superioridad, inferioridad, equivalencia y no inferioridad



importante, sólo el resultado E es clínicamente importante, aunque otros dos estudios (B y D) podrían serlo; el estudio B no permite realizar una interpretación clara por ser muy impreciso, mientras que en el D el efecto parece importante, ya que más de la mitad del intervalo está situado a la izquierda del límite de importancia clínica. Estas apreciaciones cualitativas pueden ser cuantificadas mediante cálculo de probabilidades, ofreciendo la probabilidad de que la diferencia sea clínicamente importante (área de la distribución de probabilidad que queda a la izquierda del límite de importancia clínica).

En la figura 2 se representan los IC de nueve estudios, de los que se puede juzgar su importancia clínica y significación estadística, comparándolos con la línea de valor nulo (en este caso podría corresponder a una comparación de medias o porcentajes, cuyo valor nulo sería 0) y con los límites de la diferencia clínicamente importante. Puede observarse cómo no todos los resultados estadísticamente significativos pueden ser considerados clínicamente importantes. Además, en función del área que ocupan los intervalos a un lado u otro de la diferencia clínicamente importante, pueden ser interpretados como superiores, inferiores, no inferiores o equivalentes. De igual manera a como mencionamos anteriormente, estas valoraciones cualitativas pueden ser cuantificadas mediante cálculo de probabilidades, para facilitar la interpretación clínica del efecto encontrado.

Es evidente que la elección de la magnitud de la diferencia considerada clínicamente importante va a ser fundamental en la valoración de los resultados. En el apartado de importancia clínica vimos la subjetividad y complejidad que conlleva el establecimiento de esa magnitud. Por ello, es preciso advertir que no podemos hacer una interpretación estadística estricta de estos cálculos, ya que ignoraríamos los múltiples factores implicados en el escenario clínico que tenemos que resolver.

PROCESO DE VALORACIÓN. UN EJEMPLO

Al inicio de este artículo se sugirieron los pasos a seguir en la valoración de la importancia clínica de los resultados de un estudio: valoración cuantitativa, valoración cualitativa, valoración comparativa y valoración de la relación entre beneficios, riesgos y costes (tabla 1). Esta propuesta no puede ser entendida como una norma, tan sólo como una guía que permite repasar los principales factores implicados en un orden lógico. Seguramente pueden hacerse propuestas alternativas que mejoren este proceso.

Veamos un ejemplo de valoración por pasos de la importancia clínica de los resultados de un estudio. Para facilitar la presentación hemos optado por una valoración simplificada¹³. Craig et al¹⁴ publicaron en 2009 un ensayo clínico en el que se evaluaba la eficacia de la profilaxis antibiótica tras una infección del tracto urinario (ITU), para prevenir recurrencias de ITU sintomáticas (medida principal de resultado). Compararon dos grupos de pacientes, uno con 12 meses de profilaxis con trimetoprima-sulfametoxazol (T-S) y otro con

placebo. El 13% del grupo con T-S frente al 19% del grupo con placebo presentaron recurrencia de la ITU (reducción absoluta del riesgo [RAR], 6%; intervalo de confianza del 95% [IC 95%], 1 a 13%; $p = 0,02$). También se encontró una reducción del 7% de ITU febriles, pero no hubo diferencias en el riesgo de daño renal a los 12 meses. El estudio presenta un buen diseño, aunque cuenta con un porcentaje de pérdidas muy importante (31%) que, pese a estar equilibrado entre los grupos, podría cuestionar los resultados, especialmente si éstos no fueran consistentes.

Para valorar la importancia sigamos los pasos sugeridos:

- Valoración cuantitativa: repasar la magnitud del efecto (variable de resultado principal), calculando si es necesario la medida de impacto más apropiada.

La profilaxis antibiótica con T-S durante 12 meses se asocia a una reducción del 6% (RAR, 0,06; IC 95%, 1 a 13%) en la recurrencia de ITU sintomática. Esta reducción equivale a que deben tratarse 14 pacientes para evitar una recurrencia (NNT, 14; IC 95%, 9 a 86).

- Valoración cualitativa: ponderar la importancia clínica del efecto. Buscar si los autores han definido la diferencia clínicamente importante (generalmente en el cálculo del tamaño muestral); en caso contrario, juzgar por nosotros mismos.

La magnitud del efecto parece discreta; de hecho, es menor que la estimación considerada por los autores en el cálculo del tamaño muestral (10%). Observemos que gran parte del IC está por debajo de esa magnitud. Por otra parte, no está claro si un NNT de 14 resulta útil, ya que a falta de otras ventajas (como sería una reducción del riesgo de daño renal), significa dar profilaxis a 14 pacientes durante 12 meses para evitar el tratamiento de ITU a un paciente.

- Valoración comparativa: buscar resultados de otros estudios publicados (idealmente revisiones sistemáticas). Comparar la medida principal de resultado con otras medidas alternativas; si otras medidas alternativas concuerdan, la importancia del efecto sale reforzada.

No se encontró efecto sobre prevención del daño renal, principal objetivo invocado en el planteamiento de la profilaxis; tampoco sobre la hospitalización. Una revisión sistemática de la Colaboración Cochrane¹⁵ cuestiona la utilidad y justificación de la profilaxis antibiótica sistemática, fundamentalmente por no prevenir el desarrollo de daño renal y tener un escaso impacto en la recurrencia.

- Valoración de la relación entre beneficios, riesgos y costes: ponderar los posibles efectos perjudiciales y la relación coste-beneficio.

La profilaxis se asocia a un aumento del 3,8% en las ITU con microorganismos resistentes a T-S. Es poco probable que el coste de la profilaxis supere al del tratamiento de un episodio de ITU.

Como vemos, esta aproximación permite introducir cierta sistemática en una valoración en la que no debe olvidarse nunca el sentido clínico de los resultados. Los números nunca han de suplantar el juicio sensato del clínico bien informado.

BIBLIOGRAFÍA

1. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490.
2. Hojat M, Xu G. A visitor's guide to effect sizes: statistical significance versus practical (clinical) importance of research findings. *Adv Health Sci Educ Theory Pract*. 2004;9(3):241-9.
3. Hollon SD, Flick SN. On the meaning if clinical significance. *Behav Assess*. 1988;10:197-206.
4. Lindgren BR, Wielinski CL, Finkelstein SM, Warwick WJ. Contrasting clinical and statistical significance within the research setting. *Pediatr Pulmonol*. 1993;16(6):336-40.
5. LeFort SM. The statistical versus clinical significance debate. *Image J Nurs Sch*. 1993;25(1):57-62.
6. Hujoel PP, Armitage GC, Garcia RI. A perspective on clinical significance (editorial). *J Periodontol*. 2000;71:1515-8.
7. Kingman A. Statistical vs clinical significance in product testing: can they be designed to satisfy equivalence? *J Public Health Dent*. 1992;52(6):353-60.
8. Killoy WJ. The clinical significance of local chemotherapies. *J Clin Periodontol*. 2002; (supplement 2):22-9.
9. Greenstein G. Clinical versus statistical significance as they relate to the efficacy of periodontal therapy. *J Am Dent Assoc*. 2003;134(5):583-91.
10. Martínez-González MA, Seguí-Gómez M, Delgado-Rodríguez M. ¿Cómo mejorar los intervalos de confianza? *Med Clin (Barc)*. 2009. doi:10.1016/j.medcli.2009.10.033 2009.
11. Burton PR, Gurrin LC, Campbell MJ. Clinical significance not statistical significance: a simple Bayesian alternative to p values. *J Epidemiol Community Health*. 1998;52(5):318-23.
12. Kaul S, Diamond GA. Trial and error. How to avoid commonly encountered limitations of published clinical trials. *J Am Coll Cardiol*. 2010;55(5):415-27.
13. Ochoa Sangrador C, Buñuel Álvarez JC. La profilaxis antibiótica podría disminuir discretamente la recurrencia de infección urinaria, aunque no está claro si nuestros pacientes se beneficiarían de ella. *Evid Pediatr*. 2009;5:84.
14. Craig JC, Simpson JM, Williams GJ, Lowe A, Reynolds GJ, McTaggart SJ, et al. Antibiotic prophylaxis and recurrent urinary tract infection in children. *N Engl J Med*. 2009;361:1748-59.
15. Williams G, Wei L, Lee A, Craig JC. Long-term antibiotics for preventing recurrent urinary tract infection in children. *Cochrane Database of Systematic Reviews* 2006, Issue 3. Art. No.: CD001534. DOI: 10.1002/14651858.CD001534.pub2.