



Significación estadística y relevancia clínica: diferentes formas de valorar críticamente los resultados de un estudio

Autora: Ana Belén Salamanca Castro 

* *Dirección de contacto:* nureinvestigacion@fuden.es

Diplomado y Grado en Enfermería. Máster en Cuidados Perinatales y la Infancia. Máster en Salud y Género. Experto en Metodología de la Investigación en Ciencias de la Salud. Directora de la revista NURE Investigación.

Estuvimos respondiendo, en el anterior editorial, la primera de las preguntas formuladas y concluimos que **obtener un valor que sea significativo desde un punto de vista estadístico (es decir, obtener un valor de $p < 0,05$) no es una garantía de calidad**. Queda por responder, no obstante, la pregunta acerca de las implicaciones que tiene, para la práctica clínica, un resultado estadísticamente significativo ya que, como Kittlein indica: “es claro que no hay una única forma de evaluar estadísticamente una hipótesis; que no hay una única forma de analizar los datos” (1). Concretamente, vamos a explicar en este editorial **¿qué implica el p-valor para la práctica clínica?**, o, dicho de otra forma, vamos a responder a la pregunta de si el valor de p informa sobre el grado o la fuerza de asociación (o de relación causal) entre las variables.

De entrada, es preciso tener claro que **un menor valor de p no significa una mayor relación de asociación o causalidad entre las variables**. Es decir, una $p < 0,01$ no se debe interpretar como una relación causal o asociación fuerte. El p-valor es una medida de probabilidad, no de impacto o de fuerza de asociación. Como ya se explicó en el anterior editorial, la significación estadística informa de la probabilidad de que los resultados del análisis realizado se deban al azar o a asociaciones o relaciones espurias. No obstante, y esta es otra falacia del p-valor, la magnitud de la diferencia que subyace entre las variables puede hacer que el valor de p sea estadísticamente significativo. Como apuntan Pita y Pértega, cuanto más grande sea la diferencia entre las dos variables, más fácil es demostrar que la diferencia es significativa (2), o dicho del otro modo: si existe una diferencia grande entre las variables estudiadas, la probabilidad de que las diferencias encontradas se deban al azar se reduce.

Por tanto, **si deseamos saber el grado en el que los resultados de los estudios pueden tener una relevancia clínica, es decir, si lo que queremos conocer es el grado en el que la variable de exposición impacta o influye en la variable resultado, la significación estadística no es suficiente**, puesto que no nos informa de la magnitud del efecto que tiene la variable independiente sobre la variable dependiente (que es lo que nos facilitará información útil para valorar la relevancia clínica). Para conocer la relevancia clínica son necesarios otros análisis, puesto que evaluar la relevancia clínica, como Pita y Pértega aseveran, “va más allá de cálculos aritméticos y está determinada por el juicio clínico” (2). Como estos autores indican, la relevancia clínica depende de la magnitud de la diferencia, de la gravedad del problema a investigar, de la vulnerabilidad, de la morbimortalidad generada por el mismo, de su coste y de su frecuencia, entre otros elementos (2).

Entre esos otros análisis que nos permiten evaluar la relevancia clínica se encontrarían las medidas de impacto o de efecto de la intervención sanitaria, que informan de la fuerza de asociación entre la exposición a un factor y la aparición de un evento indeseable (como una enfermedad o la muerte del paciente). Cuál de ellas estimar depende del diseño del estudio (y lógicamente, y relacionado con ello, depende de cuál sea la naturaleza o el origen de la asociación entre las variables). Así, las medidas a estimar serán diferentes cuando la exposición ocurre en un ambiente natural (como en el caso de estudios observacionales), o cuando se trata de una exposición artificial, deliberada, a una variable independiente que ha sido introducida por los investigadores (lo que ocurre en los estudios experimentales). Por ejemplo,

en el caso de ensayos clínicos aleatorizados, la reducción absoluta del riesgo (RAR), la reducción relativa del riesgo (RRR) y el número necesario de pacientes a tratar para reducir un evento indeseable (NNT) son las medidas que tienen mayor utilidad para la toma de decisiones clínicas. Otras medidas útiles son el riesgo relativo (RR), el riesgo atribuible (RA) o el odds ratio (OR), este último de aplicación en los estudios de casos y controles. A continuación, se expone de forma sucinta, qué estima cada una de estas medidas y cómo debemos interpretarlas (2-5):

El **riesgo relativo (RR)** es el cociente entre los sujetos expuestos a la variable estudiada (que puede ser un nuevo tratamiento o una intervención sanitaria, por ejemplo) y los no expuestos. Concretamente, relaciona la incidencia (casos nuevos) en expuestos con la incidencia de los sujetos no expuestos y por ese motivo, si el resultado es superior a uno ($RR > 1$), la exposición estudiada es un factor de riesgo, si es inferior a uno ($RR < 1$), la exposición estudiada es un factor de protección. Si su valor es próximo a uno ($RR = 1$), la exposición no se relaciona con la enfermedad.

El **odds ratio (OR)** también supone un cociente entre los expuestos a la variable estudiada y los no expuestos, pero en este caso se trata de una medición indirecta de la fuerza de la asociación, ya que no se estima partir de la incidencia, sino de casos ya diagnosticados (puesto que, como ya se ha indicado, es una estimación que se utiliza en los estudios de casos y controles). Como la lógica que subyace si es la misma que la del riesgo relativo (RR), es decir, los sujetos expuestos se encuentran en el numerador del cociente, se interpreta igual que el riesgo relativo (RR).

El **riesgo atribuible (RA)**, también denominado diferencia de incidencias (DI), diferencia de riesgos, o reducción absoluta de riesgo (RAR) cuando se calcula a partir de los datos de un ensayo clínico aleatorio, se define como la diferencia del riesgo (o incidencia) del grupo expuesto (o grupo experimental) y el riesgo (o incidencia) del grupo no expuesto (o grupo control). Un resultado igual (o muy próximo) a cero ($RA = 0$), indica que no hay asociación entre la presencia del tratamiento, intervención o exposición y el evento, mientras que un valor mayor de cero ($RA > 0$) indica que la presencia del tratamiento (intervención o exposición) se asocia a mayor ocurrencia del evento (y por tanto, es un factor de riesgo), y un valor menor de cero ($RA < 0$) indica que la presencia del tratamiento (intervención o exposición) se asocia a menor ocurrencia del evento (y por tanto, es un factor de protección o de prevención del evento). El inverso del RA ($1/RA$) informa del número de personas que es necesario que estén expuestas para que se produzca un daño o resultado adverso (NND), que puede utilizarse tanto en ensayos clínicos como en estudios de cohortes.

En general, se suele usar el riesgo atribuible (RA) cuando el riesgo en el grupo expuesto es mayor que en el grupo control. Sin embargo, cuando lo que se espera que la exposición al tratamiento prevenga o reduzca la aparición del evento indeseable (como en el caso de estudios experimentales), se utiliza la reducción absoluta de riesgo (RAR), que se explica a continuación.

La **reducción absoluta de riesgo (RAR)** calcula la diferencia entre el riesgo en el grupo control y el riesgo del grupo experimental. En este caso, un valor mayor de cero ($RAR > 0$) indica que la presencia del tratamiento (intervención o exposición) se asocia a una menor ocurrencia del evento (y, por tanto, es un factor de protección o prevención) y un valor menor de cero ($RAR < 0$) indica que la presencia del tratamiento (intervención o exposición) se asocia a mayor ocurrencia del evento (y, por tanto, es un factor de riesgo). Si el valor del RAR es próximo a cero ($RAR = 0$), no hay asociación entre la presencia del tratamiento (intervención o exposición) y el evento.

Generalmente el resultado del RAR se multiplica por 100 y se expresa en términos de porcentaje, para una mejor comprensión. De este modo se obtiene el porcentaje de personas que previenen o evitan un evento indeseable por haber recibido la exposición, intervención o el tratamiento que se está evaluando. Conociendo el RAR podemos también calcular el número necesario a tratar (NNT), por una simple regla de tres, como se explica a continuación.

El **número necesario a tratar (NNT)** indica cuántos individuos deberían ser tratados (o expuestos a la variable que se estudia) para prevenir la ocurrencia del evento indeseable. Equivale al inverso del RAR ($1/RAR$). Por tanto, el NNT informa del “esfuerzo” terapéutico que debe realizarse para prevenir un suceso indeseable adicional (3). Consecuentemente, cuanto menor es el valor del NNT, menos pacientes deben ser tratados para evitar el efecto indeseable (o la enfermedad o la muerte), y lógicamente, esto se interpreta como una mayor efectividad del tratamiento (exposición o intervención) implementada.

Como se puede deducir fácilmente, si el valor del RAR es próximo a cero (no hay asociación entre las variables estudiadas), el NNT sería infinito.

La **reducción relativa del riesgo (RRR)** es el complemento del RR, es decir, $(1-RR)$. Informa de la medida en que la intervención, tratamiento o factor de exposición en el grupo experimental ha reducido el riesgo de aparición del evento indeseable respecto al grupo control. Es una medida útil para tener una aproximación a la relevancia clínica y, de hecho, Pita y Pértega indican que valores de RRR del 25% con frecuencia son considerados como clínicamente relevantes independientemente de la significación estadística y, cuando la cifra aumenta hasta el 50%, casi siempre son considerados clínicamente relevantes. No obstante, el RRR debe ser considerado junto a la incidencia del evento que se intenta reducir en el grupo de referencia (el riesgo basal) y junto al NNT, ya que el RRR, al ser una medida relativa puede sobreestimar el efecto de la intervención (2).

En cualquier caso, al igual que sucede con las estimaciones de otros parámetros, es conveniente facilitar el resultado de la magnitud del efecto observado en un estudio (que no es más que una estimación puntual de la verdadera magnitud del efecto o asociación de interés) junto con intervalos de confianza, con el fin de informar sobre la incertidumbre que dicho parámetro presenta. De esa forma

sabremos no solo la dirección del efecto, sino también la significación estadística de los resultados, ya que si el intervalo de confianza incluye el valor nulo (el valor 1 en el caso de cocientes o el valor 0 en el caso de diferencias), los resultados no serán estadísticamente significativos y, por ende, no podremos asegurar que existe relación o asociación entre las variables evaluadas. Además, los intervalos de confianza nos informan de la precisión del efecto observado, que será mayor cuanto menor sea el rango de los posibles valores del efecto que contiene dicho intervalo.

Todo lo expuesto me lleva a pensar que **quizás sea la errónea interpretación del p-valor como una medición suficiente para la toma de decisiones clínicas uno de los motivos por los que la significación estadística lleva décadas siendo cuestionada**. Debemos ser conscientes de que, si bien la significación estadística puede ser entendida como una condición *sine qua non* para que se considere la implementación de una intervención sanitaria (puesto que si no hay diferencias significativas entre los grupos, si las diferencias son fruto del azar, no resulta lógico instaurar la intervención o el tratamiento), que sea un requisito indispensable no implica que sea suficiente, puesto que no puede ser la única fuente de información para la toma de decisiones clínicas basadas en la evidencia.

Además, considerando lo expuesto en este y en el anterior editorial, en los que se ha explicado qué estimamos realmente con el p-valor, cuáles son las limitaciones y/o falacias de este cálculo, y cómo es posible utilizarlo maliciosamente para obtener resultados estadísticamente significativos aumentando el tamaño muestral u obviando una gran diferencia entre las variables estudiadas (puesto que, al fin y al cabo, lo que se está calculando es la probabilidad de que el azar haya propiciado el resultado obtenido), no es de extrañar que cada vez sean más las voces que se cuestionan la robustez del p-valor y su idoneidad para servir como fundamento para la práctica basada en la evidencia.

Probablemente, la solución no consista en desaparecer del p-valor, sino que pase por incluir, en los estudios que responden preguntas orientadas a la toma de decisiones clínicas, estos otros análisis y mediciones para disponer de una información más adecuada para la toma de decisiones clínicas. Aun así, **no podemos prescindir del p-valor** puesto que, como se ha indicado, debemos acompañar los cálculos de la magnitud del efecto de nuestras intervenciones (exposiciones o tratamientos) de sus correspondientes intervalos de confianza para conocer, fundamentalmente, si los resultados obtenidos son estadísticamente significativos (puesto que también en el caso de las medidas que evalúan la magnitud del efecto, si no se puede descartar que el efecto observado sea fruto del azar, tampoco tendría sentido implementar el tratamiento o la intervención evaluada). En definitiva, debemos ser conocedores de las falacias, limitaciones, e imprecisiones que toda estimación puede conllevar. Pero además de estas nociones metodológicas, no debemos obviar nunca el juicio clínico del profesional, capaz de evaluar el resto de aspectos relacionados con la relevancia clínica (como

la gravedad del problema a investigar, su vulnerabilidad, la morbimortalidad generada por el mismo, su coste o su frecuencia) y que van más allá de los datos.

REFERENCIAS BIBLIOGRÁFICAS

1. Kittlein M. El ocaso de la significación estadística. *Mastozoología Neotropical*, 2019;26(2):217-219
2. Pita Fernández S, Pértega Díaz S. Significancia estadística y relevancia clínica. *Cad Aten Primaria* 2001;8:191-195. [Citado 29 jul 2025]. Disponible en: <https://www.fisterra.com/formacion/metodologia-investigacion/significancia-estadistica-relevancia-clinica/>
3. Argimón Pallàs JM, Guarga Rojas A, Jiménez Villa J. Medicina basada en la evidencia: guías y protocolos. [Citado 23 jul 2025]. Disponible en: <https://dsp.facmed.unam.mx/wp-content/uploads/2022/02/U-2-Anexo-1A-Argimon-J-Medicina-Basada-en-evidencias-guias-y-protocolos.pdf>
4. Pita Fernández S, López de Ullibarri Galparsoro I. Número necesario de pacientes a tratar para reducir un evento. *Cad Aten Primaria*. 1998; 96-98. Actualizado 26/01/2001. [Citado 18 jul 2025]. Disponible en: <https://www.fisterra.com/formacion/metodologia-investigacion/numero-necesario-pacientes-tratar-para-reducir-evento/>
5. Pita Fernández S, Vila Alonso MT, Carpente Montero J. Determinación de factores del riesgo. *Cad Aten Primaria*. 1998;4: 75-78. [Citado 18 jul 2025]. Disponible en: <https://www.fisterra.com/formacion/metodologia-investigacion/determinacion-factores-riesgo/#sec1>