

## Ensayos terapéuticos, significación estadística y relevancia clínica

La medicina basada en la evidencia tiene su mejor sustento científico en los ensayos clínicos terapéuticos de grandes dimensiones. La publicación en 1983 del artículo que hoy podríamos llamar el manifiesto de los megaensayos (“Why do we need ...”, traducido como *¿Por qué necesitamos estudios aleatorizados sencillos de grandes dimensiones?*) (1) orientado a los problemas de salud más frecuentes significó un cambio de paradigma en la manera de diseñar y llevar adelante ensayos científicos.

En forma resumida, lo que esta publicación proponía era desarrollar ensayos en miles o decenas de miles de pacientes para responder preguntas relevantes sobre enfermedades comunes. Esto permitía evaluar tratamientos de los que se esperaban efectos beneficiosos moderados (20% a 30% de reducción de la mortalidad) y que de ser exitosos tendrían repercusión epidemiológica. No es casual entonces que la gran vanguardia en este sentido han sido los estudios cardiovasculares y el infarto en particular, a través de las redes ISIS de Oxford y GISSI de Italia, y su extensión luego a las grandes redes científico-empresariales actuales. Eso llevó a los cardiólogos a familiarizarse quizás antes que en otras especialidades médicas con las nuevas nomenclaturas y formas de mirar los datos.

Aprendimos lentamente lo que quería expresar una *p* significativa y un intervalo de confianza, y luego una serie de conceptos como el riesgo relativo, reducción absoluta y relativa de riesgo, el *odds ratio* y el número que es necesario tratar. También el metaanálisis devino en una herramienta común en la práctica de la cardiología.

Los conceptos estadísticos han ido evolucionando tanto para el diseño como la para la interpretación de los resultados, aunque su difusión y comprensión por parte de los médicos es variable. En los últimos años ha comenzado un debate creciente, producto de un momento histórico diferente.

*¿Cuáles son las diferencias entre estos dos períodos y cómo están impactando en la interpretación de los ensayos clínicos?*

Cuento con la comprensión del lector al afirmar que la descripción que se presenta en la Tabla 1 es esquemática y orientada sólo a las grandes tendencias. En la década de los '80 no contábamos con tratamientos útiles para el infarto, salvo ante algunas complicaciones, y muy escasos recursos en la insuficiencia cardíaca. Tampoco existían conductas de rutina validadas de prevención secundaria. El surgimiento de la nueva forma de investigar y sus extraordinarios éxitos llevaron a incorporar un arsenal terapéutico valioso y a modificar notablemente la evolución natural en diferentes ámbitos de la patología cardiovascular. Era fácil y ético efectuar estudios contra placebo con diferentes intervenciones y se generó un extraordinario entusiasmo comunitario para participar en las redes de investigación. Así se llevaron adelante en la Argentina estudios de diseño local, como EMERAS, (2) GEMICA, (3) GESICA, (4) ENAI, (5) con miles de pacientes sin remuneración o mínima a los investigadores. El punto final incuestionable era la mortalidad o eventos en ese momento duros, como el infarto evolutivo (dolor, cambios electrocardiográficos y elevación enzimática). El hallazgo de diferencias significativas y su expresión en la nueva nomenclatura permitió una rápida traslación a la práctica de muchas de las medidas investigadas y el impacto sobre la calidad de vida y la sobrevida ha sido notable. Dadas la motivación comunitaria, las características de los problemas estudiados y las estructuras independientes, las normas reguladoras y el tiempo consumido en el llenado de papeles o historias eran mínimos.

En el momento actual existen múltiples terapéuticas validadas, de manera que es difícil efectuar estudios contra placebo y también pretender un impacto en la mortalidad por encima de un tratamiento con-

Tabla 1

	Década de los '80	2010
Patología cardiovascular. Tratamientos con impacto sobre mortalidad	Escasos	Múltiples
Redes para investigar	Comunitarias con vocación cooperativa	Científico-empresariales costosas
Puntos finales	Duros: mortalidad	Blandos: combinados (Diferentes definiciones de infartos relevantes, sangrados mayores, revascularización, etc.)
Significación estadística	Cercana a la relevancia clínica	Cuestionada su relevancia clínica

solidado con efectos similares. A su vez, el intervencionismo por catéter se ha transformado en una rutina en la terapéutica de la enfermedad coronaria y un aspecto relevante del gasto médico gira alrededor de sus resultados y complicaciones. La descripción se refiere a la cardiología, pero problemas similares abarcan casi todas las áreas de la medicina. Los estudios plantean así preguntas que no tienen el peso revolucionario de los años '80:

¿Un paciente que ya recibe luego de un infarto betabloqueantes, aspirina, inhibidores de la enzima convertidora y estatinas puede beneficiarse con un nuevo fármaco que actúa sobre un mecanismo antiaterosclerótico diferente?

¿Qué régimen antitrombótico-antiplaquetario es más eficaz durante la angioplastia en términos de infarto periprocedimiento y sangrado?

¿Es equivalente-no inferior un tratamiento determinado al previo validado: antagonistas de la angiotensina *vs* inhibidores, diferentes betabloqueantes, diferentes grupos antihipertensivos, diferentes hipoglucemiantes?

Como con la mayoría de los nuevos tratamientos no se podrá reducir la mortalidad respecto de los previos, se evalúan puntos combinados o subrogantes. Así, la mortalidad se descompone en cardiovascular y no cardiovascular para mejorar el poder del estudio, se define infarto con criterios que incrementan su incidencia y que todavía no se han podido uniformar o se obtienen beneficios en algunos puntos pero escaso impacto en otros más relevantes. Con números muy grandes de pacientes se obtienen resultados estadísticamente significativos, pero con una eventual repercusión clínica cuestionable, o sobre eventos de escaso significado clínico, como el *late-loss* en los estudios con *stents* coronarios.

Con este telón de fondo han surgido propuestas para enfocar los resultados y eventualmente el diseño de los ensayos clínicos desde una óptica diferente, orientada a consensuar, además de las medidas de significación estadística, algunos criterios sobre la relevancia clínica. En esta carta repasaré algunos conceptos básicos de la interpretación estadística de los ensayos clínicos y las características de estas nuevas propuestas.

## EL NIVEL DE P Y LA SIGNIFICACIÓN ESTADÍSTICA

La introducción de la estadística en el pensamiento médico ha sido dificultosa, en parte por un conflicto esencial e irresoluble: la necesidad implícita en el acto médico de colaborar en la atención de la salud del paciente que nos consulta hoy a partir de pruebas científicas basadas en ensayos que se afirman en conceptos probabilísticos de los grandes números y que nos dicen poco sobre este paciente individual.

El primer concepto complejo es comprender el nivel de *p* y a qué llamamos significación estadística.

La investigación comparativa de tratamientos sobre eventos clínicos se formula estadísticamente como

un testeo de hipótesis. Enunciamos una hipótesis de nulidad, que expresa que no habrá diferencias entre los tratamientos a evaluar en el estudio, y una hipótesis alternativa, que al rechazar la primera permite afirmar que los tratamientos tienen efectos diferentes.

Supongamos que los tratamientos son una droga A comparada con placebo sobre mortalidad:

Hipótesis de nulidad:

mortalidad A = mortalidad P;  
también puede expresarse como  
mortalidad A – mortalidad de P = 0

Hipótesis alternativa:

A <> B o A-P <> 0

Los ensayos clínicos trabajan con muestras y las evaluaciones estadísticas tratan de acotar el efecto del azar sobre los resultados, asumiendo un margen de incertidumbre propio del pensamiento probabilístico. Se asume que podrán cometerse errores en la interpretación y se trata de limitar estos errores al mínimo o a niveles aceptados por la comunidad.

## Error alfa

Así se establece el error alfa, en la práctica el nivel de *p* que se considerará significativo para el ensayo, habitualmente < 0,05 o < 0,01. Si en la comparación estadística resulta un valor de *p* menor que el error alfa establecido a priori, rechazamos la hipótesis de nulidad y abrazamos la hipótesis alternativa: existen diferencias “significativas” entre los tratamientos.

*El error alfa expresa la probabilidad de que al rechazar la hipótesis de nulidad podamos equivocarnos, es decir, de incurrir en considerar al estudio falsamente positivo.*

En otras palabras, si el valor de *p* fue < 0,05, estamos afirmando que rechazamos la hipótesis de nulidad, aceptamos que hay diferencias significativas entre los tratamientos y que el error en que podemos incurrir al afirmarlo es menor del 5%; hay una probabilidad menor del 5% de que una diferencia de esa magnitud pueda ocurrir por azar.

El valor de *p* no expresa algo sencillo o intuitivo y está influido por el número de la muestra: grandes diferencias en el resultado de tratamientos pueden dar una significación de *p* < 0,05 o no significativas en estudios pequeños; a su vez, diferencias pequeñas en estudios muy grandes pueden dar diferencias significativas de *p* < 0,01. *No hay relación directa entre el valor de *p* y la relevancia clínica.*

La elección del nivel de *p* < 0,05 como criterio de corte para la significación estadística es absolutamente convencional y ha recibido numerosas críticas técnicas y conceptuales. (6, 7) Sin embargo, tiene la ventaja de que su cálculo es sencillo a través de los programas estadísticos y goza de consenso en las publicaciones y entidades reguladoras: si la *p* es 0,048, el tratamiento fue útil y si resultó 0,052, no lo es. De acuerdo

con lo que comentamos anteriormente, la diferencia entre los dos resultados es sólo una variación del 4 por mil en la posibilidad de incurrir en un error estadístico, evidentemente trivial conceptualmente pero aceptada como dicotómica (éxito-fracaso) en el cuerpo de creencias actuales.

### Valor de p o intervalos de confianza

En la década de los '80 varios autores difundieron las limitaciones de la convencionalidad del nivel de p, su falta de relación con la relevancia clínica y la debilidad de transformar problemas complejos e inciertos en éxitos y fracasos por un parámetro convencional. Propusieron, lo que fue rápidamente aceptado por muchas revistas médicas, informar los resultados en términos de intervalos de confianza e incluso en lo posible omitir el valor de p. (8)

En la Tabla 2, tomada del programa Evicardio®, calculamos el impacto sobre la mortalidad de la estreptoquinasa contra placebo en el infarto de miocardio en el estudio GISSI.

Los intervalos de confianza permiten una estimación de la magnitud del efecto (*effect-size*), que es más cercana al pensamiento clínico, y que también desde un enfoque sanitario permiten estimar el eventual costo del tratamiento a través del número que es necesario tratar: en este caso, se requiere tratar 45 pacientes para salvar una vida.

El intervalo de confianza tiene un fundamento estadístico claro: los resultados de los estudios siempre tienen un margen de incertidumbre y posibilidad de error.

La traducción conceptual básica del intervalo de confianza es que existe un 95% de probabilidad de que el verdadero efecto esté abarcado por esos valores. En este caso, tenemos una confiabilidad del 95% o un margen de error del 5% al afirmar que la verdadera reducción de la mortalidad con la estreptoquinasa está entre 1,1 y 3,4 muertes cada 100 pacientes tratados. Existe una probabilidad menor del 2,5% de que sea

**Tabla 2.** Mortalidad con estreptoquinasa vs placebo en el estudio GISSI I

Intervención	Evento +	Total	Porcentaje
ESTREPTOQUINASA	628	5860	10,7%
CONTROL	758	5852	13,0%
<b>Global</b>	<b>1386</b>	<b>11712</b>	<b>11,8%</b>
	<b>p &lt; 0,001</b>	<b>IC 95%</b>	
<b>RRA</b>	2,2%	1,1%	3,4%
<b>NNT</b>	44,7	29,4	93,7
<b>RR</b>	0,83	0,75	0,91
<b>RRR</b>	17,3%	8,6%	25,1%
<b>OR</b>	0,81	0,72	0,90

Valor de p muy significativo. Expresado en términos de medidas de efecto del tratamiento y sus intervalos de confianza, tenemos una idea más acabada de la magnitud del impacto clínico: la mortalidad se reduce el 17% en forma global (IC 95% 8,6-25,1), cada 100 pacientes tratados podemos reducir 2,2 muertes (IC 95% 1,1-3,4). RRA reducción absoluta de riesgo, NNT número necesario para tratar, RR riesgo relativo, RRR reducción relativa de riesgo, OR odd ratio.

menor de 1,1 y también menor del 2,5% de que resulte mayor de 3,4.

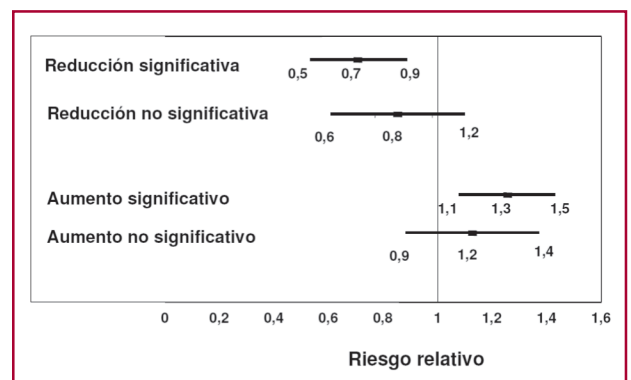
En forma gráfica, el tema se ha simplificado y nos hemos acostumbrado a un reconocimiento rápido para saber si el resultado cruza o no la línea de empate: RR = 1 cuando se trabaja con riesgo relativo, o RRA = 0 con reducción de riesgo absoluto (Figura 1).

### SIGNIFICACIÓN ESTADÍSTICA Y RELEVANCIA CLÍNICA

#### Limitaciones del valor de p y el intervalo de confianza

En la década de los '90, Alvan Feinstein y otros autores introdujeron algunas observaciones críticas sobre la limitación del valor de p y el intervalo de confianza, que han devenido actualmente en una forma diferente de enfocar los resultados de los ensayos. (9) En una de sus publicaciones más representativas afirmaron que tanto el valor de p como el intervalo de confianza referían a la incertidumbre en el rechazo de la hipótesis de nulidad, pero eran sólo dos caras de la misma moneda. Ninguna se orientaba a evaluar si el resultado era médicamente relevante, ni si tenía la magnitud para influir una conducta terapéutica en el paciente individual. Si bien los conceptos de medicina basada en la evidencia y medidas del efecto son más clínicos que el valor de p, no nos dicen mucho acerca de la relevancia de la introducción del tratamiento en la práctica.

Tomando en cuenta lo discutido en la introducción, un momento histórico en el cual proliferan trabajos con puntos combinados de diferente relevancia o estudios de no inferioridad, de grandes dimensiones con diferencias pequeñas pero estadísticamente



**Fig. 1.** Representación gráfica de los resultados comparativos de un tratamiento contra placebo respecto de un evento adverso. El rectángulo central indica el riesgo relativo (RR) observado y los brazos del intervalo de confianza de 95% (IC 95%). Si el tratamiento es superior al placebo, se expresará como un RR menor de 1 y si es perjudicial, como un RR mayor de 1. Si el IC 95% no cruza el 1, el efecto beneficioso o perjudicial es significativo (podemos asegurar que el tratamiento beneficia o perjudica). Si el IC 95% cruza la unidad, no podemos asegurar con un error menor del 5% si el tratamiento es beneficioso o perjudicial.

significativas, el tema adquiere aun una mayor importancia.

### La propuesta bayesiana

Un aporte ha surgido desde el pensamiento bayesiano proyectado al análisis de los ensayos terapéuticos. (10) La argumentación es que los valores de  $p$  y el intervalo de confianza dan una falsa sensación de seguridad frente a la lógica incertidumbre de los ensayos terapéuticos. El intento es transmitir al médico ese margen de incertidumbre a través de nuevas herramientas conceptuales.

El pensamiento bayesiano en diagnóstico nos dice que dada una prueba determinada con su sensibilidad y especificidad, al aplicarla sobre diferentes poblaciones con diferentes prevalencias podemos estimar en cuánto nos ayudará para mejorar la interpretación: las llamadas probabilidad preprueba y probabilidad posprueba cuando el resultado es positivo o negativo. En la interpretación de ensayos terapéuticos propone explicar diferentes escenarios y transformar la información en un continuo de posibilidades porcentuales. Uno de los grandes problemas de este pensamiento es establecer la "probabilidad previa", sencilla en estudios diagnósticos y hasta ahora muy difícil para ensayos terapéuticos. Los intentos de los grupos bayesianos para que la Food and Drug Administration (FDA) estadounidense acepte sus razonamientos como fundamento para la validez de los ensayos clínicos han progresado mucho en los últimos años y recientemente se ha publicado una guía oficial para su uso por parte de la industria y la propia FDA. (11) Ha llevado también a un intento conceptual de los estadísticos no bayesianos, llamados en este caso *frecuentistas*, por expresar los mismos conceptos con las herramientas del análisis clásico.

### La utilización del intervalo de confianza en el sentido bayesiano

Hemos comentado que el intervalo de confianza del 95% expresa los valores entre los cuales estimamos que se encuentra el verdadero efecto con un 95% de confiabilidad. Otra forma de interpretar el intervalo de confianza es asumir, sobre la base de los resultados, la probabilidad de diferentes efectos porcentuales, lo que ha sido propuesto por Sackett en los años recientes. (12)

Volvamos al ejemplo del GISSI:

Hemos observado una reducción de la mortalidad que redondearemos en 17% (9% a 25%). El intervalo de confianza se calcula de acuerdo con la distribución gaussiana estimando el error estándar de la reducción de mortalidad y con la fórmula:

$IC\ 95\% = \% \text{ de reducción observada} \pm 1,96 * \text{Error estándar de la reducción}$

En este caso, muy esquemáticamente, 1,96 podríamos considerarlo 2 y asumir que el intervalo de confianza es simétrico: 8% por encima y por debajo de

17%, lo que indicaría un error estándar de aproximadamente  $8\% / 2 = 4\%$ .

Conociendo la distribución gaussiana, podemos estimar diferentes puntos porcentuales de efecto y su probabilidad. Por ejemplo, si recordamos que  $\pm 1$  error estándar abarca el 68% de la probabilidad (34% por debajo y 34% por encima de la media porcentual), podríamos afirmar que existe una probabilidad menor del 16% de que el efecto sea superior a 21% o menor de 13% (calculando  $17\% \pm 4\%$ , es decir, porcentaje observado  $\pm 1$  error estándar).

### RELEVANCIA CLÍNICA Y SIGNIFICACIÓN ESTADÍSTICA

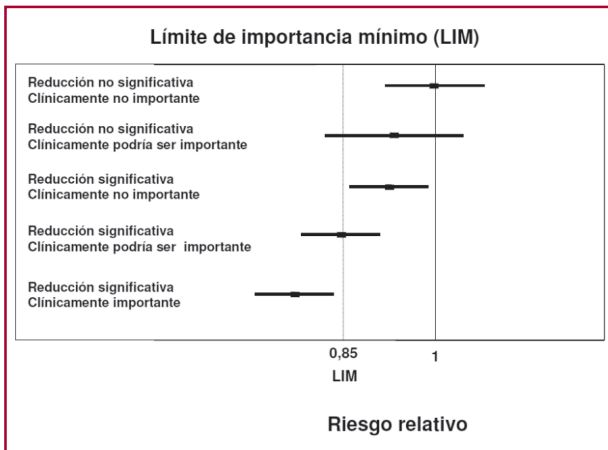
En febrero de 2010, Sanjay Kaul y George Diamond propusieron una nueva forma de graficación de los resultados, tratando de aproximarse al concepto de relevancia clínica. (13) Queda claro que la importancia o relevancia de una intervención terapéutica está fuertemente influida por el contexto del problema, sus costos, terapéuticas alternativas, riesgos implícitos, de manera que es imposible establecer un porcentaje "lógico" o deseable. En la idea de los creadores del paradigma de los ensayos de grandes dimensiones se hablaba de efectos moderados, 20-30%, sobre eventos muy relevantes. No es lo mismo bajar el 30% la mortalidad que el infarto periprocedimiento diagnosticado por elevaciones mínimas de troponina o la necesidad de una intervención tardía que no conlleva mortalidad.

A pesar de estas limitaciones, el esquema es válido como herramienta conceptual. Supongamos, como proponen los autores, que establecimos para un problema particular un nivel mínimo de impacto que consideraremos importante, en este caso una reducción del 15% de un evento. El riesgo relativo considerado efecto mínimo clínicamente importante (en inglés abreviado como MCID, que en castellano podríamos llamar límite de importancia mínimo, LIM) resulta así de 0,85. En la parte A de la Figura 1 del trabajo mencionado, (13) los autores proponen el esquema que aquí reproducimos en la Figura 2.

De arriba abajo en la Figura 2, el primer estudio no obtuvo una reducción estadísticamente significativa, es decir, no pudo descartar la hipótesis de nulidad expresada en este caso por el  $RR = 1$ . El brazo de su intervalo de confianza se aleja también del margen mínimo de relevancia clínica, LIM, establecido como  $RR\ 0,85$ , es decir, reducción del 15% del evento.

El segundo estudio no rechaza la hipótesis de nulidad, pero en su intervalo de confianza abarca el LIM. Sugeriría la necesidad de una evaluación con un número mayor de pacientes.

El tercer estudio tiene un intervalo de confianza estrecho, y si bien rechaza la hipótesis de nulidad, su beneficio sobre el evento es reducido y no se acerca al LIM que hemos establecido como relevante. La posibilidad de lograr una reducción mayor que el LIM es menor del 2,5% porque está por fuera de una de las dos colas del intervalo de confianza del 95%.



**Fig. 2.** Se han graficado cinco ensayos clínicos, con su riesgo relativo e IC 95%. Además de la vertical de RR = 1 se dibujó una línea en el RR 0,85, considerado aquí como límite de importancia mínima aceptable. Explicación en el texto. Modificado de cita 13.

El cuarto estudio rechaza la hipótesis de nulidad y abarca el LIM, por lo que podría ser relevante de acuerdo con nuestro criterio clínico. En este caso podemos estimar que la probabilidad de lograr una reducción superior al LIM es del 50% (el efecto central coincide con 0,85). Como se ve, conociendo los resultados del estudio y la magnitud de la reducción deseada, podemos calcular su probabilidad.

El quinto estudio no sólo rechaza la hipótesis de nulidad, sino que asegura reducir el evento por encima de nuestro LIM, es decir, nos brinda una respuesta estadísticamente significativa y clínicamente relevante para nuestro criterio establecido a priori.

**Estimación de la relevancia clínica y herramientas informáticas**

En su artículo, los autores toman el ejemplo del estudio TACTICS-TIMI 18 (14) para analizar los efectos de la reducción de la incidencia del evento combinado. En el protocolo se había estimado para el cálculo de la muestra una reducción esperada del 25%. La reducción real fue del 18% (IC 95% 2-32%), estadísticamente significativa con una p de 0,028. Calcularon que de acuerdo con los resultados, la probabilidad de alcanzar una reducción del 25%, considerada a priori como valedera, fue de sólo el 17%. Un elemento más para juzgar la relevancia del ensayo, que no surge directamente de la observación del intervalo de confianza. Los autores efectúan ese cálculo sobre la base del razonamiento bayesiano.

Lo mismo puede obtenerse de la estadística convencional (“frecuentista”). He construido una planilla que quedará como material adjunto de este trabajo y puede bajarse de <http://www.sac.org.ar/web/es/revista-argentina-de-cardiologia>. En la Figura 3 se desarrolla un ejemplo de su uso.

Ingresamos los datos del estudio TACTICS-TIMI 18 en las casillas sombreadas de la tabla de 2x2. El

	Evento combinado		total	%
	(+)	(-)		
Agresivo	177	937	1114	15,9%
Conservador	215	891	1106	19,4%
Global	392	1828	2220	17,7%

	IC 95%		
RRA	3,6%	0,4%	6,7%
NNT	28,2	14,9	262,6
RR	0,82	0,68	0,98
RRR	18,3%	2,1%	31,8%
OR	0,78	0,63	0,97

% reducción relevante	RR hipotético	Probabilidad de esa reducción
10%	0,90	85%
15%	0,85	66%
20%	0,80	41%
25%	0,75	18%
30%	0,70	5%
35%	0,65	1%
1%	0,99	98%
2%	0,98	98%
3%	0,97	97%
5%	0,95	95%

**Fig. 3.** Tabla para el análisis estadístico de un ensayo clínico. Ingresando los datos en las cuatro casillas sombreadas de arriba, calcula en forma automática la significación estadística de la diferencia en los efectos del tratamiento, las medidas del efecto y sus intervalos de confianza. Más abajo, en los casilleros sombreados pueden establecerse diferentes porcentajes de reducción del evento como LIM y la planilla indica la probabilidad de esa reducción.

evento combinado se presentó en 177/1.114 pacientes con el tratamiento agresivo inicial y en 215/1.106 pacientes del grupo conservador. La planilla calcula los porcentajes, las medidas del efecto y su intervalo de confianza y, más abajo, la probabilidad en este estudio de alcanzar una reducción de porcentajes diferentes del evento, que pueden cambiarse. En este caso se utilizaron 10-15-20%, etc. Como se ve, la probabilidad de una reducción del 25% estimada fue del 18% (17,6% con un decimal), muy similar al 17% calculado por los autores con métodos bayesianos.

**COMENTARIOS FINALES Y CONCLUSIONES**

La medicina basada en evidencias conceptualmente nos aporta herramientas para evitar la aplicación de medidas sin efecto demostrado y seleccionar las terapéuticas para las que existen mejores pruebas. Abordar la búsqueda de evidencias en forma individual, a pesar del imaginario de Sackett y colaboradores, es prácticamente imposible por razones de tiempo y de la complejidad creciente en la interpretación de las publicaciones.

El médico práctico debe recurrir a guías o consensos comunitarios que deberían evaluar la información en forma seria y responsable, pero que en la práctica están fuertemente influidos por conflictos de intereses que han llevado a que gran parte de las recomendaciones resulten opinables o fundamentadas en evidencias débiles.

La superación de estas limitaciones es sin duda una tarea también colectiva. Las nuevas propuestas de analizar la información desde la mirada de la relevancia e importancia clínica son un aporte en tal sentido, con criterios necesariamente subjetivos pero más cercanos a las decisiones del médico frente al paciente.

La intención de esta carta ha sido acercarnos a este pensamiento, que se encuentra hoy en sus comienzos pero será seguramente muy fructífero, y aportar una herramienta informática que espero facilite la aplicación de este concepto a la lectura de ensayos clínicos.

Establecer la relevancia e importancia clínica es una tarea comunitaria, que quizá condicione el diseño futuro de ensayos más orientados al paciente que a la intervención o al fármaco.

**Dr. Carlos D. Tajer**

Director de la Revista Argentina de Cardiología

---

## BIBLIOGRAFÍA

1. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984;3:409-22.
2. EMERAS (Estudio multicéntrico estroptoquinasa Repúblicas de América del Sur) Collaborative group. Randomized trial of late thrombolysis in patients with suspected acute myocardial infarction. *Lancet* 1993;342:767-72.
3. Elizari M, Martinez JM, Belziti C, et al. Morbidity and mortality following early administration of amiodarone in acute myocardial infarction. GEMICA study investigators, GEMA Group, Buenos Aires, Argentina. Grupo de Estudios Multicéntricos en Argentina. *Eur Heart J* 2000;21(3):198-205.
4. Doval H, Nul D, Grancelli H, et al; for the Grupo de Estudio de la Sobrevida en la Insuficiencia Cardíaca en Argentina (GESICA) Investigators. Randomised trial of low dose amiodarone in severe congestive heart failure. *Lancet* 1994;344:493-8.
5. Tajer C, Grancelli H, Hirschson Prado A, et al. Enalapril in unstable angina: a randomized double blind multicentre trial. *Eur Heart J* 1995;16:259. Enalapril en la angina inestable. Estudio multicéntrico ENAI. *Rev Argent Cardiol* 1996;64:31-47.
6. Lang J, Rothman K, Cann C. The confounded p value. *Epidemiology* 1998;9:7-8.
7. Ioannidis J. Why most published research findings are false. *PLoS Med* 2005;2:696-701.
8. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J* 1986; 292:746-50.
9. Feinstein A. P-Values and Confidence Intervals: Two Sides of the Same Unsatisfactory Coin. *J Clin Epidemiol* 1998;51:355-60.
10. Diamond GA, Kaul S. Bayesian approaches to the analysis and interpretation of clinical megatrials. *J Am Coll Cardiol* 2004;43: 1929-39.
11. Campbell G. Guidance for Industry and FDA Staff Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials . 2010. Puede obtenerse en: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf>
12. Sackett DL. Superiority, equivalence and noninferiority trials. En: Haynes RB, Sackett DL, Guyatt GH, Tugwell P, editors. *Clinical Epidemiology*. 3<sup>rd</sup> ed. Lippincott Williams & Wilkins; 2006.
13. Kaul S, Diamond G. Trial and Error How to Avoid Commonly Encountered Limitations of Published Clinical Trials. *J Am Coll Cardiol* 2010;55:415-27.
14. Wiviott SD, Braunwald E, McCabe CH, et al, for the TRITON TIMI 38 Investigators. Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med* 2007;357:2001-15.