Servicio Extremeño de Salud
Oficina de Evaluación de Medicamentos
Servicio Extremeño de Salud
Programa de la Consejería de Salud y Política Social

evalmed
evaluación de medicamentos

¿Es clínicamente relevante además de estadísticamente significativo?

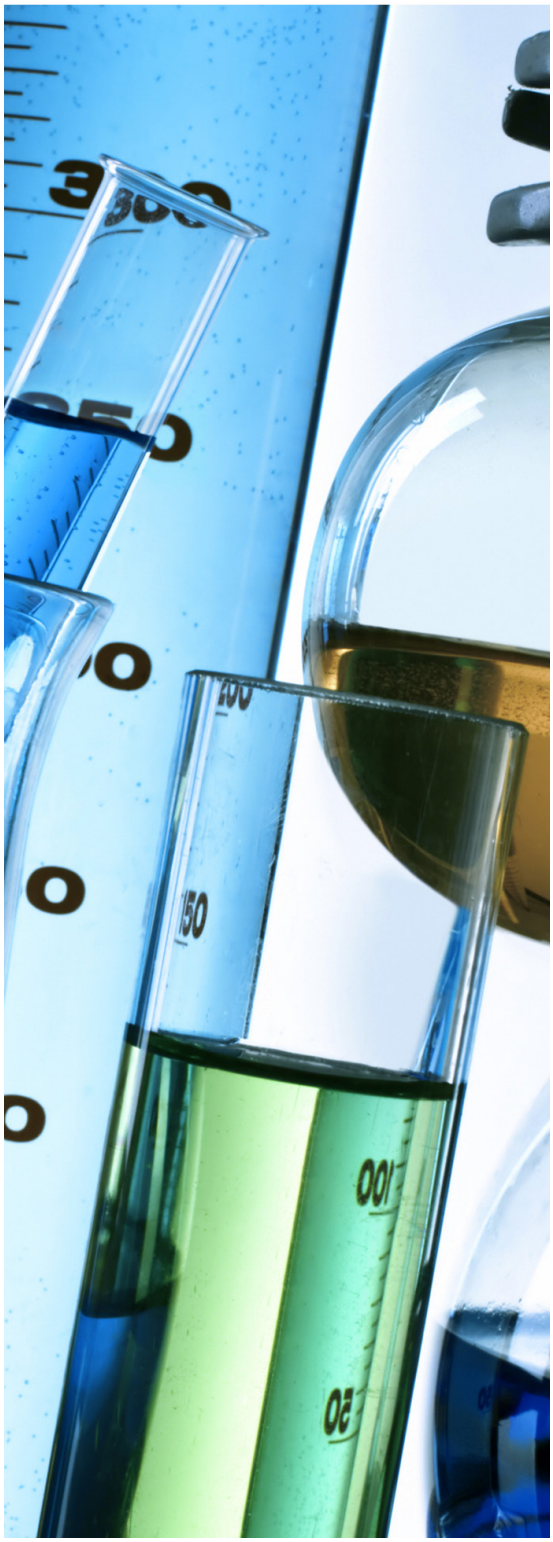
CONCLUSIONES

Los médicos, pacientes, autoridades y gestores incurren en el *efecto marco* cuando toman decisiones basándose únicamente en los valores del RR o de la reducción relativa del riesgo (RRR), porque sobreestiman el efecto. Como consecuencia, se produce un incremento significativo en la propensión de los médicos a prescribir, del deseo de los pacientes para que se les prescriba y de la inclinación de las autoridades y gestores sanitarios a financiar. Conviene destacar que la "Efectividad" de las vacunas es la RRR.

La *relevancia clínica* de una intervención de prevención de un riesgo grave, frente a la no intervención o frente a una intervención control, no se puede estimar con el valor de la *p*, ni con el RR, sino con la RAR, y aún mejor con su inverso el NNT.

***El balance de beneficios* y *riesgos* de una misma variable no se puede estimar, y aún menos calcular, con sus respectivos RR sino con la RAR del beneficio y la RAR del riesgo añadido.**

La toma de decisiones compartida mejora cuando se conoce el riesgo absoluto (RA) del grupo control, el RR, la RAR y el NNT con sus respectivos intervalos de confianza. En caso de que no estén disponibles estas medidas del efecto, hay sencillas calculadoras para obtenerlas. Una de ellas es la que proporcionamos en la web evalmed.es, pestaña HERRAMIENTAS.



ABREVIATURAS

AAR: aumento absoluto del riesgo; **ACV**: accidente cerebrovascular; **CV**: cardiovascular; **IAM**: infarto agudo de miocardio; **NND**: número necesario a tratar para dañar a 1 paciente; **NNT**: número necesario a tratar para evitar 1 evento, **RA**: riesgo absoluto; **RAR**: reducción absoluta del riesgo; **RR**: riesgo relativo; **RRR**: reducción relativa del riesgo.

NOTA: Todos los intervalos están calculados para un 95% de confianza (IC 95%), salvo que se indique expresamente otro porcentaje.

Boletín Terapéutico Extremeño (BTE)

Álvarez-Cienfuegos A, Montaño Barrientos A, Baquero Barroso MJ, Rubio Núñez PL, Candela Marroquín E, Gavilán Moral E, Gómez Santana MC, Sánchez Robles GA. ***¿Es clínicamente relevante además de estadísticamente significativo?*** Boletín Terapéutico Extremeño. 2012 Sep;3(3). Disponible en la web de la Oficina de Evaluación de Medicamentos SES: http://evalmed.es

Conflicto de intereses: Todos los autores declaran no tener conflictos de intereses.

Página web: evalmed.es **Director: Galo Agustín Sánchez Robles**

Depósito legal: GR574-2009 / ISSN: 1889-3554

¿Es clínicamente relevante además de estadísticamente significativo?

INTRODUCCIÓN

En los ensayos clínicos debemos entender por efecto tanto los "beneficios" como los "daños añadidos" en el grupo de intervención frente al grupo de control. La magnitud del efecto se mide comparando el número de pacientes que sufren un evento en el grupo de intervención frente al número de pacientes que sufren un evento en el grupo de control.

Imaginemos que los 100 habitantes de una pequeña isla de Micronesia enferman por la picadura de un insecto. Y, además, que provocará la muerte a 10 de ellos en un año. Esto significa que el riesgo basal de muertes es de 10 por cada 100 enfermos[1] ; es decir, el +10% en un año. En la isla hay una planta (un antídoto) que evita 3 muertes por cada 100 enfermos. El "beneficio" en muertes, por tanto, es del -3% en un año. Pero esta planta también provoca 1 muerte al año por paro cardíaco de cada 100 enfermos. El "daño añadido" en muertes, por tanto, es del +1% por año. El balance se establecería así: "% de riesgo basal" menos "% de beneficios" más "% de riesgos añadidos" = +10% - 3% + 1% = +8%.

Si tuviéramos que enseñar a un niño estos fenómenos, tendríamos que hacerlo así porque es la forma natural de aprendizaje (1). Y expresaríamos el "beneficio" con signo negativo porque es restar una parte del riesgo basal del que se parte, y pondríamos el "daño añadido" con signo positivo porque es añadir una nueva cantidad de riesgo al riesgo basal. Sin embargo, la forma de aprendizaje en las profesiones sanitarias invierte los signos, expresando los beneficios con signo positivo y los daños añadidos con signo negativo, porque se les otorga una entidad propia e independiente al desvincularlos del riesgo basal. Y así, tomando el ejemplo anterior, el sanitario suele decir que el beneficio de la intervención es +3% y el daño añadido el -1%, con un balance de +3% -1% = +2%, lo cual sólo tiene sentido cuando deliberada o indeliberadamente están disociados del riesgo basal +10%. Y, si bien los profesionales sanitarios, reconocen este patrón cuando buscan la referencia del riesgo basal, otros lectores pueden creer que los beneficios originan nuevas acciones, cuando los fármacos nada crean, sino que modifican (estimulando o deprimiendo) las acciones propias del organismo.

A pesar de estas premisas, nosotros no utilizaremos la forma natural, sino la universalmente difundida forma artificial de expresar los "beneficios" con signo positivo y los "daños añadidos" con signo negativo. Asimismo, por ser más habitual en la literatura biomédica, utilizaremos la expresión "riesgos añadidos" en lugar de "daños añadidos".

1 El riesgo absoluto se puede expresar en frecuencia absoluta (10 de cada 100 enfermos), pero también puede expresarse en probabilidad, casos favorables dividido por casos posibles: 10/100 = 0,1 (que es un tanto por uno). Asimismo puede expresarse en porcentaje: 10/100 = 10%. Los cálculos se hacen fácilmente con la probabilidad y con el porcentaje.

MEDIDAS DE LA MAGNITUD DEL EFECTO

Para el ejemplo sirve el ensayo clínico "4 S", un estudio aleatorizado y controlado en prevención secundaria, que comparó a un grupo de intervención con 2.221 pacientes que tomaron simvastatina frente a un grupo de control con 2.223 que tomaron un placebo.

Nos fijaremos únicamente en la variable "Muertes por cualquier causa", de las que en 5,4 años hubo 182 en el grupo de simvastatina frente a 256 en el grupo de placebo. Con este ejemplo calcularemos las medidas de la magnitud del efecto contabilizando todos los eventos acumulados en los 5,4 años (2).

- Riesgo Absoluto o Incidencia grupo de intervención: RAi = 182 / 2221 = 8,2%
- Riesgo Absoluto o Incidencia grupo control: RAc = 256 / 2223 = 11,5%
- Riesgo Relativo: RR = RAi / RAc = 8,2% / 11,5% = 0,71
- Reducción Relativa del Riesgo: RRR = 1 – RR = 1 – 0,71 = 0,29 (= 29%)
- Reducción Absoluta del Riesgo: RAR = RAc – RAi = 11,5% - 8,2% = 3,32%
- Número Necesario a Tratar para evitar 1 evento: NNT = 100 / RAR = 100 / 3,32 = 30

Situamos estos cálculos en la tabla 1.

El Riesgo Absoluto (RA) o incidencia es la probabilidad que tiene un sujeto de sufrir un evento en un período de tiempo.

El Riesgo Relativo (RR) es la ratio entre el riesgo absoluto del grupo de intervención en el numerador y el riesgo absoluto del grupo de control en el denominador. Y se puede interpretar como el número de casos incidentes en el grupo de intervención por cada uno que ocurre en el control, y así RR 0,71 nos dice que inciden 0,71 casos por cada 1 que incide en el control. Pero quizás sea más fácil decir que inciden 71 casos en el grupo de intervención por cada 100 que inciden en el control. Esta última forma de explicarlo nos lleva directamente a la Reducción Relativa del Riesgo (RRR), que son los casos que evita la intervención

Tabla 1: Hombre / Mujer 58 años, con historia angina o infarto agudo de miocardio, y colesterol y triglicéridos elevados						
Estudio 4S; seguimiento 5,4 años	Simvastatina n_i= 2221	Placebo n_c= 2223	Cálculos por riesgos acumulados			
	RA_i = eventos/n_i	RA_c = eventos/n_c	RR = RA_i / RA_c	RRR = 1 - RR	RAR = RA_c - RA_i	NNT = 100% / RAR
Mortalidad por cualquier causa	182 / 2221 = 8,2%	256 / 2223 = 11,5%	8,2% / 11,5% = 0,71	1 - 0,71 = 0,29	11,5% - 8,2% = 3,32%	100 %/ 3,32% = 30

RA_i: Riesgo Absoluto del grupo de intervención; **RA_c**; Riesgo Absoluto del grupo de control; **RR:** Riesgo Relativo; **RRR:** Reducción Relativa del Riesgo; **RAR:** Reducción Absoluta del Riesgo; **NNT:** Número Necesario a Tratar para evitar 1 evento; **n_i:** número de pacientes totales del grupo de intervención; **n_c:** número de pacientes totales del grupo dec control.

por cada 100 casos que inciden en el control, a saber 100 - 71 = 29: es decir el 29%. Y se puede calcular más fácilmente como el complementario del RR así: RRR = 1 - RR = 1 - 0,71 = 0,29 (= 29%).

La Reducción Absoluta del Riesgo (RAR) es la diferencia entre el riesgo absoluto del grupo de control (RAc) menos el riesgo absoluto del grupo intervención (RAi). El Número Necesario a Tratar (NNT) de una intervención expresa el número de pacientes que hay que tratar para evitar un evento más que con el control, y se puede calcular mediante una regla de tres simple así: Si evito 3,32 eventos tratando a 100 pacientes, entonces para evitar "1" evento necesito tratar a 100 x 1 / 3,32 = 30, aunque puede hacerse más rápidamente calculando el inverso del RAR así: 100% / 3,32% = 30. Un NNT 30 significa que simvastatina evita una "muerte por cualquier causa" por cada 30 pacientes tratados, y también significa que es eficaz en 1 paciente por cada 30 pacientes tratados durante 5,4 años, pero no eficaz en 29 de cada 30.

Cuando el medicamento de intervención protege más que el medicamento de control, el RR es menor de 1, y por tanto la RAR y el NNT serán positivos.

Cuando el medicamento de intervención daña más que el medicamento de control, el RR es mayor de 1, y por tanto la RAR y el NNT serán negativos. Y si bien se comprende fácilmente un RR mayor de 1, se concibe con dificultad una RAR y un NNT con signos negativos, de modo que para mejorar la comprensión se cambia el signo negativo por un signo positivo, pasando a denominarse respectivamente "Aumento Absoluto del Riesgo (AAR)" y "Número Necesario para Dañar (NND)".

Cuando el RR = 1, y por tanto la RAR = 0, significa que no hay diferencia entre el medicamento de intervención y el de control, porque el porcentaje de eventos de ambos es el mismo, y al dividirlos da uno (RR = 1) y al restarlos da cero (RAR = 0).

EL INTERVALO DE CONFIANZA DEL PROMEDIO MUESTRAL "p_i" DE UNA MUESTRA (m_i) TOMADA ALEATORIAMENTE CONTIENE EL PROMEDIO POBLACIONAL "P" DE LA POBLACIÓN TOTAL

A modo de ejemplo, en el estudio RELY los 6.022 pacientes del grupo de warfarina constituyen una muestra (m_i) aleatoriamente tomada de la Población Mundial de pacientes con fibrilación auricular no valvular.

Supongamos que la Población española de pacientes con fibrilación auricular no valvular que toman anticoagulantes es de 350.000 personas. A nosotros lo que nos interesa es la tasa de accidente cerebrovascular (ACV) que tiene en un año la Población española con fibrilación auricular. Si estuviera en nuestra mano, seguiríamos a estas 350.000 personas durante un año para registrar la incidencia de ACV en ese período. Imaginemos que registramos 5.250 ACV, entonces el Promedio Poblacional "P" resultaría 5.250 / 350.000 = 1,5% en un año. Este "P" es el riesgo absoluto o incidencia basal, no tiene error estándar, es 1,5% y nos da toda la información con esa única cantidad. Podríamos afirmar que en la Población española con fibrilación auricular el Promedio Poblacional de ACV es 1,5%, (P = 1,5% en un año), que también podemos ver escrito como el verdadero valor del parámetro.

Pero nosotros no podemos tomar a los 350.000 pacientes que nos interesan, sino que, basándonos en el Teorema del Límite Central (Lyapunov, 1901), tomamos una muestra aleatoria "m_1" con 6.022 pacientes de la Población española de 350.000 pacientes. Con la muestra aleatoria "m_1" nosotros vamos buscando el Promedio Poblacional "P" de la Población española con fibrilación auricular, que es lo que nos interesa. Si "m_1" es verdaderamente una muestra aleatoria tomada de la Población española, entonces podremos acercarnos al conocimiento del Promedio Poblacional "P". Y para eso hacemos un seguimiento

¿Es clínicamente relevante además de estadísticamente significativo?

de un año a nuestra muestra "m1" de 6.022 pacientes. Al final del año registramos 97 ACV. El promedio muestral "p_1" sería 97 / 6.022 = 1,61%. Pero no es ese promedio muestral p_1 = 1,61% lo que estamos buscando, sino el Promedio Poblacional "P". Ahora bien, como suponemos que "m1" es una muestra aleatoriamente tomada de la Población española, su promedio muestral puede completarse con los intervalos de confianza al 95%, de modo que p_1 = 1,61% (IC 95%, 1,29% a 1,93%).

Ahora bien, si repitiéramos la extracción de la muestra 100 veces, en las mismas condiciones, es muy probable que en 95 veces el promedio poblacional "P" se encuentre situado entre 1,29% y 1,93%. Podemos decir también que tenemos una confianza del 95%, de la misma manera que tenemos una confianza del 2,5% de que será menor del 1,29% y una confianza del 2,5% de que será mayor del 1,93%.

En realidad a nosotros no nos interesa el promedio muestral p_1 = 1,61% de la muestra "m1". A nosotros lo que nos interesa es el Promedio Poblacional "P" de la Población española, también llamado el verdadero valor del parámetro.

Volviendo al caso real del estudio RELY, recordemos que los investigadores tomaron una muestra aleatoria (m1) de 6.022 pacientes representativos de la Población Mundial de pacientes con fibrilación auricular no valvular, y, tras un año, registraron 94 ACV. El promedio muestral fue p_1 = 94 / 6.022 = 1,56% (IC 95%, 1,25% a 1,87%). Con esta muestra (mi) los investigadores van buscando el Promedio Poblacional "P" de toda la Población Mundial de pacientes con fibrilación aurículas no valvular, y pueden afirmar con un 95% de confianza que "P" se encuentra entre 1,25% y 1,87%. Y puede estar tanto en el 1,25%, como en el 1,26%, como en el 1,29%, como en el 1,56%, como en el 1,86%, como en el 1,87%. Puede estar en cualquiera de todos los valores del intervalo de confianza. Por tanto, no debe interpretarse que la estimación puntual 1,56% está más cerca del Promedio Poblacional "P" que todos los demás valores del intervalo de confianza.

Si los investigadores toman 100 muestras aleatorias (p_i) de 6.022 sujetos para hacer cien veces el experimento, lo más probable es que tengan el Promedio Poblacional "P" que van buscando entre los intervalos de confianza de 95 promedios muestrales (p_i), pero en los 5 restantes no. Vale la pena dar por bueno el acertar en 95 de cada 100 veces cuando tenemos delante de nosotros un promedio con su intervalo de confianza. Aunque hay excepciones, este 95% es el mínimo que se asume en la metodología de la investigación, pero en muchas ocasiones es necesario arriesgarse sólo en el 1% estableciendo una exigencia del 99% de confianza.

El intervalo de confianza al 95% de una proporción (p_i) es igual a la proporción ± Z multiplicado por el error estándar de la proporción. Calculamos el intervalo de confianza al 95% de los

Cálculo por riesgos acumulados de RR, RAR, NNT con sus IC 95%, potencia estadística y valor de p

Abreviaturas: **RA**: Riesgo Absoluto; **RR**: Riesgo Relativo; **RAR**: Reducción Absoluta del Riesgo; **NNT**: Número Necesario a Tratar para evitar un evento; **IC 95%**: intervalo de confianza al 95%

Escribimos sobre el amarillo pálido el Nº total de muestra de cada grupo y el nº de eventos de cada grupo, y los resultados salen automáticamente…

	Enferman Con eventos	No enferman Sin eventos	Total
Fármaco intervención	182	2.039	2.221
Fármaco control	256	1.967	2.223
Total	438	4.006	4.444

Nº eventos Interv (%)	Nº eventos Control (%)	RR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)	Potencia	Valor de *p* para la diferencia
182/2221 (8,19%)	256/2223 (11,52%)	0,71 (0,59-0,85)	3,32% (1,56% a 5,07%)	30 (20 a 64)	96,02%	0,0002

Aquí se obtiene el Riesgo absoluto del grupo de intervención, RA(i)

Y aquí el Riesgo absoluto del grupo control, RA(c)

Aquí se obtiene el RR
Aquí la RAR
Aquí el NNT

Aquí se obtiene la potencia estadística resultante

Aquí el valor de *p* para la diferencia

Figura 1: Hoja de cálculo para obtener RR, RAR, NNT, potencia y valor de p por riesgos acumulados. Disponible en http://evalmed.es/

94 eventos que incidieron en la muestra aleatoria (m_I) de 6022 pacientes representativos de la Población Mundial de pacientes con fibrilación auricular no valvular. La proporción p_I = 94 / 6022 = 0,0156 (=1,56%); entonces $1-p_I$ = 0,9844. En las tablas encontramos que para el 95% de confianza la Z = 1,96.

IC 95% = $p_I \pm 1{,}96 \times$ Raíz $[p_I \times (1-p_I) / n_I)]$ = 0,0156 ± 1,96 x Raíz [0,0156 x (0,9844) / 6022)] = 0,0031 (= 0,31%).

De esta forma se aplica a derecha e izquierda de la proporción, obteniendo el límite inferior así: 1,56% – 0,31% = 1,25%, y el límite superior así: 1,56% + 0,31% = 1,87%, y lo expresamos como 1,56% (IC 95%, 1,25% a 1,87%). De cada 100 veces que repitiéramos el experimento, tenemos la confianza de que en 95 de ellos el Promedio de la Población que perseguimos se encuentra entre 1,25% y 1,87%.

De manera similar, previo cálculo del error estándar de la Ratio de Proporciones (RR) y del error estándar de la Diferencia de Proporciones (RAR), se calculan los intervalos de confianza del RR y de la RAR respectivamente. Y los del NNT son los inversos de la RAR.

Para facilitar su cálculo, proporcionamos una sencilla calculadora en la web evalmed.es, pestaña HERRAMIENTAS, con la que es muy fácil introducir los datos. Por ejemplo, del anteriormente mencionado estudio "4S", en el que tras 5,4 años, se introducen los 182 pacientes con evento de "Muerte por cualquier causa" en el grupo de simvastatina y 256 en el grupo placebo, introduciendo también sus respectivos tamaños de muestra de 2.221 y 2.223 pacientes totales. Hecho esto, se obtienen automáticamente el RR, la RAR y el NNT con sus intervalos de confianza (figura 1).

Una diferencia no es estadísticamente significativa cuando concurren alguna de las siguientes circunstancias: a) se solapan los intervalos de confianza del %RAi y %RAc; b) el intervalo de confianza del RR atraviesa el uno; c) el intervalo de confianza de la RAR atraviesa el cero; o d) el intervalo de confianza del NNT "atraviesa el infinito". Se comprende que, como se ha expresado más arriba, cuando el intervalo de confianza del RR es uno, es porque el % de eventos del grupo de intervención (en el numerador) es igual al % de eventos del grupo de control (en el denominador), lo que se traduce en que no hay diferencia entre ambos. De la misma manera se comprende que cuando el intervalo de confianza de la RAR es cero, es porque el % de eventos del grupo de control (en el minuendo) es igual al % de eventos del grupo de intervención (en el sustraendo), lo que se traduce en que no hay diferencias entre ambos.

REPRESENTACIÓN GRÁFICA DEL RR

En la figura 2 simulamos un ensayo clínico en el que se comparan los infartos de los Fármacos A, B, C y D (intervención) frente a Placebo (control). Representamos los RR con sus intervalos de confianza al 95% de los 4 Fármacos de intervención frente al control (Placebo), para lo cual el control se sitúa en el "1".

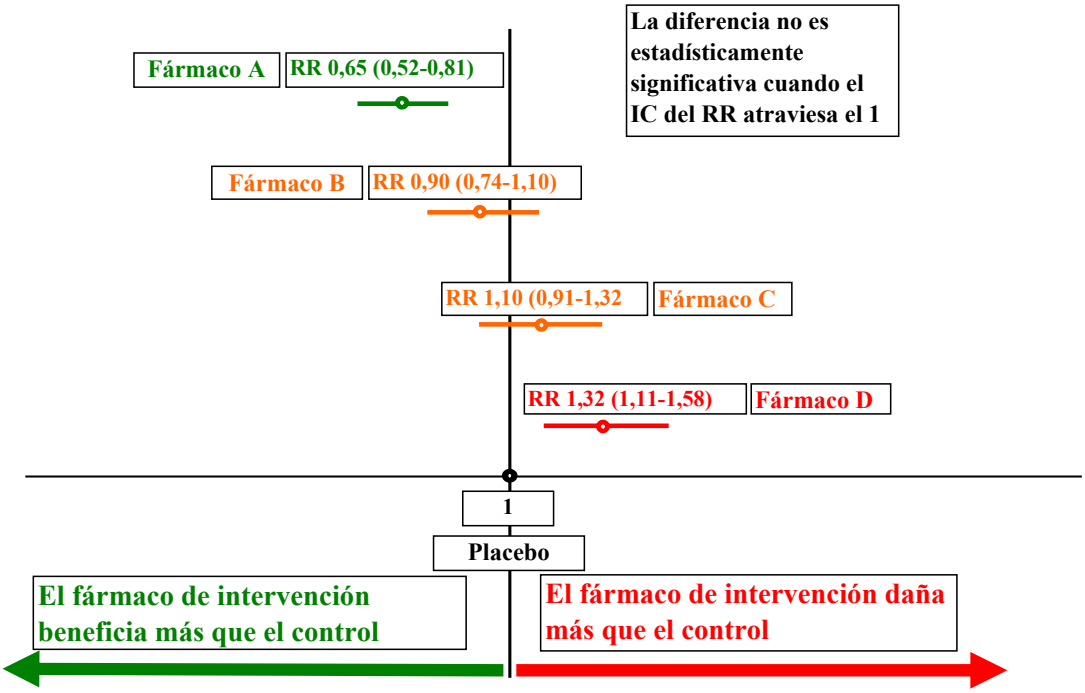
Se encuentra un Beneficio del Fármaco A frente a Placebo que es estadísticamente significativo. Al comparar el "% de infartos" con el Fármaco A frente a Placebo se obtuvo un RR 0,65 (0,52-0,81), cuyo intervalo de confianza no atraviesa el "1". Decimos por tanto que hay diferencia estadísticamente significativa a favor del Fármaco A, ya que tiene menos "% de infartos" en todo su intervalo de confianza que Placebo.

No se encuentra diferencia estadísticamente significativa entre los Fármacos B y C frente a Placebo. Al comparar el "% de infartos" con el Fármaco B frente a Placebo se obtuvo un RR 0,90 (0,74-1,10), que sí atraviesa el "1". Decimos entonces que no hay diferencia estadísticamente significativa entre ambos, ya que el Fármaco B tiene menos, igual y más "% de infartos" dentro de su intervalo de confianza que Placebo. Efectivamente, este intervalo de confianza es compatible al mismo tiempo con: a) menos "% de infartos" en la parte del intervalo que va desde 0,74 hasta el 1; b) igual "% de infartos" justo en el 1; y c) más "% de infartos" en la parte que va desde el 1 hasta el 1,10. Y eso mismo sucede con el Fármaco C frente a Placebo, con un RR 1,10 (0,91-1,31).

Se encuentra un Riesgo añadido del Fármaco D frente a Placebo que es estadísticamente significativo. Al comparar el "% de infartos" con el Fármaco D frente a Placebo se obtuvo un RR 1,32 (1,11-1,58), cuyo intervalo de confianza no atraviesa el "1". Por ello decimos que hay diferencia estadísticamente significativa en contra del Fármaco D, ya que tiene más "% de infartos" en todo su intervalo de confianza que Placebo.

¿Es clínicamente relevante además de estadísticamente significativo?

Figura 2: Representación gráfica del riesgo relativo (RR)



REPRESENTACIÓN GRÁFICA DE LA RAR

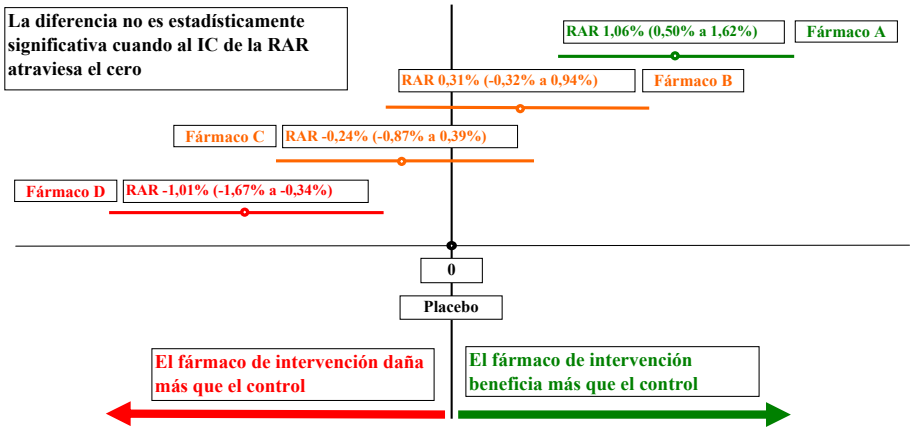
Simulamos el mismo ejemplo anterior y representamos las RAR con sus intervalos de confianza al 95% de los 4 Fármacos de intervención frente al control (Placebo), para lo cual el control se sitúa en el "0" (figura 3).

Se encuentra un beneficio estadísticamente significativo del Fármaco A frente a Placebo. Al comparar el "% de infartos" con el Fármaco A frente a Placebo se obtuvo una RAR 1,06% (0,5% a 1,62%), cuyo intervalo de confianza no atraviesa el "0". Decimos por tanto que hay diferencia estadísticamente significativa a favor del Fármaco A, ya que tiene menos "% de infartos" en todo su intervalo de confianza que Placebo.

No se encuentra diferencia estadísticamente significativa entre los Fármacos B y C frente a Placebo. Al comparar el "% de infartos" con el Fármaco B frente a Placebo se obtuvo una RAR 0,31% (-0,32% a 0,94%), que sí atraviesa el "0". Decimos entonces que no hay diferencia estadísticamente significativa entre ambos, ya que el Fármaco B tiene menos, igual y más "% de infartos" dentro de su intervalo de confianza que Placebo. Efectivamente, este intervalo de confianza es compatible al mismo tiempo con: a) más "% de infartos" en la parte del intervalo que va desde -0,32% hasta el 0%; b) igual "% de infartos" justo en el 0%; y c) menos "% de infartos" en la parte que va desde el 0% hasta el 0,94%. Y eso mismo sucede con el Fármaco C frente a Placebo, con una RAR -0,24% (-0,87% a 0,39%).

Se encuentra un Daño añadido estadísticamente significativo del Fármaco D frente a Placebo. Al comparar el "% de infartos" con el Fármaco D frente a Placebo se obtuvo una RAR -1,01% (-1,67% a -0,34%), cuyo intervalo de confianza no atraviesa el "1". Por ello decimos que hay diferencia estadísticamente significativa en contra del Fármaco D, ya que tiene más "% de infartos" en todo su intervalo de confianza que Placebo. Y como todo su intervalo de confianza es negativo nos resultará más fácil comprenderlo cambiando el signo negativo a positivo, pasando a denominarse Aumento Absoluto del Riesgo, AAR 1,01% (0,34% a 1,67%).

Figura 3: Representación gráfica de la reducción absoluta del riesgo (RAR)



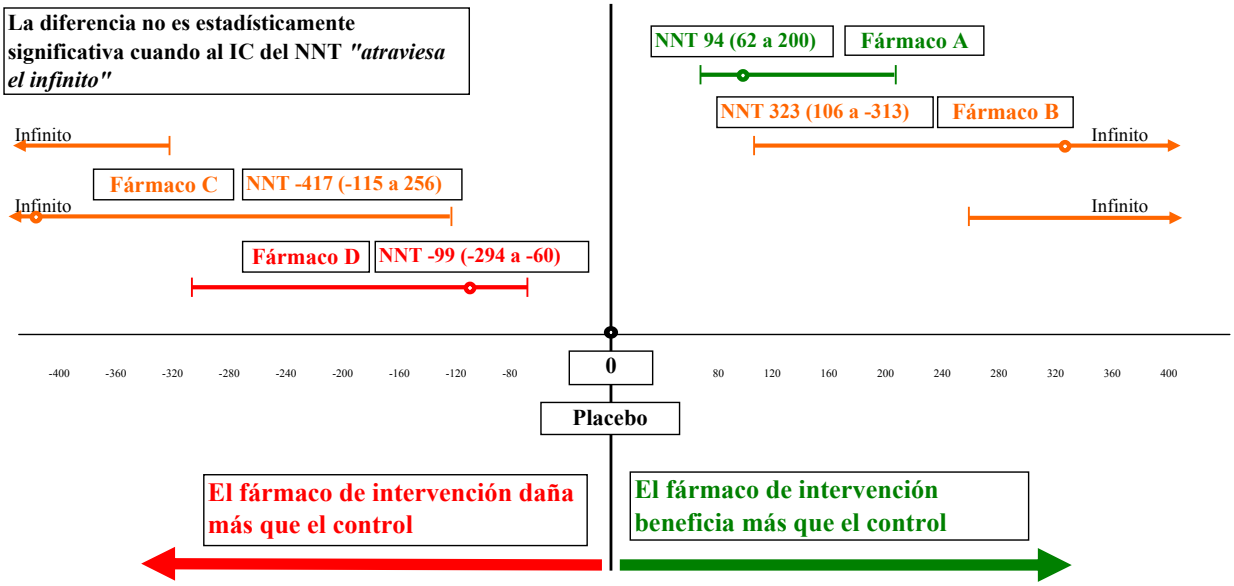
REPRESENTACIÓN GRÁFICA DEL NNT

Representamos los NNT con sus intervalos de confianza al 95% de los 4 Fármacos de intervención frente al control (Placebo), calculando los inversos de las RAR anteriores, para lo cual el control se sitúa en el "0" (figura 4).

Se encuentra un beneficio estadísticamente significativo del Fármaco A frente a Placebo. Al comparar el "% de infartos" con el Fármaco A frente a Placebo se obtuvo una NNT 94 (62 a 200), cuyo intervalo de confianza "no atraviesa el infinito". Decimos por tanto que hay diferencia estadísticamente significativa a favor del Fármaco A, ya que hay que tratar entre 62 y 200 pacientes para evitar un infarto.

No se encuentra diferencia estadísticamente significativa entre los Fármacos B y C frente a Placebo. Al comparar el "% de infartos" con el Fármaco B frente a Placebo se obtuvo un NNT

Figura 4: Representación gráfica del Número Necesario a Tratar para evitar"un" evento (NNT)



¿Es clínicamente relevante además de estadísticamente significativo?

323 (106 a -313), que "sí atraviesa el infinito". Decimos entonces que no hay diferencia estadísticamente significativa entre ambos, porque es compatible con evitar 1 infarto más que con placebo si tratamos entre 106 e infinitos pacientes, y con producir 1 infarto más que con Placebo por cada 313 e infinitos pacientes tratados. Lo mismo sucede con el Fármaco C frente a Placebo, con un NNT -417 (256 a -115).

Se encuentra un Daño añadido estadísticamente significativo del Fármaco D frente a Placebo. Al comparar el "% de infartos" con el Fármaco D frente a Placebo se obtuvo un NNT -99 (-60 a -294), cuyo intervalo de confianza "no atraviesa el infinito". Por ello decimos que hay diferencia estadísticamente significativa en contra del Fármaco D, porque produce un infarto más que Placebo por cada 60 a 294 pacientes tratados. Y como todo su intervalo de confianza es negativo nos resultará más fácil comprenderlo cambiando el signo negativo a positivo, pasando a denominarse Número Necesario para Dañar a un paciente más que con Placebo, NND 99 (60 a 294).

UTILIDADES Y LIMITACIONES DE LAS MEDIDAS DEL EFECTO PARA PERCIBIR LA RELEVANCIA CLÍNICA.

El RR expresa la relación entre el porcentaje de pacientes con eventos en ambos grupos, pero no nos informa si ambos porcentajes son insignificantes o altos. Un valor de RR = 0,5 no informa al médico si la relevancia clínica es alta o muy baja, pues puede ser el resultado de un 20% de eventos del grupo de intervención en el numerador frente a un 40% del grupo de control en el denominador, que es un resultado muy satisfactorio, pero también de un 0,2% frente a un 0,4%, que es un resultado insignificante. Un inconveniente aún mayor es que el RR no permite establecer un balance de beneficios y riesgos añadidos, porque todo balance es un ejercicio de sumas y restas para obtener un saldo, y esto puede hacerse para una misma variable con los porcentajes de eventos, pero no puede hacerse con los resultados de dividir dos porcentajes de eventos. La RRR tiene estos mismos inconvenientes, y conviene destacar que la "Efectividad" en las vacunas es la RRR. Sin embargo en algunas ocasiones hay exhortaciones desde agencias reguladoras, departamentos de la Administración y gestión sanitaria, fichas técnicas y Guías de Práctica Clínica que, aportando únicamente los riesgos relativos, terminan sus comunicaciones de riesgos y advertencias con la frase: "el médico valorará la instauración de tal terapia después de un balance de riesgos y beneficios", lo cual es imposible como mostramos en el ejemplo del siguiente párrafo. Imaginemos un medicamento cuya finalidad es evitar [ACV totales] y el informe dice que un medicamento frente a placebo es más eficaz en la variable [ACV isquémicos], con un RR = 0,50, y menos eficaz en la variable [ACV hemorrágicos], con un RR = 1,50. Pues bien, con estos RR el médico no puede estimar un balance de riesgos y beneficios, pues el RR = 0,50 no se puede restar del RR = 1,50.

La RAR sí proporciona al médico la magnitud del efecto y le permite establecer un balance de riesgos y beneficios. Efectivamente, completando al ejemplo anterior, si el informe dice que el medicamento frente a placebo es más eficaz para la variable [ACV isquémicos], con una RAR = 1,80%, y menos eficaz en la variable [ACV hemorrágicos], con una RAR = –0,24% (que es lo mismo que decir AAR = 0,24%), entonces, el médico sí puede establecer un balance de riesgos y beneficios, pues para [ACV totales (isquémicos y hemorrágicos)] se obtiene una RAR = 1,80% - 0,24% = 1,56%.

Verifíquese esta afirmación comprobando que la simple suma de las RAR situados en la vertical coincide con la RAR obtenida operando con los ACV totales situados en la horizontal (tabla 2).

El inconveniente de la RAR es que se trata de un porcentaje con dos decimales y esto dificulta al médico su recuerdo a la hora de incorporarlo a su práctica clínica. Y es que los dos decimales son importantes cuando el número entero de la RAR es igual o menor de 3, que es una franja muy abundante en la farmacoterapia de atención primaria y abundante en atención hospitalaria. Pero este inconveniente se palia con el NNT.

El NNT fue un invento de Laupacis y col en 1988, sencillamente como el inverso de la RAR, con el objetivo de estimar el número de pacientes que un médico tiene que tratar para conseguir evitar 1 evento, y tal y como ellos argumentan el NNT resume las medidas que son relevantes porque pone de manifiesto la relación del esfuerzo terapéutico (refiriéndose al número de personas que hay que tratar) con el rendimiento clínico (salvar del evento a una). Es, por tanto, la medida que mejor informa la magnitud del efecto de la intervención, es fácil de recordar para el médico y le permite establecer una sencilla relación entre el esfuerzo para conseguir beneficios de distintas opciones terapéuticas (3).

Cuando se utiliza el NNT como método para la toma de decisiones durante las visitas a la cabecera del enfermo, el procedimiento es muy simple, pues consiste en comparar el riesgo basal de mi paciente con el riesgo basal promedio del prototipo de paciente publicado en el ensayo clínico. Si el riesgo basal de mi paciente es un factor de f veces el riesgo inicial del prototipo del estudio, y el riesgo relativo se mantiene constante, la reducción

absoluta del riesgo para el paciente se escala de acuerdo con el mismo factor f. Por ejemplo si el NNT es 25 en un ensayo clínico con un riesgo basal de 6%, y mi paciente tiene un riesgo basal del 3%, como el factor f = 6% / 3% = 2, entonces el NNT para mi paciente es "NNT inicial" x "f" = 25 x 2 = 50 (4).

A pesar de su bondad, la debilidad del NNT es que no nos dice cuánto es el riesgo basal. Acudiendo de nuevo a la "Mortalidad por cualquier causa" del estudio 4 S, su NNT en 5,4 años es 30, el médico tiene que tratar a 30 pacientes para evitar la muerte a uno, pero como no contiene el riesgo basal, nada sabemos del destino de los otros 29 pacientes restantes (3, 5). Efectivamente, entre esos 29 restantes habrá algunos que morirán incluso tomando simvastatina y los restantes sobrevivirán sin tomar nada, o mejor dicho tomando un placebo. Y es por ello que Hutton y col nos exhortan a comunicar los resultados informando los riesgos en términos absolutos y relativos, haciendo referencia al riesgo basal. Por ejemplo «Se observan 33 eventos por cada 1.000 pacientes sin tratamiento. El tratamiento reduce la tasa en 20/1.000, es decir de 33/1.000 a 13/1.000. A pesar de que la reducción de eventos de 250/1.000 a 230/1.000 es la misma reducción absoluta (20/1.000), el nivel de ansiedad del paciente y su familia es diferente si saben que el riesgo basal es de 33/1.000 que si es de 250 /1.000» (6).

Nosotros, conscientes de la utilidad de esta información, la hemos incorporado completa en lo que venimos llamando la REGLA DEL 1, que nos muestra: por cada 1 paciente en que el medicamento es efectivo, en cuántos no es efectivo, discriminando dentro de éstos: a) el número de los que padecerán el evento incluso tomando el fármaco; y b) el número de los que permanecerán sanos (sin evento) incluso sin tomar nada (o mejor dicho tomando un placebo, en el caso de que éste sea el control). Tal interpretación es la correcta cuando el control es un placebo.

Tabla 2: Ejemplo simulado para mostrar que con los RR no puede establecerse un balance de beneficios y riesgos, y sí se puede con las RAR.

Seguimiento 1 año	Nº event Interv (%)	Nº event Control (%)	RR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)
ACV isquémicos	180 / 10000 (1,8%)	360 / 10000 (3,6%)	0,50 (0,42-0,60)	1,80% (1,35% a 2,25%)	56 (45 a 74)
ACV hemorrágicos	72 / 10000 (0,72%)	48 / 10000 (0,48%)	1,50 (1,04-2,16)	-0,24% (-0,45% a -0,02%)	-418 (-220 a -5088)
ACV totales	252 / 10000 (2,52%)	408 / 10000 (4,08%)	0,62 (0,53-0,72)	1,56% (1,06% a 2,05%)	64 (49 a 94)

Cuando el control es otro medicamento activo, la REGLA DEL 1 nos muestra: por cada 1 paciente en el que el medicamento de intervención es más efectivo que el medicamento control, en cuántos no es más efectivo, discriminando dentro de éstos: a) el número de los que tendrán el evento con el medicamento de intervención (igual que con el medicamento de control); y b) el número de los que permanecerán sin evento (igual que con el medicamento de control).

Hasta ahora todo lo dicho es de aplicación cuando el NNT es positivo, es decir cuando el beneficio es estadísticamente significativo, pero el mismo esquema se puede utilizar cuando el NNT es negativo, es decir cuando el daño añadido es estadísticamente significativo.

Para facilitar el cálculo y una representación gráfica, proporcionamos una sencilla calculadora en la web evalmed.es, pestaña HERRAMIENTAS, archivos "Regla del 1 si el control es placebo" y "Regla del 1 si el control es un medicamento activo".

UN EJEMPLO DE LA REGLA DEL UNO CUANDO EL CONTROL ES PLACEBO

Veamos el ejemplo de la variable principal combinada [Muerte CV, IAM, ACV Hospitalización por angina o Revascularización] del estudio JUPITER en el que, tras 1,9 años, se produjeron 142 primeros eventos en el grupo de rosuvastatina (n= 8.901) frente a 251 en el grupo placebo (n= 8.901) en pacientes de 66 años, normolipémicos, normotensos con PCR 4,2 mg/dl. Introducimos los datos en la hoja de cálculo "Regla del 1, cuando el control es placebo", que puede extraerse de la web evalmed.es, pestaña HERRAMIENTAS, y obtenemos RA con rosuvastatina 1,6%, RA con placebo 2,8%; RR 0,57 (0,46-0,69); RAR 1,22% (0,79% a 1,65%); NNT 82 (60 a 127); potencia 99,98%, valor de p < 0,001 (figura 5).

¿Es clínicamente relevante además de estadísticamente significativo?

Cálculo por riesgos acumulados de RR, RAR, NNT con sus IC 95%, potencia estadística y valor de *p*

Abreviaturas: **RA**: Riesgo Absoluto; **RR**: Riesgo Relativo; **RAR**: Reducción Absoluta del Riesgo; **NNT**: Número Necesario a Tratar para evitar un evento; **NND**: Número Necesario para Dañar a un paciente más que con el control; **IC 95%**:intervalo de confianza al 95%.

En las casillas con fondo amarillo se introducen el Nº de eventos y el Nº de pacientes de cada grupo.

	Nº de personas con evento	Nº personas sin evento	Total
Mto de Intervención	142	8.759	8.901
Mto de control	251	8.650	8.901
Total	393	17.409	17.802

Automáticamente se obtiene RR, RAR y NNT con sus intervalos de confianza. Y además la potencia y el valor de *p* .

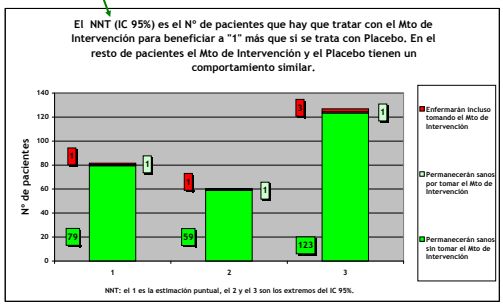
Nº event Interv (%)	Nº event Control (%)	RR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)	Potencia
142 (1,6%)	251 (2,82%)	0,57 (0,46-0,69)	1,22% (0,79% a 1,65%)	82 (60 a 127)	99,98%

Valor de *p* para la diferencia
0,0000

Aplíquese únicamente cuando el NNT y sus intervalos de confianza son POSITIVOS	Estimación puntual	Límite inferior del IC	Límite superior del IC
NNT =	82	60	127
Permanecerán sanos sin tomar el Mto de Intervención	79	59	123
Permanecerán sanos por tomar el Mto de Intervención	1	1	1
Enfermarán incluso tomando el Mto de Intervención	1	1	3

Aplíquese únicamente cuando el NNT y sus intervalos de confianza son NEGATIVOS	Estimación puntual	Límite inferior del IC	Límite superior del IC
NND =	82	127	60
Permanecerán sanos sin tomar el Mto de Intervención	80	124	60
Enfermarán por tomar el Mto de Intervención	1	1	1
Enfermarán incluso sin tomar el Mto de Intervención	2	4	2

Automáticamente se obtiene la "regla del 1" [Por cada 1 paciente en el que es efectivo, en cuántos no es efectivo].
La tabla superior explica el beneficio, pero únicamente puede aplicarse cuando el NNT y sus IC son positivos.
La tabla inferior explica el daño añadido, pero únicamente puede aplicarse cuando el NNT y sus IC son negativos, y entonces lo denomimamos NND.



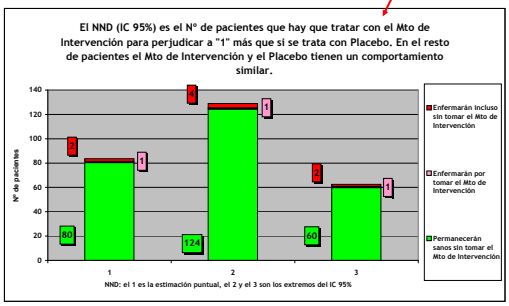


Figura 5: Hoja de cálculo para obtener los resultados y el gráfico de la Regla del 1. Disponible en http://evalmed.es

1) Obsérvese que en el grupo que tomó rosuvastatina hubo un 1,6% con evento de la variable en 1,9 años, lo que significa que hay 1,6 pacientes de cada 100 en los que el fármaco no es efectivo.

2) Obsérvese que en el grupo que tomó placebo hubo un 2,8% con evento de la variable en 1,9 años, lo que significa que hay 100-2,8% = 97,2% pacientes que permanecen sanos (entiéndase sin evento).

3) La diferencia entre los eventos con placebo y con rosuvastatina es 2,8%-1,6%= 1,22%. Éste es el beneficio asociado con rosuvastatina.

4) El mérito (el efecto) de rosuvastatina es únicamente el 1,22%, pues no tiene ningún efecto sobre el 1,6% de personas que tienen el evento incluso tomando el fármaco, ni tampoco tiene ningún efecto sobre el 97,2% de los pacientes que permanecen sanos, pues ésos permanecen sanos tomando placebo.

Esto mismo lo podemos expresar más fácilmente relacionando el número de pacientes en los que el fármaco no es efectivo por cada 1 en el que es efectivo. Los cálculos se hacen muy fácilmente sobre el NNT 82. Permanecerán sanos 82 x 97,2% = 79,4 (que se redondea a 79); Enfermarán incluso con fármaco: 82 x 1,22% = 1,3 (que se redondea a 1).

Si hacemos esto mismo con los límites del intervalo de confianza del NNT 82 (60 a 127), averiguamos que "por cada 1 paciente en el que la rosuvastatina es efectiva", permanecerán sanos sin tomar rosuvastatina entre 59 y 123 pacientes y sufrirán un evento tomando rosuvastatina entre 1 y 3 pacientes.

De lo observado podemos extraer un corolario. Un gran número de personas atribuye como mérito de rosuvastatina frente a placebo el número de pacientes que permanece "sano" por tomar dicho fármaco. Centrándonos en la estimación puntual para facilitar la explicación, diremos que el único beneficio real atribuido a rosuvastatina frente a placebo es de 1 paciente por cada 81 (sin beneficiar a los 79 que permanecen sanos igual que con placebo, ni a 1 que enferma incluso con rosuvastatina). Sin embargo, la falta de precisión le llevan a "creer" que beneficia a 1+79 = 80 por cada 81. La diferencia entre la creencia (realidad ingenua) y la realidad real es enorme, pues la creencia lleva a atribuir erróneamente un mérito de 80 por cada 81, cuando en realidad es 1 por cada 81.

RELEVANCIA CLÍNICA DE LAS INTERVENCIONES PREVENTIVAS

La diferencia entre una intervención y el control tiene que ser estadísticamente significativa para demostrar que es efectiva. Pero además de cumplir esa primera condición tiene que cumplir una segunda: ser clínicamente relevante. Porta Serra y col nos mostraron en 1998 un buen ejemplo de una intervención que era estadísticamente significativa (porque la "*p*" era menor de 0,05) pero no era clínicamente relevante (7).

La "relevancia clínica" es la combinación de una sensación y una percepción. Por una parte la sensación es una cualidad afectiva: me gusta o me disgusta, me acerco o me alejo, colaboro o deserto; es interna y subjetiva, y lo es para mí. Por otra, la percepción es externa y objetiva, y lo es para todos.

Como la percepción objetiva y su correspondiente sensación subjetiva se producen al mismo tiempo, en nuestra experiencia nunca las encontramos separadas. De ahí que nuestra creencia nos lleve a considerarlas una sola cosa, a darles un solo nombre y a confundir sus diferentes atributos. Resulta muy difícil distinguirlas en el pensamiento, atenderlas por separado y no atribuir a una de ellas nada que pertenezca a la otra.

Cuando me muerdo la lengua experimento la sensación de dolor. Esta experiencia existe para mí solo, y si tratase de describir a qué se parece, podría hacerlo sólo de una manera vaga y metafórica. El dolor que experimento tiene asociado un tiempo (ahora mismo), un lugar (mi lengua), una intensidad (suave) y un tono afectivo (desagradable), pero en la mayor parte de los demás aspectos parece situado más allá del alcance de la descripción física. De hecho mi dolor no es parte del mundo objetivo, el mundo de la materia física. En síntesis, difícilmente puede considerarse un acontecimiento físico (8).

Profundicemos más con el siguiente ejemplo. Ana y Pedro son dos de los veinticinco alumnos de química. Ayer en el laboratorio el profesor dibujó en la pizarra la forma plana de la molécula de limoneno, cuya nomenclatura, 1-metil, 4-isopropenil, 1-ciclohexeno, supieron formular todos, lo cual les sirvió para calcular que un mol pesa 136 gramos. Poco después sacó un frasco que contenía limoneno y pesó 1 gramo en una balanza. A la pregunta de cuántas moléculas contenía ese gramo, todos supieron que contenía $4{,}4 \times 10^{21}$, porque lo calcularon a partir del número de Avogadro. Seguidamente lo dividió en dos partes iguales, depositó 0,5 gramos en la palma de la mano de Ana y 0,5 gramos en la de Pedro, les dijo que aspiraran su aroma durante 10 segundos y les pidió que hablaran un minuto sobre sus sensaciones. El número de moléculas aspiradas fue el mismo, pero Ana habló de una agradable sensación que asociaba al patio fresco, limpio y luminoso de su abuela, y Pablo habló de una desagradable sensación por el dolor que le causó un amigo muy curioso cuando le exprimió el jugo de un limón encima de una herida.

Se entiende que respecto al limoneno todos los alumnos experimentan una percepción objetiva, constituida por su representación plana, su nomenclatura y su número de moléculas, que está pasando fuera de cada uno de ellos y que es igual para todos. Simultáneamente Ana experimenta una sensación subjetiva, difícil de comunicar, que está pasando dentro (de sí misma). Pedro también la experimenta, pero su sensación es distinta a la de Ana, sin que podamos saber cómo de distinta es.

Respecto al beneficio de una intervención, cada médico experimenta una sensación respecto a la gravedad de la enfermedad o evento, y simultáneamente experimenta una percepción respecto a su eficacia, es decir al número de personas a las que beneficia frente a las que no. Aunque no conocemos el mecanismo de cómo emerge la conciencia en los humanos, según la investigación de Nicholas Humphey, la mente estimaría la relevancia clínica combinando la sensación y la percepción. Y siendo plausible que así se forme, nosotros podemos descomponer la relevancia clínica en dos componentes: 1) la sensación del grado del riesgo; y 2) la percepción de la magnitud del efecto.

El primer componente. La sensación del riesgo actualmente es imposible de estandarizar porque, frente al mismo estímulo, cada médico puede experimentar una sensación subjetiva distinta a otros médicos, e incluso distinta a sí mismo en diferentes momentos y situaciones de su vida. Sin embargo, sí puede estandarizarse una escala ordinal de grados de aversión al riesgo, cuyo orden de aversión o preferencia desde el punto de vista biológico[1] es universal porque mantiene la impronta evolutiva. La mayor aversión de la biología es la muerte, por lo cual ésta ocupa universalmente el escalón superior en el orden de aversión.

Un observador puede no saber exactamente cómo de distintas son las sensaciones subjetivas de diez médicos ante un riesgo como la angina de pecho, pero sí sabrá que no será distinto el orden de su aversión a los siguientes riesgos: muerte, infarto de miocardio no fatal, angina de pecho, claudicación intermitente y cefalea leve.

La metodología GRADE utiliza el grado de aversión al riesgo para construir su escala ordinal de importancias para los pacientes, así: Son riesgos graves los que causan la muerte, incapacidad o amenaza de la vida, y les asigna las puntuaciones ordinales de importancia 9, 8 ó 7. Son riesgos moderados los que causan deterioro importante de la calidad de vida o son susceptibles de convertirse en graves, y les asigna las puntuaciones ordinales 6, 5

1 Lo referimos inicialmente a lo biológico, porque las circunstancias de la cultura circundante que desnaturalizan la biología introducen elementos que necesitan una incursión más profunda en la ética.

¿Es clínicamente relevante además de estadísticamente significativo?

ó 4. Y son riesgos leves los que causan deterioro no importante en la calidad de vida y no son susceptibles de convertirse en graves, y les asignan las puntuaciones ordinales 3, 2 ó 1. El inconveniente es que al ser números ordinales no nos permiten colegir que el ordinal "8" significa el doble de aversión que el ordinal "4".

El segundo componente. La percepción de la magnitud del efecto es un número cardinal, porque nos informa del número de personas a las que beneficia frente a las que no, que es justamente el NNT.

Una vez entendido que intuimos la relevancia clínica mediante una inadvertida combinación de ambos componentes, nos encontramos que en la práctica los médicos otorgan valoraciones muy diversas a un mismo NNT y/o a un mismo riesgo. Y como esta variabilidad puede derivar en arbitrariedad, conviene minimizarla consensuando una escala que oriente al médico en el esfuerzo terapéutico para conseguir un resultado en salud. Tal consenso no existe directamente expresado en la literatura biomédica (9).

Nosotros, formando un panel, hemos consensuado una escala orientativa sobre la magnitud del efecto de las intervenciones preventivas[2] para tres escalones cualitativos de riesgos graves, concretamente para las puntuaciones ordinales de importancia 9, 8 y 7 según la nomenclatura GRADE. Se trata de un consenso interno únicamente para poner adjetivos calificativos a la magnitud del efecto, que no pretende ser normativo fuera de nuestro panel. El elemento de consenso primero fue la calificación de magnitud del efecto alta para un NNT ≤ 150 en 1 año para una intervención preventiva de la mortalidad, es decir de importancia 9. A la calificación de magnitud del efecto alta para un NNT ≤ 133 en 1 año para una intervención preventiva de importancia 8 (como por ejemplo infarto no mortal), acordamos llegar mediante una regla de 3 simple (9 es a 8 como 150 es a 133), a pesar de que, tal como decimos, los números ordinales no son para operaciones proporcionales. Y de la misma manera llegamos al NNT ≤ 117 en 1 año para una intervención preventiva de importancia 7, como por ejemplo la angina de pecho (tabla 3).

2 Según el objetivo terapéutico distinguimos entre prevención y curación. Prevención es la reducción de un riesgo basal ya existente (en términos de probabilidad). Curación es la restitución de la salud perdida, la sustitución de la pérdida de sustancias indispensables para la vida o el alivio de los síntomas.

LA RAR Y EL NNT SE PUEDEN INTERPOLAR PERO NO EXTRAPOLAR

Cometiendo un error asumible, la RAR se puede interpolar desde el tiempo de seguimiento de un ensayo clínico hacia atrás. Así por ejemplo, si un ensayo clínico tuvo un tiempo de seguimiento de 2 años, y para el ACV la RAR fue 2%, entonces podemos interpolarlo a 1 año dividiéndolo por 2, obteniendo una RAR = 1% en 1 año, asumiendo que la incidencia de eventos fue constante en el tiempo. Ahora bien, si el ensayo clínico duró 2 años, no podemos extrapolar los resultados a 3 años.

De la misma manera, cometiendo error un asumible, el NNT se puede interpolar desde el tiempo de seguimiento de un ensayo clínico hacia atrás. Así por ejemplo, si un ensayo clínico tuvo un tiempo de seguimiento de 2 años, y para el ACV el NNT fue 25, entonces podemos interpolarlo a 1 año multiplicándolo por 2, obteniendo un NNT = 50 en 1 año, asumiendo que la incidencia de eventos fue constante en el tiempo. Ahora bien, si el ensayo clínico duró 2 años, no podemos extrapolar los resultados a 3 años (10).

EL EFECTO MARCO

El efecto marco es un sesgo cognitivo descubierto y estudiado por el programa de investigación llevado a cabo hace varias décadas por Amos Tversky y Daniel Kahneman (premio Nobel de economía en 2003), que demostraron experimentalmente que el cerebro humano confunde en muchas situaciones lo psicológico con lo lógico, afectando con ello a la toma de decisiones. Se define al efecto marco como la discrepancia en las respuestas proporcionadas por los sujetos según que una misma información se presente desde un marco positivo o desde un marco negativo (11, 12). Autores posteriores clasifican el efecto marco en tres diferentes tipos: el marco del riesgo, el marco del atributo y el marco de objetivo. Pues bien, las decisiones sanitarias incurren en los tres.

Un ejemplo del marco de riesgo (risk framing) puede ser: Elegir el programa A para 600 personas con el que morirán 200 frente al programa B con el que vivirán 400 (cuyo resultado reve-

ESCALA DE CONSENSO: CALIFICATIVOS DE LA MAGNITUD DEL EFECTO EN EVENTOS GRAVES DE IMPORTANCIA 9 (*) PARA INTERVENCIONES PREVENTIVAS						
Magnitud del efecto de una intervención con resultados en salud	NNT					
	en 1 año	en 2 años	en 3 años	en 4 años	en 5 años	en 10 años
alta	≤ 150	≤ 75	≤ 50	≤ 38	≤ 30	≤ 15
moderada	151 a 300	76 a 150	51 a 100	39 a 75	31 a 60	16 a 30
baja	301 a 450	151 a 225	101 a 150	76 a 113	61 a 90	31 a 45
muy baja	> 450	> 225	> 150	> 113	> 90	> 45
Magnitud del efecto de una intervención con resultados en salud	RAR					
	en 1 año	en 2 años	en 3 años	en 4 años	en 5 años	en 10 años
alta	≥ 0,67%	≥ 1,33%	≥ 2,00%	≥ 2,67%	≥ 3,33%	≥ 6,67%
moderada	0,33% a 0,67%	0,67% a 1,33%	1,00% a 2,00%	1,33% a 2,67%	1,67% a 3,33%	3,33% a 6,67%
baja	0,22% a 0,32%	0,44% a 1,66%	0,67% a 0,99%	0,89% a 1,32%	1,11% a 1,66%	2,22% a 3,32%
muy baja	< 0,22%	< 0,44%	< 0,67%	< 0,89%	< 1,11%	< 2,22%

(*) Los calificativos son los consensuados por el panel firmante de este artículo.

ESCALA DE CONSENSO: CALIFICATIVOS DE LA MAGNITUD DEL EFECTO EN EVENTOS GRAVES DE IMPORTANCIA 8 (*) PARA INTERVENCIONES PREVENTIVAS						
Magnitud del efecto de una intervención con resultados en salud	NNT					
	en 1 año	en 2 años	en 3 años	en 4 años	en 5 años	en 10 años
alta	≤ 133	≤ 67	≤ 44	≤ 33	≤ 27	≤ 13
moderada	134 a 267	68 a 133	45 a 89	34 a 67	28 a 53	14 a 27
baja	268 a 400	134 a 200	90 a 133	68 a 100	54 a 80	28 a 40
muy baja	> 400	> 200	> 133	> 100	> 80	> 40
Magnitud del efecto de una intervención con resultados en salud	RAR					
	en 1 año	en 2 años	en 3 años	en 4 años	en 5 años	en 10 años
alta	≥ 0,75%	≥ 1,50%	≥ 2,25%	≥ 3,00%	≥ 3,75%	≥ 7,50%
moderada	0,38% a 0,75%	0,75% a 1,50%	1,13% a 2,25%	1,50% a 3,00%	1,88% a 3,75%	3,75% a 7,50%
baja	0,25% a ,37%	0,40% a 0,74%	0,59% a 1,12%	1,00% a 1,49%	1,25% a 1,87%	2,50% a 3,74%
muy baja	< 0,25%	< 0,50%	< 0,75%	< 1,00%	< 1,15%	< 2,50%

(*) Los calificativos son los consensuados por el panel firmante de este artículo.

ESCALA DE CONSENSO: CALIFICATIVOS DE LA MAGNITUD DEL EFECTO EN EVENTOS GRAVES DE IMPORTANCIA 7 (*) PARA INTERVENCIONES PREVENTIVAS						
Magnitud del efecto de una intervención con resultados en salud	NNT					
	en 1 año	en 2 años	en 3 años	en 4 años	en 5 años	en 10 años
alta	≤ 117	≤ 58	≤ 39	≤ 29	≤ 23	≤ 12
moderada	117 a 233	58 a 117	39 a 78	29 a 58	23 a 47	12 a 23
baja	234 a 350	118 a 175	79 a 117	59 a 88	48 a 70	24 a 35
muy baja	> 350	> 175	> 117	> 88	> 70	> 35
Magnitud del efecto de una intervención con resultados en salud	RAR					
	en 1 año	en 2 años	en 3 años	en 4 años	en 5 años	en 10 años
alta	≥ 0,86%	≥ 1,71%	≥ 2,57%	≥ 3,43%	≥ 4,29%	≥ 8,57%
moderada	0,43% a 0,86%	0,86% a 1,71%	1,29% a 2,57%	1,71% a 3,43%	2,14% a 4,29%	4,29% a 8,57%
baja	0,29% a 0,42%	0,57% a 0,85%	0,86% a 0,1,28%	1,14% a 1,70%	1,43% a 2,13%	2,86% a 4,28%
muy baja	< 0,29%	< 0,57%	< 0,86%	< 1,14%	< 1,43%	< 2,86%

(*) Los calificativos son los consensuados por el panel firmante de este artículo.

¿Es clínicamente relevante además de estadísticamente significativo?

ló que es más frecuente la aversión al riesgo en las decisiones que implican ganancias y más frecuente la tendencia a correr riesgos en las decisiones que implican pérdidas). Un caso de marco de atributo (attribute framing) puede ser: Elegir un yogur libre de grasa en un 99% frente a un yogur con grasa en un 1%. Un ejemplo en el marco de objetivo (goal framing) puede ser: Elegir entre "si se hace una mamografía usted obtendrá las ventajas A, B y C" frente a "si usted no se hace una mamografía dejara de tener las ventajas A, B y C".

Siendo una variante de "efecto marco", según algunos autores (13, 14), o siendo sus efectos similares a él, hay numerosos estudios que muestran que un mismo beneficio clínico presentado en dos marcos distintos afecta a las elecciones y decisiones de individuos sanos, pacientes, médicos clínicos, becarios de epidemiología, profesores de epidemiología, autoridades y gestores sanitarios. Mayoritariamente cuando la información se presenta en forma de RRR la magnitud del efecto terapéutico se sobrestima y se incrementa significativamente la propensión de los médicos a prescribir, de los pacientes a desear que se les prescriba y de las autoridades y gestores sanitarios a financiar (15), que cuando se presenta en forma de RAR o de NNT. Esto puede conducir al sobretratamiento en pacientes con bajo riesgo basal, exponiéndolos a efectos adversos con baja o nula expectativa de beneficios (16, 17). Asimismo, cuando se comparan la RAR y su inverso el NNT, con este último decrece la propensión de los médicos a prescribir y de los pacientes a desear que se les prescriba (14, 18-30). Otro tanto sucede cuando un mismo dato de terapia oncológica se presenta como probabilidad de vivir frente a probabilidad de morir, resultando estadísticamente más atractiva la elección en el primer marco (31).

En la tabla 4 mostramos 5 ejemplos simulados que muestran el mismo RR = 0,5, con distinta relevancia clínica en términos de NNT. El RR = 0,5 es lo mismo que la RRR = 1 - RR = 1 – 0,5 = 0,5 (=50%), lo cual significa que "la intervención reduce el riesgo relativo en un 50%". En los cinco casos este significado común produce una misma impresión emocional en el lector y, sin embargo, el esfuerzo terapéutico de los cinco supuestos va desde tener que tratar 77 hasta tener que tratar a 6 pacientes para conseguir el mismo redimiento: evitar a 1 paciente el evento grave.

Tabla 4: Cinco ejemplos similados que muestran el mismo RR = 0,5, con distinta relevancia clínica en términos de NNT.

Prevención de evento grave en 1 año	Nº event Interv (%)	Nº event Control (%)	RR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)
Supuesto 1	13 (1,3%)	26 (2,6%)	0,5	1,3%	77
Supuesto 2	53 (5,3%)	106 (10,6%)	0,5	5,3%	19
Supuesto 3	93 (9,3%)	186 (18,6%)	0,5	9,3%	11
Supuesto 4	133(13,3%)	266 (26,6%)	0,5	13,3%	8
Supuesto 5	173 (17,3%)	346/34,6%)	0,5	17,3%	6

EL VALOR "*p*"

Muchos autores se han decantado claramente a favor de los intervalos de confianza y adversamente hacia el valor p de los contrastes de hipótesis. Nosotros coincidimos con esta recomendación porque el contraste de hipótesis es engorroso y requiere estimar la probabilidad de negar una negación. No obstante es conveniente calcularla habitualmente y es necesario cuando no se pueden calcular los intervalos de confianza.

El contraste de hipótesis supone una comparación. Al valor numérico que se compara se le denomina "efecto". Se suele relacionar este efecto en el numerador con la variabilidad aleatoria esperada ("error esperado por el muestreo o error estándar") en el denominador. Para no perdernos, el "efecto" es la diferencia entre lo observado en la muestra y lo que sería esperable como consecuencia del azar si H_0 fuese cierta. El efecto es la señal y el error estándar es el ruido ambiente. Si el efecto observado en la muestra (numerador) es muy superior al error esperado (denominador), se concluye diciendo que la muestra apoya la existencia del efecto en la población. Si el efecto observado en la muestra es similar o inferior al error esperado, se concluye diciendo que la muestra no apoya la existencia de dicho efecto en la población.

Para tomar una u otra decisión, establecemos primero dos hipótesis (a nivel de la población):

a) Hipótesis nula (H_0): mantiene que el efecto de interés no existe (es nulo, es decir, vale 0) en la población de la que procede la muestra.

b) Hipótesis alternativa (H_1): mantiene que existe algún efecto distinto de 0 en la población de la que procede la muestra.

En sentido estricto, la interpretación de un "valor de p" es la probabilidad de que las diferencias esperadas por el muestreo sean iguales o mayores a las observadas en la muestra si la hipótesis nula (H_0) fuera cierta. Valor p = p (dif esperada ≥ dif observada | H_0), y se expresa así porque es una probabilidad condicionada a que Ho sea cierta. Aunque nos resulte incómodo estamos obligados a formular la hipótesis nula (H_0) de que "no hay diferencias entre los promedios de intervención y control", y la hipótesis alternativa (H_1) de que "sí hay diferencias entre los promedios de intervención y control".

Cuando p < 0,05 (ó < 5%) se suele rechazar la hipótesis nula porque sería muy raro encontrar tales datos si H_0 fuera cierta, y se dice que la diferencia es estadísticamente significativa (se acepta la hipótesis alternativa). Cuando p > 0,1 no se suele rechazar la hipótesis nula porque no es raro encontrar tales datos si H_0 es cierta, y se dice que la diferencia no es estadísticamente significativa. Algunos autores consideran "tierra de nadie" cuando p está entre 0,05 y 0,1, lo cual exige una interpretación aparte.

Veamos un ejemplo simulado. Necesitamos saber si una vacuna es efectiva para prevenir una infección, para lo que asignamos aleatoriamente 200 individuos sanos al grupo de vacuna y 200 al grupo de placebo. Veamos tres supuestos:

Supuesto 1. Al cabo de un año los individuos que han tenido infección han sido 40 (20%) en el grupo de vacuna y 40 (20%) en el grupo de placebo. No hay necesidad de calcular para saber que el valor de p = 1 (ó 100%), es decir que la probabilidad de que no haya diferencia es del 100%.

Supuesto 2. Al cabo de un año han tenido infección 36 (18%) individuos en el grupo de vacuna y 40 (20%) en el grupo de placebo. Se encuentra una diferencia 20% - 18% = 2%, pero esta diferencia ¿es mayor, igual o menor a la esperada por azar? Para responderlo, calculamos y obtenemos una p = 0,6 (ó 60%); lo que significa que hay una probabilidad del 60% de que la diferencia esperada por azar sea igual o mayor del 2%, por lo que tenemos pocos argumentos para rechazar la hipótesis nula, y decimos que la diferencia encontrada no es estadísticamente significativa.

Supuesto 3. Al cabo de un año han tenido infección 20 (10%) individuos en el grupo de vacuna y 40 (20%) en el grupo de placebo. Se encuentra una diferencia 20% - 10% = 10%, y tras calcular obtenemos una p = 0,005 (ó 0,5%); lo que significa que hay una probabilidad del 0,5% de que la diferencia esperada por azar sea igual o mayor del 20%, por lo que tenemos muchos argumentos para rechazar la hipótesis nula, y decimos que la diferencia encontrada si es estadísticamente significativa.

Es importante señalar que con el valor de p no se puede estimar la relevancia clínica. Un mismo valor de p puede corresponder a un efecto alto o muy bajo. En la siguiente tabla observamos un supuesto simulado de dos vacunas para prevenir una infección grave en 1 año. Los resultados de la primera son 20/200 (10%) frente a 40/200 (20%) y los de la segunda 500/10.000 (5%) frente a 590/10.000 (5,9%). Ambas ofrecen un valor de p = 0,005 (0,5%). Sin embargo la primera tiene un NNT 10 (6 a 33), cuyo efecto estimamos de magnitud alta, y la segunda tiene un NNT 111 (65 a 371), cuyo efecto estimamos de magnitud moderada a baja. La relevancia clínica es muy distinta (tabla 5).

Tabla 5: Ejemplo simulado de dos vacunas que muestran el mismo valor de *p* = 0,005 (0,5%), con muy distinta relevancia clínica en términos de NNT.

Prevención de infección grave en 1 año	Nº event Interv (%)	Nº event Control (%)	RR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)	Potencia	Valor de *p*
Vacuna A	20/200 (10%)	40/200 (20%)	0,5 (0,3-0,82)	10% (3,01% a 17%)	10 (6 a 33)	80,0%	0,0051
Vacuna B	500/10000 (5%)	590/10000 (5,9%)	0,85 (0,75-0,95)	0,9% (0,27% a 1,53%)	111 (65 a 371)	80,1%	0,0051

¿Es clínicamente relevante además de estadísticamente significativo?

BIBLIOGRAFÍA

1. Pinker S. Cómo funciona la mente. Cap 5: Buenas ideas. Ferrán Meler-Orti (trad.). Barcelona. Ed. Destino. 2001. p. 436-7.
2. Pedersen TR, Kjekshus J, Berg K on behalf of Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet. 1994; 344(8934):1383-9.
3. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med. 1988;318:1.728-33.
4. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ. 1995;310(6977):452-4.
5. Tramèr MR, Walder B. Number needed to treat (or harm). World J Surg. 2005;29(5):576-81.
6. Hutton JL. Number needed to treat and number needed to harm are not the best way to report and assess the results of randomised clinical trials. Br J Haematol. 2009;146(1):27-30.
7. Porta Serra M, Plasencia A, Sanz F. La calidad de la información clínica (y III): ¿estadísticamente significativo o clínicamente relevante? Med Clín. 1998;90:463-8.
8. Humphrey N. Una historia de la mente: La evolución y el nacimiento de la conciencia. Cap 4: La doble provincia de los sentidos. José Mª Lebrón (trad.). Barcelona. Ed Gedisa. 1995. p. 49-56.
9. Librero J, Peiró S. Comparaciones abstrusas en los resúmenes de ensayos clínicos en revistas médicas españolas. Med Clín. 2001;117:732-3.
10. Palomo L, Sánchez-Robles G. Extrapolating the NNT jeopardises biostatistics and leads to immortality. Aten Primaria. 2011;44(5):296-7.
11. Tversky A, Kahneman D. The framing of decisions and the rationality of choice. 1981;211(4481):453-8.
12. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science. 1974;185(4157):1124-31.
13. Malenka DJ, Baron JA, Johansen S, Wahrenberger JW, Ross JM. The framing effect of relative and absolute risk. J Gen Intern Med. 1993;8(10):543-8.
14. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? Ann Intern Med.1992;117(11):916-21.
15. Fahey T, Griffiths S, Peters TJ. Evidence based purchasing: understanding results of clinical trials and systematic reviews. BMJ. 1995;311(7012):1056-9; discussion 1059-60.
16. Smith GD, Egger M. Who benefits from medical interventions? BMJ. 1994;308(6921):72-4.
17. Brett AS. Treating hypercholesterolemia. How should practicing physicians interpret the published data for patients? N Engl J Med. 1989;321(10):676-80.
18. Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. Am J Med. 1992;92(2):121-4.
19. Figon G, Boissel JP, Peyrieux JC. Doctors' perception of pertinent information. Results of a survey of a random sample of French general practitioners. GEP (Groupe d'Etude de la Prescription). Eur J Clin Pharmacol. 1992;43(2):113-6.
20. Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. Lancet. 1994;343(8907):1209-11.
21. Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. BMJ. 1994;309(6957):761-4.
22. Meneu R, Peiró S, Márquez S. Influencia de la presentación de los resultados de los ensayos clínicos en la intención de prescribir: relativizando el riesgo relativo. Aten Primaria. 1998; 21: 446-450.
23. Nikolajevic-Sarunac J, Henry DA, O'Connell DL, Robertson J. Effects of information framing on the intentions of family physicians to prescribe long-term hormone replacement therapy. J Gen Intern Med. 1999;14(10):591-8.
24. McGettigan P, Sly K, O'Connell D, Hill S, Henry D. The effects of information framing on the practices of physicians. J Gen Intern Med. 1999;14(10):633-42.
25. Lacy CR, Barone JA, Suh DC, Malini PL, Bueno M, Moylan DM, Kostis JB. Impact of presentation of research results on likelihood of prescribing medications to patients with left ventricular dysfunction. Am J Cardiol. 2001;87(2):203-7.
26. Moxey A, Dip G, O'Connell D, McGettigan P. Describing treatment effects to patients: How they are expressed makes a difference. J Gen Intern Med. 2003;18: 948–95.
27. Carling CL, Kristoffersen DT, Montori VM, Herrin J, Schünemann HJ, Treweek S, Akl EA, Oxman AD. The effect of alternative summary statistics for communicating risk reduction on decisions about taking statins: a randomized trial. PLoS Med. 2009 Aug;6(8):e1000134.
28. Goodyear-Smith F, Kenealy T, Wells S, Arroll B, Horsburgh M. Patients' preferences for ways to communicate benefits of cardiovascular medication. Ann Fam Med. 2011;9(2):121-7.
29. García-Retamero R, Galesis M. Who profits from visual aids: Overcoming challenges in people's understanding of risks. Soc Sci Med. 2010;70(7):1019-25.
30. Perneger TV, Agoritsas T. Doctors and patients' susceptibility to framing bias: a randomized trial. J Gen Intern Med. 2011;26(12):1411-7.
31. McNeil BJ, Pauker SG, Sox HC Jr, Tversky A. On the elicitation of preferences for alternative therapies. N Engl J Med. 1982;306(21):1259-62.