

¿Es clínicamente relevante además de estadísticamente significativo?

Álvarez-Cienfuegos A, Montaña 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? Web evalmed.es. 31-10-2014. Disponible en: <http://evalmed.es/2020/04/20/modulo-1-1-estimacion-de-las-medidas-de-la-efectividad/>



1. LA MISIÓN (*telos*) DE LAS INTERVENCIONES SANITARIAS

LA MISIÓN (telos) DE LAS INTERVENCIONES SANITARIAS

La misión (finalidad propositiva) de toda intervención sanitaria es **disminuir en una magnitud relevante los riesgos [1,2] basales graves y moderados de una persona** que forma parte de una comunidad de personas [3], **sin que tal intervención añada riesgos que igualen o superen los de la situación inicial**. El resultado del balance entre los **Beneficios** (riesgos evitados) y los **Riesgos** añadidos además debe justificar los **Inconvenientes** y los **Costes (BRIC)**, en el marco de los valores y preferencias de la persona informada (hasta garantizar su comprensión), de modo que con su autodeterminación (autonomía), codeterminada dentro de la comunidad de personas de la que forma parte, pueda tomar la mejor decisión para sí [4].

Los clínicos, además de tener como misión la general de todas las intervenciones sanitarias, también deben acompañar al enfermo grave o moderado, especialmente cuando la mejor intervención tiene un beneficio cero o no compensa los riesgos añadidos.

[1] Con el término “riesgo” nos referimos a la probabilidad de incidencia de un evento en un tiempo determinado. No debe confundirse por tanto “riesgo” con “factores de riesgo”. Efectivamente, los factores de riesgo son asociaciones estadísticas y no las causas, por lo cual la intervención artificial sobre ellos no significa que disminuirá el riesgo con el que está asociado estadísticamente.

[2] Ante una misma percepción de un riesgo, como por ejemplo “hay una probabilidad de que 1 individuo de cada 100 con la condición AAA tenga un ictus en 1 año”, las personas con y sin la condición AAA tienen distinta sensación subjetiva, influida por la cultura, valores y preferencias, todo ello mediatizado por los heurísticos y sesgos cognitivos de la mente humana.

[3] Los humanos siempre nacen y viven en grupos de personas. El individuo no existe más que como una abstracción de las unidades que hay en los grupos, comunidades, sociedades políticas realmente existentes. Una persona no puede darse normas (*nomos*) a sí misma (*autós*). De lo que sí dispone es de auto-determinación, pero ésta siempre está limitada por las codeterminaciones dentro de los grupos a los que pertenece.

[4] Inevitablemente toda intervención sanitaria parte de las necesidades (no deseos) que emite el individuo sano o enfermo (*input*). Y es el profesional el que posteriormente absorbe esas necesidades (*output*).

2. RIESGO, EFECTO, MAGNITUD DEL EFECTO Y BALANCE DE BENEFICIOS Y DAÑOS

RIESGO, EFECTO, MAGNITUD DEL EFECTO Y BALANCE DE BENEFICIOS Y DAÑOS.

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 personas que sufren un evento en el grupo de intervención frente al número de personas 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 enfermos¹; 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%**.

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.

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 algo 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”.

3. LAS TRES MEDIDAS MÁS UNIVERSALES DE LA EFECTIVIDAD DE LAS INTERVENCIONES: RR, RAR y NNT

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 (2).

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 (tabla 1).

Riesgo Absoluto o Incidencia grupo de intervención: $RA_i = 182 / 2221 = 8,2\%$

Riesgo Absoluto o Incidencia grupo de control: $RA_c = 256 / 2223 = 11,5\%$

Riesgo Relativo: $RR = RA_i / RA_c = 8,2\% / 11,5\% = 0,71$

Reducción Relativa del Riesgo: $RRR = 1 - RR = 1 - 0,71 = 0,29 (= 29\%)$

Reducción Absoluta del Riesgo: $RAR = RA_c - RA_i = 11,5\% - 8,2\% = 3,3\%$

Número Necesario a Tratar para evitar 1 evento: $NNT = 100 / RAR = 100\% / 3,3\% = 30$

Tabla 1.a: Hombre-Mujer de 58 años, con historia de angina o IAM y con Col y TG elevados.

ECA 4S; mediana de seguimiento 5,4 años	Grupo de intervención Simvastatina; N° eventos (%), n _i = 2221	Grupo de control Placebo; N° eventos (%), n _c = 2223	Cálculo por incidencias acumuladas			
	RA _i = n° eventos/ n _i	RA _c = n° eventos/ n _{ic}	RR (IC 95%)	RRR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)
Muerte por cualquier causa	182/ 2221 = 8,2%	256/ 2223 = 11,5%	8,2%/ 11,5% = 0,71	1 - 0,71 = 0,29 = 29%	11,5% - 8,2% = 3,3%	100% / 3,3% = 30
Muertes coronarias						
IAM probado						
Mialgias						

Abreviaturas: ECA: estudio controlado y aleatorizado; NNT: número necesario a tratar para evitar 1 evento; RA_c: Incidencia o % de riesgo en el grupo de control; RA_i: incidencia o % de riesgo absoluto en el grupo de intervención; RAR: reducción absoluta del riesgo; RR: riesgo relativo; RRR: reducción relativa del riesgo.

El Riesgo Absoluto (RA) o incidencia es la probabilidad que tiene un sujeto de sufrir un evento en la unidad de tiempo, porque se dispone de una prueba de que ha sucedido.

Puede expresarse más habitualmente en porcentaje así: el RA del grupo de control en “Mortalidad por todas las causas” fue $256 / 2223 = 11\%$ en 5,4 años.

Esto significa que, de cada 100 pacientes similares a los del grupo de control, y cuyo tratamiento es igual al del grupo de control (placebo, que es lo más parecido al no tratamiento), al cabo de 5,4 años habrá 11 que morirán, y $100 - 11 = 89$ permanecerán vivos, sin que podamos saber de antemano quienes son unos y otros.

El **Riesgo Relativo (RR)** es la **ratio** entre el % de riesgo absoluto del grupo de intervención en el numerador y el % de riesgo absoluto del grupo de control en el denominador. Y así un RR 0,71 es una fracción que contiene 0,71 en el numerador (grupo de intervención) y 1 en el denominador (grupo de control), así: $RR = 0,71 / 1$, lo cual significa que la relación entre incidencias es 0,71 con la intervención cada vez que se produce 1 con el control.

Lógicamente el RR es 0,71 cuando el numerador y denominador mantienen esa relación, como por ejemplo si multiplicamos ambos por dos así: $(0,71 \times 2) / (1 \times 2) = 1,42 / 2 = 0,71 / 1 = 0,71$.

La eliminación de los pacientes en riesgo (representados en el %) en la operación de la división, incapacita al RR para estimar la relevancia clínica de las intervenciones, como veremos también más adelante. El RR es muy útil, pero no para estimar la relevancia clínica.

La **Reducción Relativa del Riesgo (RRR)**, es la diferencia que hay entre el 1 del denominador (grupo de control) y el 0,71 del numerador (grupo de intervención), así: $RRR = 1 - 0,71 = 0,29$ (= 29%). Es más fácil de recordar como $1 - RR$, cuyo significado es el complementario del RR hasta el 1.

La **Reducción Absoluta del Riesgo (RAR)** es la diferencia entre el riesgo absoluto del grupo de control (RA_c) menos el riesgo absoluto del grupo intervención (RA_i).

El **Número Necesario a Tratar (NNT)** de una intervención expresa el número de pacientes que hay que tratar para evitar un evento más que si son tratados con el control, y se puede calcular mediante una regla de tres simple así:

Si evito 3,3 eventos tratando a 100 pacientes, entonces para evitar “1” evento necesito tratar a $100 \times 1 / 3,3 = 30$, aunque puede hacerse más rápidamente calculando el inverso del RAR así: $100\% / 3,3\% = 100 \text{ pacientes} / 3,3 \text{ eventos} = 30 \text{ pacientes} / 1 \text{ evento} = 30 \text{ pacientes/evento}$.

Un NNT 30 significa que si tenemos 30 sillas delante de nuestra consulta con esos pacientes, y tratamos a los 30 con simvastatina durante 5,4 años, evitaremos 1 muerte más que si los tratamos con placebo. Beneficiamos a 1 de cada 30 pacientes, sin beneficiar a los 29 restantes.

INTERPRETACIÓN PARA SABER SI BENEFICIA, DAÑA, O ES NEUTRAL

Cuando el medicamento de intervención beneficia más que el medicamento de control, el $RR < 1$, y por tanto la RAR y el NNT serán positivos.

Cuando el medicamento de intervención daña más que el medicamento de control, el $RR > 1$, y por tanto la RAR y el NNT serán negativos. Y si bien se comprende fácilmente un $RR > 1$, 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 ($RAR = 0$).

Veamos los beneficios y daños añadidos sobre más variables de resultados en salud en la siguiente tabla...

Tabla 1.a: Hombre-Mujer de 58 años, con historia de angina o IAM y con Col y TG elevados.

ECA 4S; mediana de seguimiento 5,4 años	Grupo de intervención Simvastatina; N° eventos (%), n _i = 2221	Grupo de control Placebo; N° eventos (%), n _c = 2223	Cálculo por incidencias acumuladas			
	RA _i = n° eventos/ n _i	RA _c = n° eventos/ n _{ic}	RR (IC 95%)	RRR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)
Muerte por cualquier causa	182/ 2221 = 8,2%	256/ 2223 = 11,5%	8,2%/ 11,5% = 0,71	1 - 0,71 = 0,29 = 29%	11,5% - 8,2% = 3,3%	100% / 3,3% = 30
Muertes coronarias	111/ 2221 = 5%	189/2223 = 8,5%	5%/ 8,5% = 0,59	1 - 0,59 = 0,41 = 41%	8,5% - 5% = 3,5%	100% / 3,5% = 29
IAM probado	164 / 2221 = 7,4%	270/ 2223 = 12,1%	7,4%/ 12,1% = 0,61	1 - 0,61 = 0,39 = 39%	12,1% - 7,4% = 4,76%	100% / 4,76% = 21
Mialgias	222/ 2221 = 10%	111/ 2223 = 5%	10%/ 5% = 2	1 - 2 = -1 = -100%	5% - 10% = -5%	100% / -5% = -20

Abreviaturas: ECA: estudio controlado y aleatorizado; NNT: número necesario a tratar para evitar 1 evento; RA_c: Incidencia o % de riesgo en el grupo de control; RA_i: incidencia o % de riesgo absoluto en el grupo de intervención; RAR: reducción absoluta del riesgo; RR: riesgo relativo; RRR: reducción relativa del riesgo.

Y, como la RRR es la medida del efecto que más distorsiona la relevancia clínica (significación práctica) de las intervenciones sanitarias, nosotros no solemos ponerla, quedando nuestras tablas así:

Tabla 1.b: Hombre-Mujer de 58 años, con historia de angina o IAM y con Col y TG elevados.

ECA 4S; mediana de seguimiento 5,4 años	Grupo de intervención Simvastatina; Nº eventos (%), n _i = 2221	Grupo de control Placebo; Nº eventos (%), n _c = 2223	Cálculo por incidencias acumuladas		
	RA _i = nº eventos/ n _i	RA _c = nº eventos/ n _{ic}	RR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)
Muerte por cualquier causa	182/ 2221 = 8,2%	256/ 2223 = 11,5%	8,2%/ 11,5% = 0,71	11,5% - 8,2% = 3,3%	100% / 3,3% = 30
Muertes coronarias	111/ 2221 = 5%	189/2223 = 8,5%	5%/ 8,5% = 0,59	8,5% - 5% = 3,5%	100% / 3,5% = 29
IAM probado	164 / 2221 = 7,4%	270/ 2223 = 12,1%	7,4%/ 12,1% = 0,61	12,1% - 7,4% = 4,76%	100% / 4,76% = 21
Mialgias	222/ 2221 = 10%	111/ 2223 = 5%	10%/ 5% = 2	5% - 10% = -5%	100% / -5% = -20

Abreviaturas: ECA: estudio controlado y aleatorizado; NNT: número necesario a tratar para evitar 1 evento; RA_c: Incidencia o % de riesgo en el grupo de control; RA_i: incidencia o % de riesgo absoluto en el grupo de intervención; RAR: reducción absoluta del riesgo; RR: riesgo relativo.

Una vez completada la tabla, vemos que en la última variable “Mialgias”, simvastatina daña más que placebo.

En los casos de riesgos añadidos (daños), tal como hemos dicho anteriormente, para entender mejor los números, debemos cambiar denominación de la **Reducción Absoluta del Riesgo (RAR)** por su inverso, es decir por el **AAR (Aumento Absoluto del Riesgo)**, cambiando de signo el número. Diremos por tanto que el **AAR es del 5%**.

Esto significa que de cada 100 pacientes que tratamos con simvastatina, 5 padecerán una mialgia “por causa” de simvastatina.

Simultáneamente cambiaremos la denominación **Número Necesario a Tratar (NNT)** por su inverso, es decir por el **NND (Número Necesario para Dañar)**. Diremos por tanto que el **NND es 20**.

Esto significa que si tenemos 20 sillas delante de nuestra consulta con esos pacientes, y tratamos a los 20 con simvastatina durante 5,4 años, dañaremos con mialgia a 1 paciente más que si los tratamos con placebo. Dañaremos a 1 de cada 20 pacientes, sin dañar a los 19 restantes.

4. LOS INTERVALOS DE CONFIANZA

EL INTERVALO DE CONFIANZA DEL PROMEDIO MUESTRAL “ p_i ” DE UNA MUESTRA (m_i) [tomada aleatoriamente de una urna con la POBLACIÓN UNIVERSO], CONTIENE EL PROMEDIO POBLACIONAL “ π ” DE ESA POBLACIÓN TOTAL.

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 “ π ” resultaría $5.250 / 350.000 = 1,5\%$ en un año. Este “ π ” 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%, ($\pi = 1,5\%$ 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 “ π ” 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 “ π ”. Y para eso hacemos un seguimiento de un año a nuestra muestra “ m_1 ” 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 “ π ”.

Ahora bien, como suponemos que “ m_1 ” es una muestra aleatoriamente tomada de la Población española, su promedio muestral puede completarse con los intervalos de confianza al 95%, de modo que $p_i = 1,61\%$ (IC 95%, 1,29% a 1,93%).

Este teorema dice que si repetimos la extracción aleatoria 100 veces, en las mismas condiciones, de una muestra de 6.022 pacientes, es esperable que en 95 veces el promedio poblacional " π " 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 " m_1 ". A nosotros lo que nos interesa es el Promedio Poblacional " π " de la Población española, también llamado *el verdadero valor del parámetro*.

COMO EJEMPLO REAL, EL ESTUDIO RELY.

Recordemos que los investigadores tomaron una muestra aleatoria (m_1) de 6.022 pacientes representativos de la “Población Universo” 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 (m_i), los investigadores van buscando el Promedio Poblacional “ π ” de toda la “Población Universo” de pacientes con fibrilación auricular no valvular de la que se toma la muestra, y pueden afirmar con un nivel de confianza del 95% que “ π ” 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 con la misma probabilidad 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 “ π ” que todos los demás valores del *intervalo de confianza*.

Si los investigadores toman de la Población Universo 100 muestras aleatorias de 6.022 sujetos (m_i) para hacer cien veces el experimento, lo más probable es que tengan el Promedio Poblacional " π " que van buscando entre los 95 intervalos de confianza que corresponden a 95 promedios muestrales (p_i), pero en los 5 experimentos 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 nivel de confianza que se asume en la metodología de la investigación clínica (porque se está dispuesto a asumir un 5% de error como máximo), pero en muchas ocasiones es necesario asumir un máximo de un 1% de error, lo cual es lo mismo que decir que se estamos dispuestos a un nivel de confianza mínimo del 99% de confianza. Error alfa del 5% es el complementario del nivel de confianza del 95%.

Cálculo del intervalo de confianza de una proporción			
IC 95% = $p \pm Z \times \text{Raíz} [p \cdot (1-p) / n]$			
número de eventos	n (tamaño de la muestra)	intervalo de confianza al	en cuántos años
94	6022	95%	1 año
Método de Wald			en cuántos años
Promedio con IC al	95%	1,56% (IC 95%; 1,25% a 1,87%)	1 año

<http://evalmed.es/2020/04/24/05-calculadoras-para-ic-de-a-una-proporcion-y-b-de-una-media/>

Cálculo del intervalo de confianza de una proporción

$$IC\ 95\% = p \pm Z \times \text{Raíz} [p \cdot (1-p) / n]$$

número de eventos	n (tamaño de la muestra)	intervalo de confianza al	en cuántos años
94	6022	90%	1 año
94	6022	95%	1 año
94	6022	99%	1 año
94	6022	99,9%	1 año
Método de Wald			en cuántos años
Promedio con IC al	90%	1,56% (IC 90%; 1,3% a 1,82%)	1 año
Promedio con IC al	95%	1,56% (IC 95%; 1,25% a 1,87%)	1 año
Promedio con IC al	99%	1,56% (IC 99%; 1,15% a 1,97%)	1 año
Promedio con IC al	99,9%	1,56% (IC 99,9%; 1,04% a 2,09%)	1 año

<http://evalmed.es/2020/04/24/05-calculadoras-para-ic-de-a-una-proporcion-y-b-de-una-media/>

Cálculo del intervalo de confianza de una media (mediante t de Student)

$$\text{IC 95\%} = \text{media} \pm t_{\text{teórica según alfa}} \times \text{DE/Raíz}(n)$$

Media	Desviación estándar	Tamaño (n)	intervalo de confianza al
43,12	18	150	95%
		<i>t</i> de Student	
		Media (IC en cada línea)	
Media con IC al	95%	43,12 (IC 95%; 40,22 a 46,02)	

<http://evalmed.es/2020/04/24/05-calculadoras-para-ic-de-a-una-proporcion-y-b-de-una-media/>

Cálculo del intervalo de confianza de una media (mediante t de Student)

$$\text{IC 95\%} = \text{media} \pm t_{\text{teórica según alfa}} \times \text{DE/Raíz}(n-1)$$

Media	Desviación estándar	Tamaño (n)	intervalo de confianza al
43,12	18	150	80%
43,12	18	150	95%
43,12	18	150	99%
43,12	18	150	99,9%

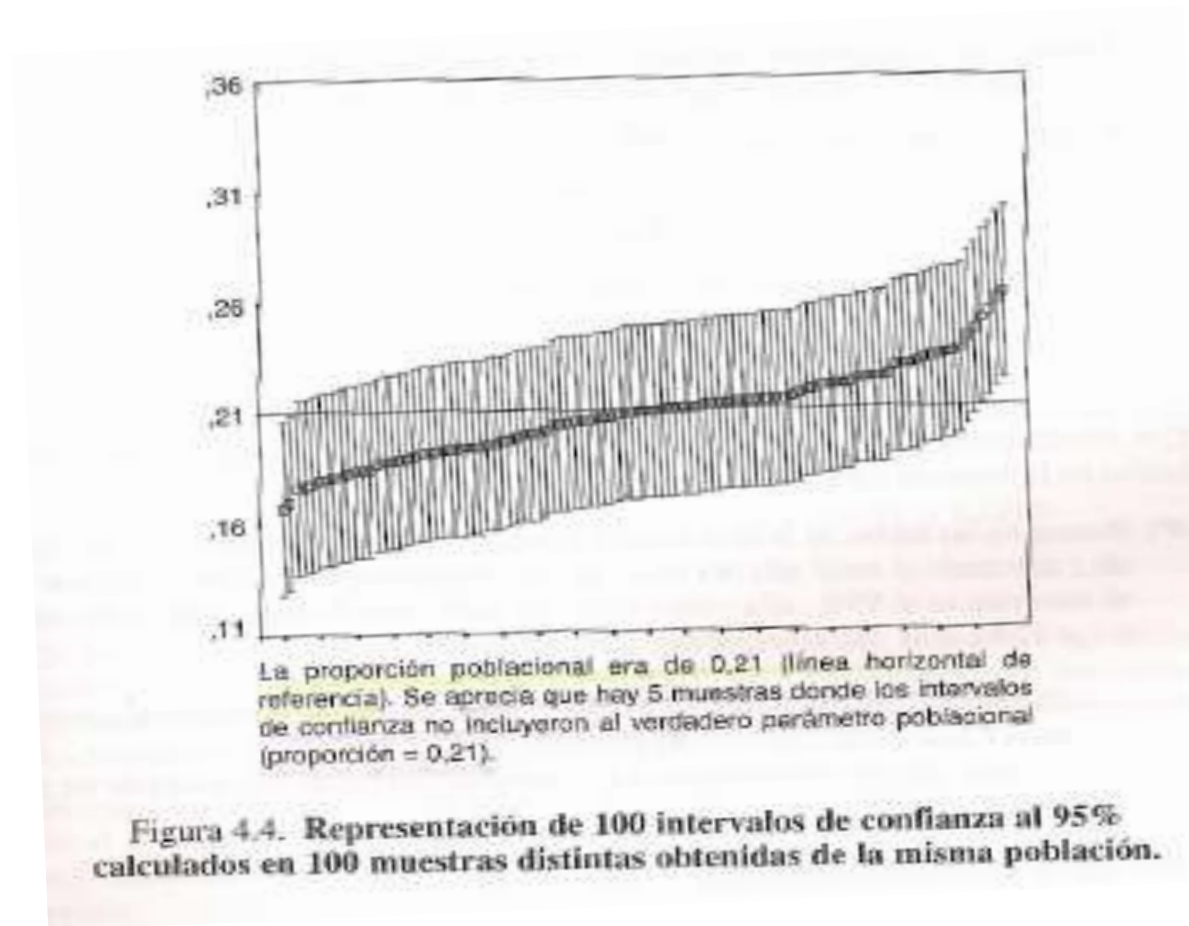
t de Student

Media (IC en cada línea)

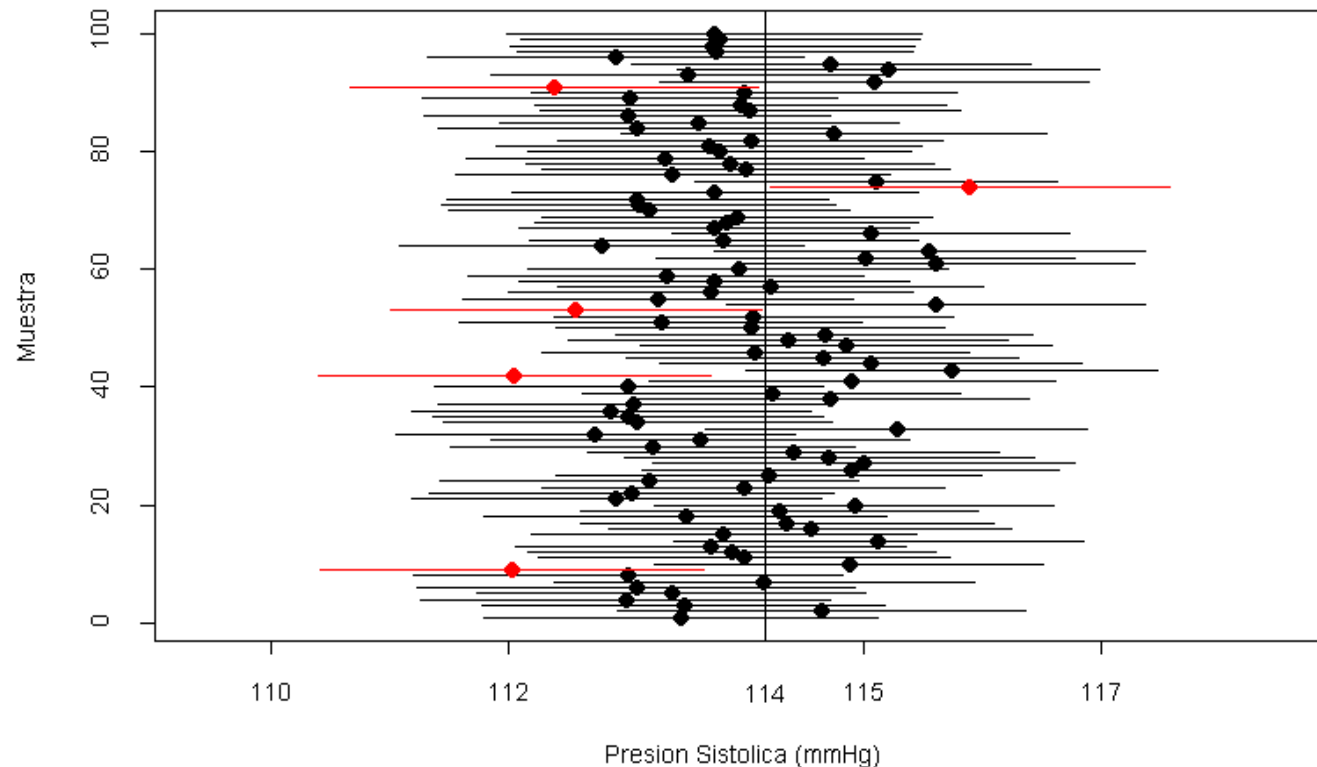
Media con IC al	80%	43,12 (IC 80%; 41,23 a 45,01)
Media con IC al	95%	43,12 (IC 95%; 40,22 a 46,02)
Media con IC al	99%	43,12 (IC 99%; 39,29 a 46,95)
Media con IC al	99,9%	43,12 (IC 99,9%; 38,19 a 48,05)

Abramos un paréntesis para ver dos ejemplos (...

Cien intervalos de confianza al 95% de cien promedios muestrales p_i corroboran el Teorema del Límite Central, en un ejemplo cuyo “Promedio de la Población Universo” $\pi = 21\%$



De una Población Universo $N= 2696$ sujetos calculó la Media Poblacional, que resultó ser $\mu= 114$ mm Hg. A continuación, de esa Población, tomó 100 muestras de $n= 500$ sujetos cada una (con 100 medias muestrales con sus respectivos IC 95%). De las 100 muestras, 95 contenían en su IC la Media Poblacional $\mu= 114$, pero 5 no la contenían.



Cortesía de Emilio Sánchez-Cantalejo (Escuela Andaluza de Salud Pública)

... y cerremos el paréntesis), para continuar.

Cómo se calcula el Intervalo de Confianza de una muestra tomada aleatoriamente:

El intervalo de confianza al 95% de una proporción (p_i) es igual a la **proporción $\pm Z$ multiplicado por el error estándar²** de la proporción. Calculamos el intervalo de confianza al 95% de los 94 eventos que incidieron en la muestra aleatoria (m_1) de 6022 pacientes representativos de la “Población Universo” de pacientes con fibrilación auricular no valvular. La proporción $p_1 = 94 / 6022 = 0,0156$ (=1,56%); entonces $1 - p_1 = 0,9844$. En las tablas encontramos que para el nivel de confianza del 95% la $Z = 1,96$.

$$\text{IC 95\%} = p_1 \pm 1,96 \times \text{Raíz} [p_1 \times (1-p_1) / n_1] = 0,0156 \pm 1,96 \times \text{Raíz} [0,0156 \times (0,9844) / 6022] = 0,0031 (= 0,31\%).$$

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 teórica de que en 95 de ellos el Promedio de la Población Universo π (que es el que perseguimos) se encuentra entre 1,25% y 1,87%.

² El error estándar, error típico o error de muestreo es un error aleatorio que vendría a estimar la variabilidad esperable por el proceso de muestreo. El error estándar es a la muestra (dentro de una distribución de 100 muestras), lo que la desviación estándar es al individuo (dentro de una distribución de 100 individuos).

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 intervalos de confianza del NNT son los inversos de los intervalos de confianza de la RAR.

Para facilitar su cálculo, proporcionamos una sencilla calculadora en <http://evalmed.es/2020/04/28/01-calculadora-variables-dicotomicas/>, con la que es muy fácil introducir los datos (dentro de las casillas con fondo amarillo). 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).

Figura 1: Calculadora para variables dicotómicas: RR, HR, RAR y NNT

Disponible en: <http://evalmed.es/2020/04/28/01-calculadora-variables-dicotomicas/>

Cálculo por incidencias acumuladas 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: intervalo de confianza.

IC: 95%

Aquí elegimos el nivel (o intervalo) de confianza.

	Enferman Con eventos	No enferman Sin eventos	Total
Intervención	83	917	1.000
Control	135	865	1.000
Total	218	1.782	2.000

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...

Nº event Interv (%)	Nº event Control (%)	RR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)	Potencia	Valor de p para la diferencia	% Intervención (Fact Box)	% Control (Fact Box)
83/1000 (8,3%)	135/1000 (13,5%)	0,61 (0,47-0,8)	5,2% (2,46% a 7,93%)	19 (13 a 41)	96,17%	0,000	8,30%	13,50%

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 de contraste (potencia estadística) resultante
Aquí el valor de p para la diferencia.

Aquí obtenemos la hoja de información al usuario (Fact Box), que no se manejan con el intervalo de confianza. [OJO: Es una simplificación].

Para un nivel de confianza, pongamos por ejemplo del 95%, la diferencia entre dos grupos no es estadísticamente significativa cuando concurren alguna de las siguientes circunstancias:

- a) se solapan los intervalos de confianza al 95% del $\%RA_i$ y $\%RA_c$;
- b) el intervalo de confianza al 95% del RR atraviesa el uno;
- c) el intervalo de confianza al 95% de la RAR atraviesa el cero;
- d) el intervalo de confianza al 95% del NNT “atraviesa el infinito”.

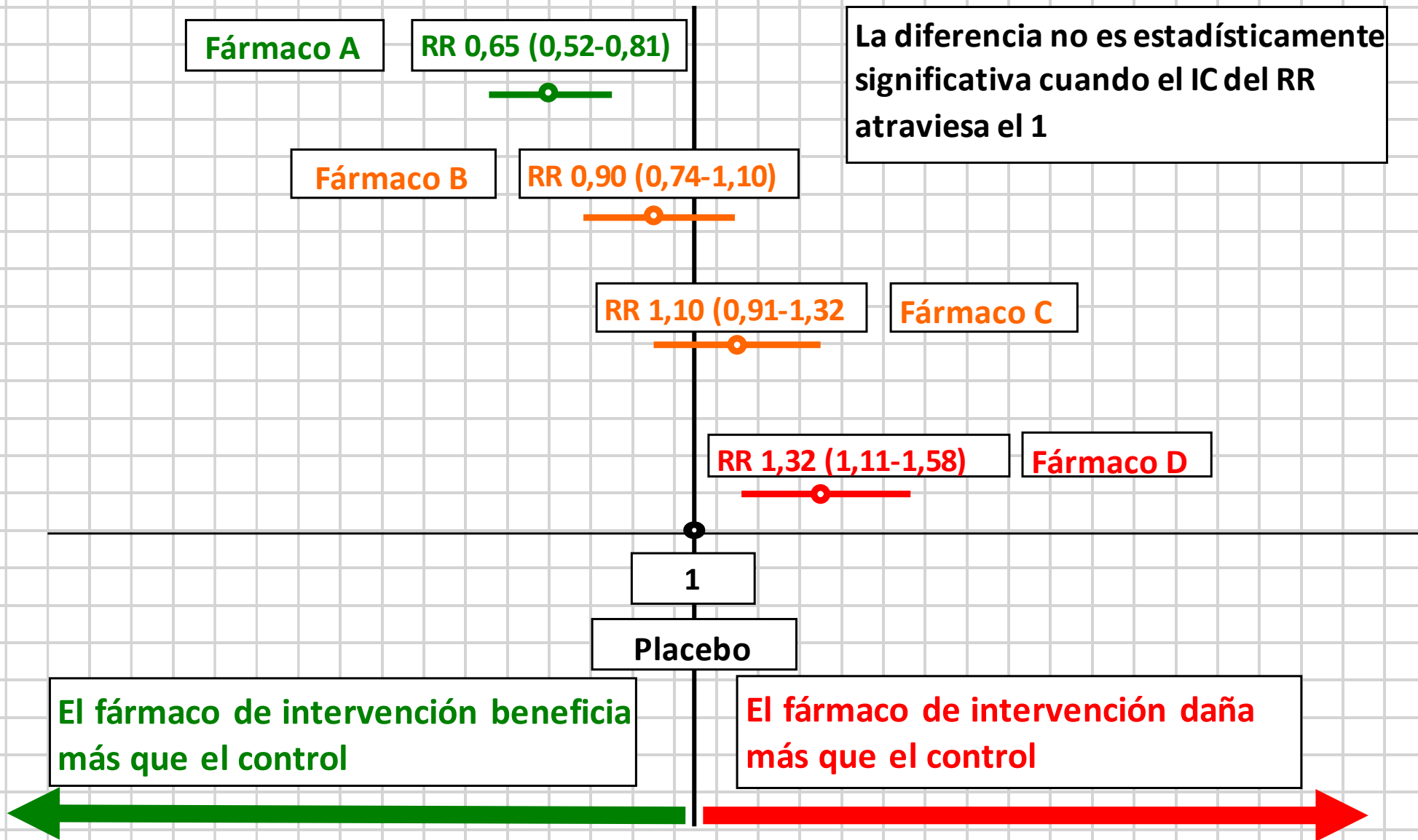
Como se ha expresado más arriba, cuando el intervalo de confianza del RR atraviesa el 1, se dice que no se ha encontrado una diferencia estadísticamente significativa entre el numerador ($\%RA_i$) y el denominador ($\%RA_c$); lo cual quiere decir que no puede descartarse que el numerador y el denominador pertenezcan a un mismo grupo, y que, aunque no sean exactamente iguales, sus diferencias son compatibles con las diferencias que hay en dos muestras aleatorias de un mismo grupo.

De la misma manera, cuando el intervalo de confianza de la RAR atraviesa el cero, se dice que no se ha encontrado una diferencia estadísticamente significativa entre el minuendo ($\%RA_c$) y el sustraendo ($\%RA_i$); lo cual quiere decir que no puede descartarse que el minuendo y el sustraendo pertenezcan a un mismo grupo, y que, aunque no sean exactamente iguales, sus diferencias son compatibles con las diferencias que hay en dos muestras aleatorias de un mismo grupo

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”.

Figura 2: Representación gráfica del riesgo relativo. $RR = \%Ra_i / \%Ra_c$



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.

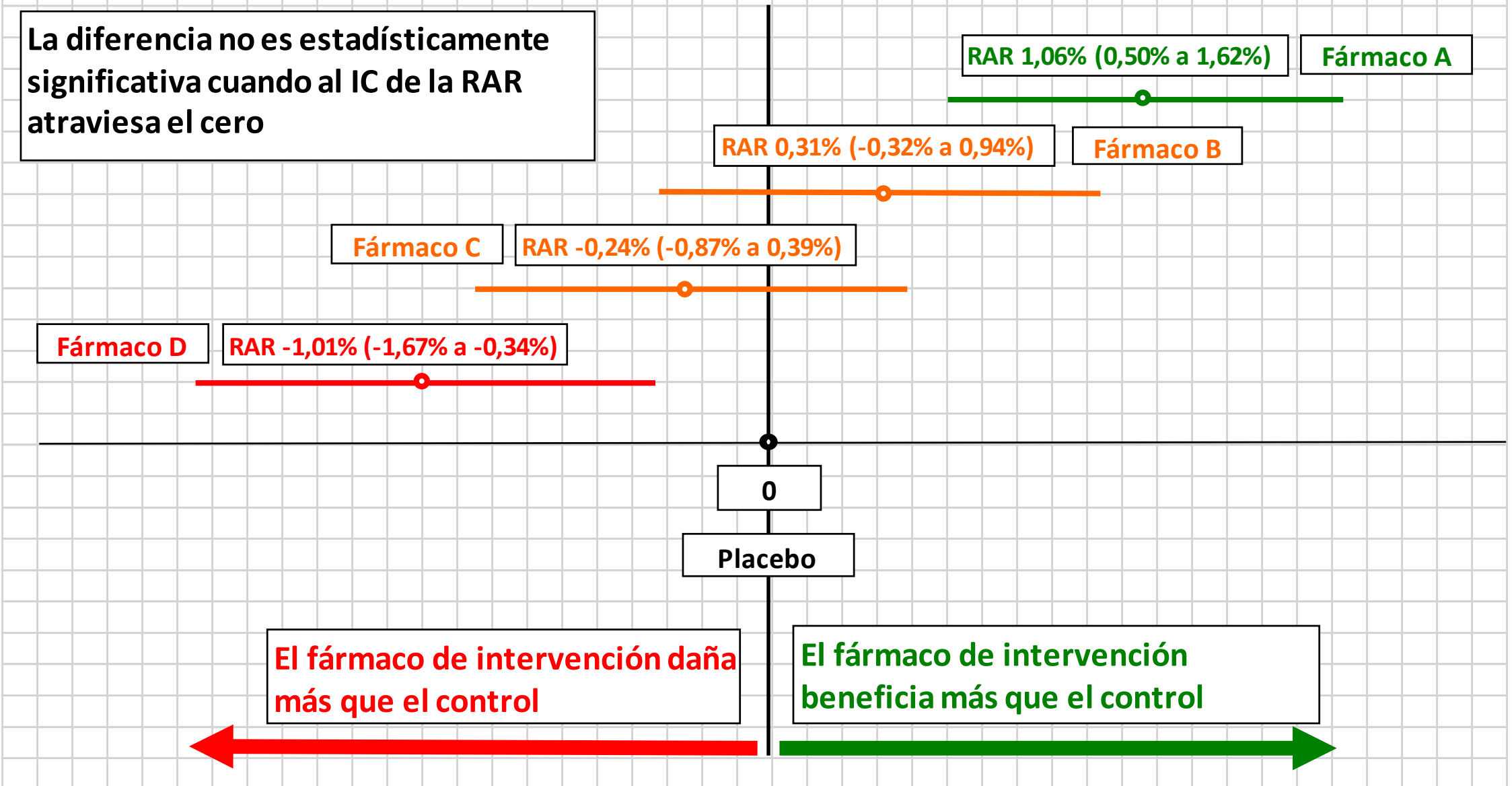
No se encuentra diferencia estadísticamente significativa entre los Fármacos B y C frente a Placebo. Al comparar el “% de infartos” con el Fármaco B frente a Placebo se obtuvo un **RR 0,90 (0,74-1,10)**, que sí atraviesa el “1”. Decimos entonces que no hay diferencia estadísticamente significativa entre 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.

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).

Figura 3: Representación gráfica del riesgo relativo. $RAR = \%RA_c - \%RA_i$



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%)**.

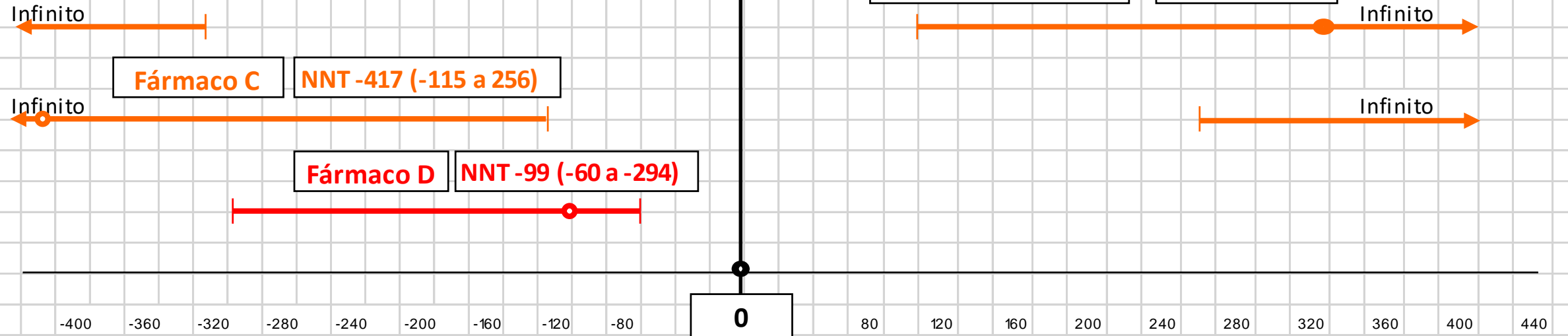
Se encuentra un Daño añadido estadísticamente significativo del Fármaco D frente a Placebo. Al comparar el “% de infartos” con el Fármaco D frente a Placebo se obtuvo una **RAR -1,01% (-1,67% a -0,34%)**, cuyo intervalo de confianza no atraviesa el “1”. Por ello decimos que hay diferencia estadísticamente 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%)**.

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).

Figura 4: Representación gráfica del Número Necesario a Tratar para evitar "un" evento. $NNT = 1 / RAR$

La diferencia no es estadísticamente significativa cuando al IC del NNT "atraviesa el infinito"



El fármaco de intervención daña más que el control

El fármaco de intervención beneficia más que el control



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 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)**.

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

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 el resultado no nos informa de si ambos porcentajes eran 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.

Para demostrar de otra forma adicional la imposibilidad de estimar la relevancia clínica, obsérvese que con el RR la única referencia que hay respecto a población inicial son los denominadores del Riesgo Absoluto del grupo de intervención y del Riesgo Absoluto del grupo de control. Nos resulta cómodo ajustar los numeradores de ambos al igualar la población de referencia siempre a 100, con el fin de obtener dos sencillos tantos por ciento. Se ve claramente que ahora ambas poblaciones están representadas en el 100 de ambos denominadores. Pues bien, al ser ambos denominadores iguales a 100, en una ratio éstos se autoeliminan, lo que trae como consecuencia que perdemos toda referencia a la población inicial, y esto significa que no puede estimarse si la relevancia clínica es alta, moderada, baja o insignificante.

Esto también se ve claramente incluso manteniendo los denominadores originales, como puede observarse en la [tabla 2](#).

Tabla 2: Cuatro ejemplos simulados que muestran cómo en la división se pierde la población de referencia, con lo que ante un mismo RR = 0,5, no puede inferirse si la relevancia clínica es alta, moderada, baja o insignificante.

Prevención de evento grave en 1 año	Nº event Interv (%)	Nº event Control (%)	RR	RAR	NNT
Supuesto 1	20/100 (20%)	40/100 (40%)	0,5	20%	5
Supuesto 2	20/1.000 (2%)	40/1.000 (4%)	0,5	2%	50
Supuesto 3	20/10.000 (0,2%)	40/10.000 (0,4%)	0,5	0,2%	500
Supuesto 4	20/100.000 (0,02%)	40/100.000 (0,04%)	0,5	0,02%	5000

$$RR = \frac{20/100}{40/100} = \frac{20}{40} = 0,5 \qquad RR = \frac{20/100}{40/100} = \frac{20/\cancel{100}}{40/\cancel{100}} = 0,5$$

$$RR = \frac{20/1.000}{40/1.000} = \frac{20}{40} = 0,5 \qquad RR = \frac{20/1.000}{40/1.000} = \frac{2/\cancel{100}}{4/\cancel{100}} = 0,5$$

$$RR = \frac{20/10.000}{40/10.000} = \frac{20}{40} = 0,5 \qquad RR = \frac{20/10.000}{40/10.000} = \frac{0,2/\cancel{100}}{0,4/\cancel{100}} = 0,5$$

$$RR = \frac{20/100.000}{40/100.000} = \frac{20}{40} = 0,5 \qquad RR = \frac{20/100.000}{40/100.000} = \frac{0,02/\cancel{100}}{0,04/\cancel{100}} = 0,5$$

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, porque la suma o la resta de porcentajes mantienen la población de referencia en el 100 de sus denominadores, y el resultado de la suma o la resta mantiene el 100 en el denominador como población de referencia. Pero esto no puede hacerse con los resultados de dividir dos porcentajes de eventos, es decir con dos RR.

La RRR tiene estos mismos inconvenientes.

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.

1) 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.

2) Sin embargo, 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%.

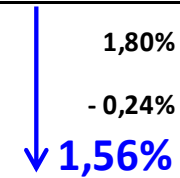
Verifíquese esta afirmación comprobando que la simple suma de las RAR situados en la vertical coincide con la RAR obtenida operando con los ACV totales situados en la horizontal (tabla 3).

Tabla 3: 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)

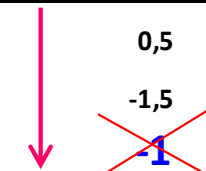
Obsérvese que por el camino vertical (RAR) se llega al mismo **1,56%** que el obtenido por el horizontal.

$$4,08\% - 2,52\% = 1,56\%$$



Obsérvese que por el camino vertical (RR) no se llega al mismo **0,62** que el obtenido por el horizontal.

$$2,52 / 4,08 = 0,62$$



A pesar de las ventajas para calcular (o estimar) el balance de beneficios y riesgos, 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 una construcción 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).

**Abramos un paréntesis para practicar con el
estudio ACCORD (...)**

... y cerremos el paréntesis) para continuar.

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 4S, su NNT en 5,4 años es 30; es decir, **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 los eventos desde 33/1.000 a 13/1.000, es decir en $20/1.000 = 2\%$, con lo cual tenemos una $RAR = 2\%$ y un $NNT = 50$.*

A pesar de que la reducción de eventos de 250/1.000 a 230/1.000 es la misma reducción absoluta (20/1.000), es decir una $RAR = 2\%$ y un $NNT = 50$, el nivel de ansiedad del paciente y su familia es diferente si saben que el riesgo basal es de 33/1.000 (3,3%) que si es de 250/1.000 (25%)» (6).

6. LOS TRES DESTINOS DEL NNT MEDIANTE LA REGLA DEL 1

Nosotros, conscientes de la utilidad de esta información, la hemos incorporado completa en lo que venimos llamando **los tres destinos del NNT** mediante 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).

Cuando el control es otro medicamento activo, los tres destinos del NNT mediante la REGLA DEL 1 nos dicen que: a) por cada 1 paciente en el que el medicamento de intervención es más efectivo que el medicamento control; b) el número de los que tendrán el evento con el medicamento de intervención (igual que con el medicamento de control); y c) 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 <http://evalmed.es/> , concretamente en los siguientes vínculos:

“Los tres destinos del NNT mediante la Regla del 1, cuando el control es placebo”:

<http://evalmed.es/2020/04/26/03-la-regla-del-1-cuando-el-control-es-placebo/>

y “Los tres destinos del NNT mediante la Regla del, cuando el control es un medicamento activo”:

<http://evalmed.es/2020/04/25/04-la-regla-del-1-cuando-el-control-es-un-medicamento-activo/>

Un ejemplo de los tres destinos del NNT mediante la Regla del 1, cuando el control es un 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 en: <http://evalmed.es/2020/04/26/03-la-regla-del-1-cuando-el-control-es-placebo/>, 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).

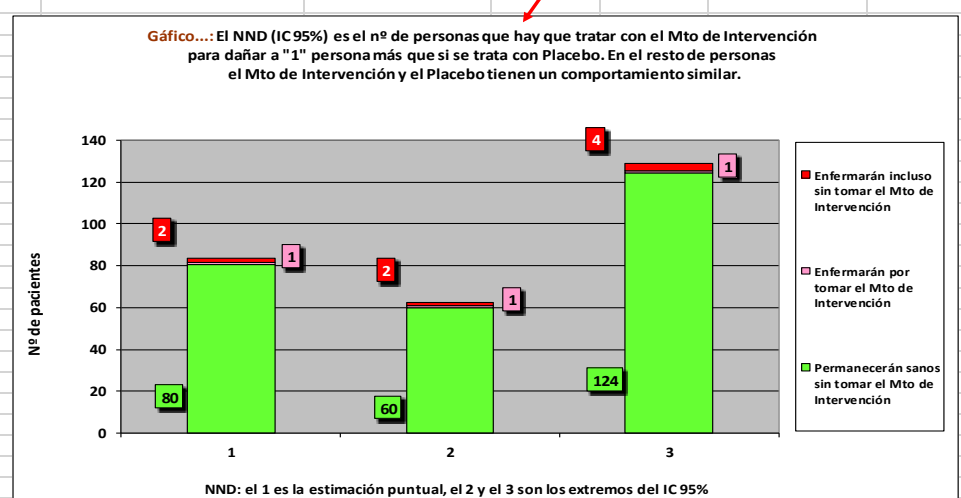
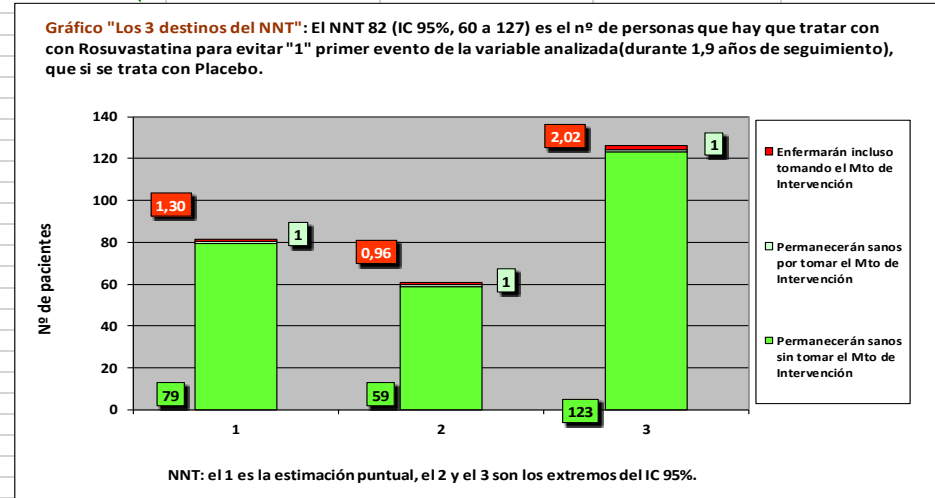
Figura 5: Hoja de cálculo para obtener los tres destinos del NNT mediante la Regla del 1. Disponible en: <http://evalmed.es/2020/04/26/03-la-regla-del-1-cuando-el-control-es-placebo/>

Cálculo por incidencias acumuladas 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%.						
	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			
Nº event Interv (%)	Nº event Control (%)	RR (IC 95%)	RRR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)	Potencia
142/8901 (1,5953%)	251/8901 (2,8199%)	0,57 (0,46-0,69)	43,43% (53,85% a 30,64%)	1,2246% (0,7878% a 1,6541%)	82 (60 a 127)	99,98%
Valor de p para la diferencia						
0,00						
Aplicase 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 libres de evento sin tomar el Mto de Intervención			79	59	123	
Permanecerán libres de evento por tomar el Mto de Intervención			1	1	1	
Tendrán el evento incluso tomando el Mto de Intervención			1,30	0,96	2,0	
Aplicase únicamente cuando el NNT y sus intervalos de confianza son NEGATIVOS			Estimación puntual	Límite inferior del IC	Límite superior del IC	
NND =			82	60	127	
Permanecerán sanos sin tomar el Mto de Intervención			80	60	124	
Enfermarán por tomar el Mto de Intervención			1	1	1	
Enfermarán incluso sin tomar el Mto de Intervención			2	2	4	

En las casillas con fondo amarillo se introducen el Nº de eventos

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

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 denominamos NND.



Cálculo por incidencias acumuladas 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%.

	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

En las casillas con fondo amarillo se introducen el Nº de eventos

Automáticamente se obtiene RR, RRR, 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%)	RRR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)	Potencia	Valor de p para la diferencia
142/8901 (1,5953%)	251/8901 (2,8199%)	0,57 (0,46-0,69)	43,43% (53,85% a 30,64%)	1,2246% (0,7878% a 1,6541%)	82 (60 a 127)	99,98%	0,00

Aplicase 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 libres de evento sin tomar el Mto de Intervención	79	59	123
Permanecerán libres de evento por tomar el Mto de Intervención	1	1	1
Tendrán el evento incluso tomando el Mto de Intervención	1,30	0,96	2,0

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 denominamos NND.

Aplicase únicamente cuando el NNT y sus intervalos de confianza son NEGATIVOS

	Estimación puntual	Límite inferior del IC	Límite superior del IC
NND =	82	60	127
Permanecerán sanos sin tomar el Mto de Intervención	80	60	124
Enfermarán por tomar el Mto de Intervención	1	1	1
Enfermarán incluso sin tomar el Mto de Intervención	2	2	4

Tabla...: Persona de 66 años (IQR 60-71), normolipémica, normotensa, con PCR de alta sensibilidad 4,2 mg/dl (IQR 2,8-7,2)						
ECA JUPITER, media de seguimiento 1,9 años	Rosuvastatina, n= 8901	Placebo, n= 8901	Cálculo por incidencias acumuladas en 1,9 años			
	Nº eventos (%)	Nºeventos (%)	RR (IC, 95%)	RAR (IC, 95%)	NNT (IC, 95%) en 1,9 años	Potencia
Primer evento de [Muerte CV, IAM no fatal, ACV no fatal, hospitaliz por angina y revascularización]	142/8901 (1,6%)	251/8901 (2,82%)	0,57 (0,46-0,69)	1,22% (0,79 a 1,65)	82 (60 a 127)	99,98%

- 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 $2,8\% - 1,6\% = 1,22\% \Rightarrow$ **Éste es el beneficio asociado con rosuvastatina**.

Conclusión: 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 éstos 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 $82 \times 97,2\% = 79,4$; Enfermarán incluso con fármaco: $82 \times 1,6\% = 1,3$.

	Estimación puntual
NNT	81,7
Permanecerán sanos sin tomar el Mto de Intervención	79,4
Permanecerán sanos por tomar el Mto de Intervención	1,0
Enfermarán incluso tomando el Mto de Intervención	1,3

Rosuvast		Placebo
Sí evento	1	Sí evento
No evento	2	Sí evento
No evento	3	No evento
No evento	4	No evento
No evento	5	No evento
No evento	6	No evento
No evento	7	No evento
No evento	8	No evento
No evento	9	No evento
No evento	10	No evento
No evento	11	No evento
No evento	12	No evento
No evento	13	No evento
No evento	14	No evento
No evento	15	No evento
No evento	16	No evento
No evento	17	No evento
No evento	18	No evento
No evento	19	No evento
No evento	20	No evento
No evento	21	No evento
No evento	22	No evento
No evento	23	No evento
No evento	24	No evento
No evento	25	No evento
No evento	26	No evento
No evento	27	No evento
No evento	28	No evento
No evento	29	No evento
No evento	30	No evento
No evento	31	No evento
No evento	32	No evento
No evento	33	No evento
No evento	34	No evento
No evento	35	No evento
No evento	36	No evento
No evento	37	No evento
No evento	38	No evento
No evento	39	No evento
No evento	40	No evento
No evento	41	No evento
No evento	42	No evento
No evento	43	No evento
No evento	44	No evento
No evento	45	No evento
No evento	46	No evento
No evento	47	No evento
No evento	48	No evento
No evento	49	No evento
No evento	50	No evento
No evento	51	No evento
No evento	52	No evento
No evento	53	No evento
No evento	54	No evento
No evento	55	No evento
No evento	56	No evento
No evento	57	No evento
No evento	58	No evento
No evento	59	No evento
No evento	60	No evento
No evento	61	No evento
No evento	62	No evento
No evento	63	No evento
No evento	64	No evento
No evento	65	No evento
No evento	66	No evento
No evento	67	No evento
No evento	68	No evento
No evento	69	No evento
No evento	70	No evento
No evento	71	No evento
No evento	72	No evento
No evento	73	No evento
No evento	74	No evento
No evento	75	No evento
No evento	76	No evento
No evento	77	No evento
No evento	78	No evento
No evento	79	No evento
No evento	80	No evento
No evento	81	No evento
No evento	82	No evento

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 rosuvastatina es efectiva”, permanecerán sanos sin tomar rosuvastatina entre 59 y 123 pacientes y sufrirán un evento tomando rosuvastatina entre 1 y 3 pacientes.

	Estimación puntual	Límite inferior del IC	Límite superior del IC
NNT	81,7	60	127
Permanecerán sanos sin tomar el Mto de Intervención	79,4	59	123
Permanecerán sanos por tomar el Mto de Intervención	1,0	1	1
Enfermarán incluso tomando el Mto de Intervención	1,3	1	3

De lo observado **podemos extraer un corolario.**

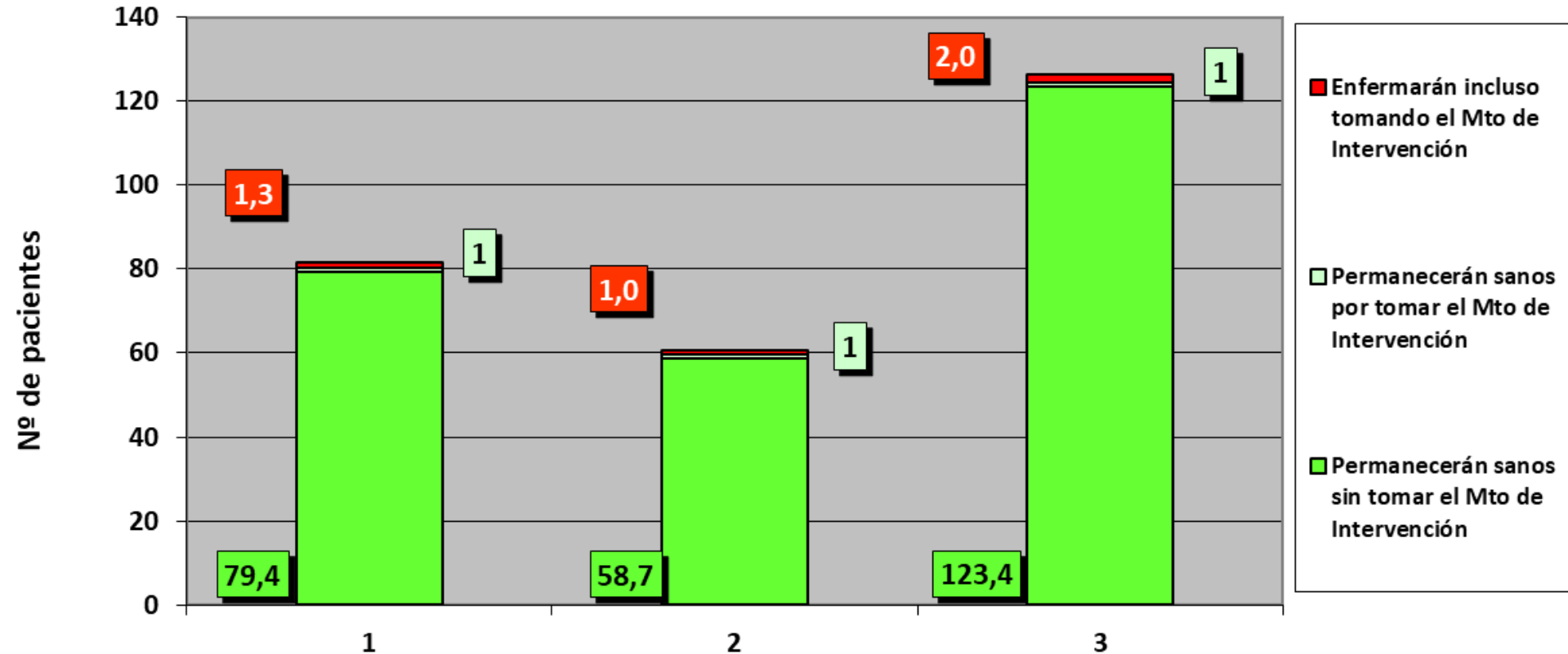
Gran parte de las personas atribuye como mérito de rosuvastatina frente a placebo el número de pacientes que permanece “sano” por tomar dicho fármaco.

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 82** (sin beneficiar a los 80 que permanecen sanos igual que con placebo, ni a 1 que enferma incluso con rosuvastatina).

Sin embargo, la falta de precisión les llevan a “creer” que beneficia a $1+80 = 81$ por cada 82.

La diferencia entre la creencia (*realidad ingenua*) y la *realidad real* es enorme, pues la creencia lleva a atribuir erróneamente un mérito de 81 por cada 82, cuando en realidad es 1 por cada 82.

Gráfico "Los 3 destinos del NNT": El NNT 82 (IC 95%, 60 a 127) es el nº de personas que hay que tratar con Rosuvastatina para evitar "1" primer evento de la variable analizada(durante 1,9 años de seguimiento), que si se trata con Placebo.



NNT: el 1 es la estimación puntual, el 2 y el 3 son los extremos del IC 95%.

7. RELEVANCIA CLÍNICA DE LAS INTERVENCIONES PREVENTIVAS

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 fenomenológicamente objetiva, y *lo es para todos*.

Como la *percepción* fenomenológicamente 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, todos supieron formular, 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 escribieran y 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 Pedro 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 Humphrey, **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 proponerse una escala ordinal de grados de aversión al riesgo, cuyo orden de aversión o preferencia desde el punto de vista biológico³ sea lo más universal posible (bajo la suposición de que mantiene la impronta evolutiva).

La mayor aversión de la biología es la muerte. En la civilización humana además, preservar la vida es una tarea suprema que no sólo resulta moralmente intachable sino que es el fundamento de todos los preceptos morales (9). Y en una escala del 1 al 9, la aversión a la muerte ocuparía el 9, a sabiendas de que (más allá de la biología) las personas pueden llegar a tener o imaginar estados más aversivos que la propia muerte (a los que asignaríamos genéricamente como “mayores de 9”)⁴.

³ 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.

⁴ Las escalas propuestas son constructos cuya fiabilidad y validez deben medirse.

Un observador puede no saber exactamente cómo de distintas son las sensaciones subjetivas de cien 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, revascularización coronaria y hospitalización (10).

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 aversiva 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 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.

Un inconveniente es que al ser números ordinales no nos permiten colegir que el ordinal “8” significa el doble de aversión que el ordinal “4”. Pero afortunadamente existen métodos estadísticos que nos permiten relacionar matemáticamente los rangos que ocupan las variables ordinales en una escala de rangos.

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 (11).

Nosotros, formando un panel, hemos consensuado una escala orientativa sobre la magnitud del efecto de las **intervenciones preventivas**⁵ 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**, como prototipo de importancia 9.

⁵ Según el objetivo terapéutico distinguimos entre prevención y curación. La **prevención** es la reducción del riesgo basal de un evento que está en potencia, y así **un 3%/año de riesgo basal de fractura de cadera** en un tipo de individuos significa que, de cada 100, probablemente 3 individuos de ese tipo tendrán fractura de cadera en un año, y 97 permanecerán sin fractura. Una intervención preventiva pretenderá reducir el riesgo basal en un 1% (o NNT = 100), es decir de 3 a 2 individuos con fractura por año. Esos 2 individuos que sufren el evento de fractura de cadera pasan a la clase “pérdida de salud respecto a su situación anterior”. La **curación** pretenderá restituir total o parcialmente esa pérdida de salud en los 2 (dos de cada dos, es decir en el 100%, o NNT = 1), o al menos en 1 (uno de cada dos, es decir en el 50%, o NNT = 2).

TOMANDO COMO REFERENCIA NUESTRO CONSENSO SOBRE EL NNT 150 EN MORTALIDAD (como prototipo de puntuación ordinal 9), PODRÍAMOS ATREVERNOS A PONER CALIFICATIVOS A LA REDUCCIÓN DE RIESGOS BASALES GRAVES CON PUNTUACIÓN ORDINAL 8 y 7.

A la calificación de magnitud del efecto alta para un $NNT \leq 133$ 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 $NNT \leq 117$ en 1 año para una intervención preventiva de importancia 7, como por ejemplo la angina de pecho (tabla 3).

Tabla 3: Escalas de consenso interno.

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.

**8. LA RAR Y EL NNT SE PUEDEN
INTERPOLAR PERO NO EXTRAPOLAR**

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 3%, entonces podemos interpolarlo a 1 año dividiéndolo por 2, obteniendo una RAR = 1,5% 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 un error 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 32, entonces podemos interpolarlo a 1 año multiplicándolo por 2, obteniendo un NNT = 64 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 (12).

LA FALACIA DE LA INMORTALIDAD POR EXTRAPOLACIÓN AL FUTURO DEL EFECTO OBSERVADO EN DOS AÑOS.

Las medidas del efecto RAR y el NNT no se pueden extrapolar por encima de los años que duró el ensayo clínico, **porque además se llegaría a la inmortalidad**. Mostramos el siguiente ejemplo, que es similar al atentado contra la bioestadística que utiliza Ridker para la propaganda de la rosuvastatina en prevención primaria. El estudio JÚPITER se clausuró en 1,9 años. Por motivos pedagógicos, consideremos que eran 2 años. Sabemos que este fármaco demostró en el JÚPITER un NNT de 182 para la Mortalidad por todas las causas. Vamos a ponérselo mucho mejor a Ridker, y en lugar de ese NNT tan irrelevante clínicamente, que el NNT hubiera sido 32. Matemáticamente, cuando los años se multiplican por 2, entonces la RAR se multiplicaría por 2 y el NNT se dividiría por 2.

Falacia de la inmortalidad por extrapolación al futuro del efecto obtenido sobre la mortalidad en 2 años.		
años	RAR	NNT
1	1,5%	64
2	3%	32
4	6%	16
8	13%	8
16	25%	4
32	50%	2
64	100% (*)	1 (*)

(*) Al violar la regla de no extrapolar el NNT, la falacia nos da la apariencia de la inmortalidad. Un NNT de 1 significa la inmortalidad, porque por cada 1 en el que actúa, evita la muerte en ese 1.

9. EL EFECTO MARCO

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 ser humano confunde en muchas situaciones lo lógico con lo psicológico (psicologismo), 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 (13, 14).

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 reveló que **es más frecuente la aversión a correr riesgos en las decisiones que se enmarcan como ganancias (vivirán 400 de 600) y más frecuente la tendencia a correr riesgos en las decisiones que se enmarcan como pérdidas (morirán 200 de 600).**

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: Si mi objetivo es que la población sana participe en el programa de mamografía: Elegir entre “si usted se hace una mamografía, tendrá un x% menos de cáncer” frente a “si usted no se hace una mamografía, tendrá un x% más de cáncer”.

Siendo una variante de “efecto marco”, según algunos autores (15, 16), 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 (16), que cuando se presenta en forma de RAR o de NNT.

NOTA: Esto mismo sucede con las pruebas diagnósticas cuando éstas se presentan como medida de su efectividad pronóstica con la sensibilidad.

Esto puede conducir al **sobreatamiento (y sobrediagnóstico)** en pacientes (o poblaciones) **con un riesgo basal bajo**, exponiéndolos a efectos adversos con baja o nula expectativa de beneficios (18, 19).

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 (16, 20-32).

Otro tanto sucede cuando un mismo dato de terapia oncológica se presenta como probabilidad de vivir (supervivencia) frente a probabilidad de morir, resultando estadísticamente más atractiva la elección en el primer marco (33).

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 6 hasta tener que tratar a 77 pacientes para conseguir el mismo rendimiento: evitar a 1 paciente el evento grave.

Tabla 4: Cinco ejemplos simulados 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 (%), n= 1000	Nº event Control (%), n= 1000	RR	RAR	NNT
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

Propaganda comercial para inducir el “efecto marco”, mediante la RRR del 35%, y provocar una sobreestimación del efecto. Tras realizar los cálculos, para la variable [ACV o Embolismo sistémico] los eventos fueron: 1,11%/año en el grupo de DAB 300 frente a 1,71%/año en el grupo de WARF; RR 0,65 (0,52-0,81); RAR 0,60% (0,32% a 0,82%); NNT 168 (122 a 310) por año.

Más grave es la alusión de que DAB tiene este efecto frente al grupo de WARF con buen control del INR, lo cual no es cierto, pues no se encontraron diferencias cuando WARF tiene un %TRT $\geq 64,5\%$.

POR PRIMERA VEZ, EN MÁS DE 50 AÑOS, SE HA DESARROLLADO UN NUEVO ANTICOAGULANTE ORAL QUE HA DEMOSTRADO SER MÁS EFICAZ Y SEGURO QUE LA TERAPIA EXISTENTE¹

Pradaxa® 150 mg c/12h ha demostrado una mayor reducción del ictus al compararlo con warfarina^{1,2}

Reducción en un 35% del riesgo relativo de ictus o embolia sistémica frente a warfarina bien controlada (INR 2,0 - 3,0)^{1,2}

Tratamiento	Tasa anual de ictus o embolia sistémica (%)
warfarina (INR 2,0 - 3,0)	1,71
Pradaxa® 150 mg c/12h	1,11

Gráfico construido a partir de Connolly S et al. N Engl J Med 2009; 361:1138-47 y Connolly SJ et al. N Engl J Med 2010; 363:1875-8.

Pradaxa® 150 mg c/12h previene 3 de cada 4 ictus asociados a fibrilación auricular³

En el estudio RE-LY³ la media de edad de los pacientes fue de 71 años, y el 63,6% fueron hombres³

POR PRIMERA VEZ, EN MÁS DE 50 AÑOS, SE HA DESARROLLADO UN NUEVO ANTICOAGULANTE ORAL QUE HA DEMOSTRADO SER MÁS EFICAZ Y SEGURO QUE LA TERAPIA EXISTENTE¹

Pradaxa® 150 mg c/12h ha demostrado una mayor reducción del ictus al compararlo con warfarina^{1,2}

Reducción en un 35% del riesgo relativo de ictus o embolia sistémica frente a warfarina bien controlada (INR 2,0 - 3,0)^{1,2}

LA MEDIA VERDAD QUE FALTA
 Cálculos en time to event (en 1 año)
 Para la variable [ACV o Embolismo sistémico] los eventos fueron: 1,11% en el grupo de DAB 300 frente a 1,71% en el grupo de WARF; RR 0,65 (0,52-0,81); RAR 0,60% (0,32% a 0,82%); NNT 168 (122 a 310) por año.
 Cálculos por riesgos acumulados (en 2 años)
 Para la variable [ACV o Embolismo sistémico] los eventos fueron: 134 (2,21%) para el grupo de DAB 300 frente a 202 (3,35%) para el grupo de WARF. RR 0,65 (0,52-0,81); RAR 1,15% (0,56% a 1,73%); NNT 87 (58 a 180) en años, y potencia resultante 97,1%.
 Ambos resultados son similares, el primero expresado en 1 año y el segundo en 2 años de seguimiento. Tal resultado lo estimamos de magnitud del efecto baja en términos de NNT.

No citaron que en un análisis de subgrupos por tiempo en rango terapéutico con warfarina (TRT en INR 2-3) realizado por los investigadores del RELY, para la variable [hemorragia intracranial] no se encontraron diferencias estadísticamente significativas cuando se comparó DAB 300 con WARF con buen control del INR (TRT $\geq 64,5\%$), pues efectivamente los eventos fueron 65 / 2898 (2,3%) frente a 83 / 2898 (2,84%), RR 0,81 (0,59-1,11), RAR 0,54% (-0,28% a 1,84%), NNT 187 (75 a -360). Omisieron decir que con DAB 300 no se encontró diferencia estadísticamente significativa en la variable [ACV o Embolismo sistémico].

Tratamiento	Tasa anual de ictus o embolia sistémica (%)
warfarina (INR 2,0 - 3,0)	1,71
Pradaxa® 150 mg c/12h	1,11

Gráfico construido a partir de Connolly S et al. N Engl J Med 2009; 361:1138-47 y Connolly SJ et al. N Engl J Med 2010; 363:1875-8.

Pradaxa® 150 mg c/12h previene 3 de cada 4 ictus asociados a fibrilación auricular³

En el estudio RE-LY³ la media de edad de los pacientes fue de 71 años, y el 63,6% fueron hombres³

Propaganda comercial (¿e institucional?) para inducir el “efecto marco”, mediante la efectividad o RRR del 83% de la vacuna 7vPnC, y provocar una sobreestimación del efecto. Tras realizar los cálculos, para la variable [Enfermedad Neucomócica Invasiva total] en menores de 2 años en USA los eventos fueron: 309/969.103 (0,03189%) en el postvacunal año 2009 frente a 943/512.769 (0,18390%) en el prevacunal año 1999; RR 0,17 (0,15-0,2); RRR 83% (80% a 85%); RAR 0,15% (0,14% a 0,16%); NNT 658 (610 a 717).

En la **Mortalidad por ENI** no se encontraron estadísticamente significativas.

Tabla 1: Tasas por 100.000 de enfermedad neumocócica invasiva (ENI) y mortalidad por ENI 1997-2010, extraídas de los informes del Núcleo de Vigilancia Activa Bacteriológica (ABCs) del CDC, constituido por 9 estados de Estados Unidos (*)										
Evolución Estados Unidos	Hasta los 2 años		Hasta los 5 años		De 18 años en adelante		De 50 años en adelante		Extrapolado a toda la población de Estados Unidos	
	Tasas ENI	Tasas Mortalidad por ENI	Tasas ENI	Tasas Mortalidad por ENI	Tasas ENI	Tasas Mortalidad por ENI	Tasas ENI	Tasas Mortalidad por ENI	Tasas ENI	Tasas Mortalidad por ENI
Año 1997	160,7	2,5	82,76	2,69	22,9	2,9	40,89	6,50	23,10	2,27
Año 1998	183,8	1,7	95,67	1,67	22,3	2,9	40,54	6,22	23,30	2,20
Año 1999	183,9	1,0	93,67	0,70	22,8	3,4	40,83	7,65	23,63	2,70
Año 2000 (#)	154,7	3,2	79,87	1,58	20,7	2,9	38,13	6,32	38,13	2,30
Año 2001	60,2	0,9	40,93	0,66	18,4	3,0	32,87	6,71	17,10	2,40
Año 2002	35,0	0,8	23,75	0,44	16,0	2,5	28,88	5,34	14,20	2,00
Año 2003	38,5	0,8	23,41	0,43	15,7	2,3	28,02	5,19	13,90	1,90
Año 2004	34,4	1,1	21,26	0,55	14,5	2,1	26,37	4,58	12,90	1,70
Año 2005	35,9	0,6	21,78	0,44	15,9	2,0	28,07	4,33	14,00	1,60
Año 2006	35,9	0,6	21,89	0,40	16,1	2,2	28,10	4,52	13,80	1,70
Año 2007	36,5	0,6	22,53	0,28	15,8	1,9	27,76	3,88	13,90	1,50
Año 2008	34,0	0,9	20,81	0,34	16,5	1,9	28,37	3,78	14,50	1,50
Año 2009	31,9	0,6	21,28	0,26	15,8	1,9	27,99	4,13	14,30	1,60
Año 2010	28,0	0,2	18,77	0,18	14,6	1,6	26,31	3,32	12,90	1,30

(*) Disponible en URL: <http://www.cdc.gov/abcs/reports-findings>

(#) La inmunización de rutina en niños con la vacuna 7vPnC en Estados Unidos comenzó en febrero del año 2001

ENI: enfermedad neumocócica invasiva

10. CON EL VALOR DE LA p SE PUEDE ESTIMAR LA SIGNIFICACIÓN ESTADÍSTICA, PERO NO SE PUEDE ESTIMAR LA SIGNIFICACIÓN PRÁCTICA (RELEVANCIA CLÍNICA)

El “valor de la p ” de una muestra respecto a la media de una distribución de muestras aleatoriamente tomadas

Muchos autores se han decantado claramente a favor de los intervalos de confianza y adversamente hacia el “valor de la 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 por azar (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 por azar, 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):

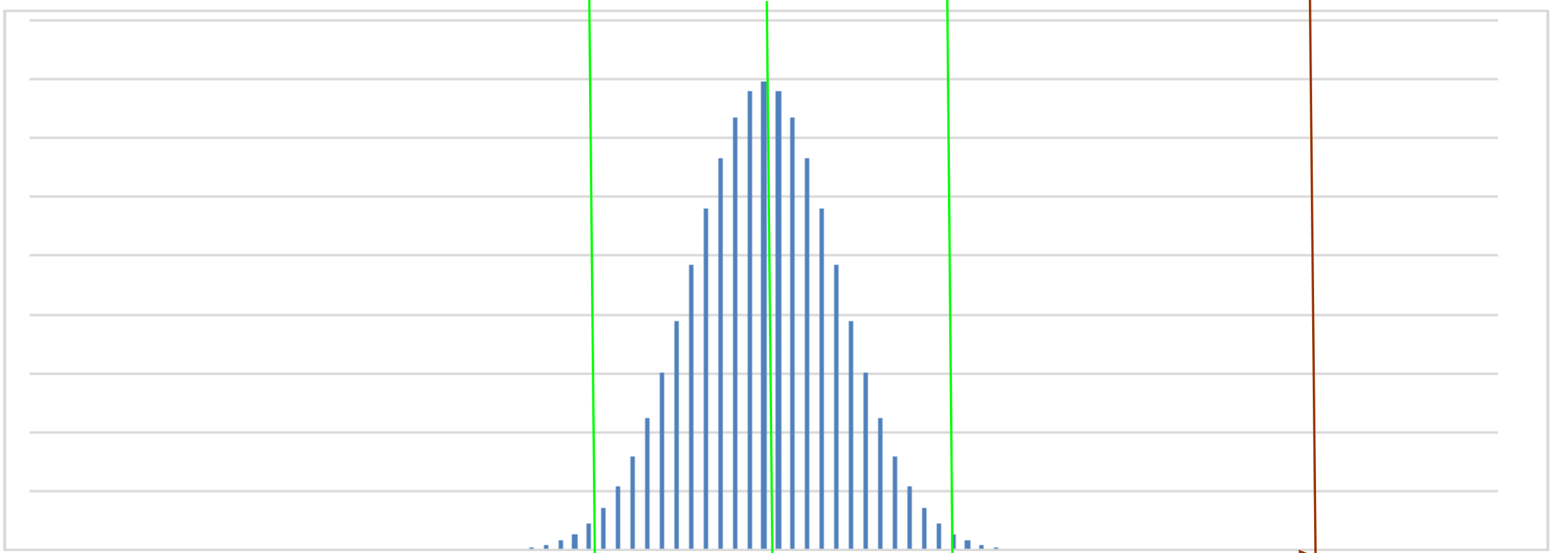
a) Hipótesis nula (H_0): mantiene que el efecto de interés no existe (es nulo, es decir, vale 0) en la Población de la que procede la muestra.

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

Para verlo mejor, tomemos por ejemplo una muestra de 6.022 personas con fibrilación auricular

En una muestra de 6.022 personas con fibrilación auricular, el porcentaje de ACV en un año fue el 1,5% (IC 95%, 1,2% a 1,8%). Esto significa que el % de ACV por año de la historia natural de la enfermedad se encuentra entre 1,2% y 1,8% con un nivel de confianza del 95%.

Si de otra muestra de tamaño similar, el promedio de ACV por año es del 2%, la probabilidad de que ésta sea una muestra de la historia natural de la enfermedad es $p = 0,002$. La hipótesis nula (H_0) es que esta muestra pertenece a la historia natural de la enfermedad. Pero la probabilidad de que eso sea verdad es tan baja, que estoy dispuesto a apostar por la hipótesis alternativa (H_1) de que es una muestra de personas con más ACV/año que la fibrilación auricular.



> 1,20% 1,20% 1,50% 1,80% > 1,80% 2%

Error alfa 2,5% <-] [-> Nivel de confianza 95% [-> Error alfa 2,5% ----->]

2,5%	2,0%	1,5%	1,0%	0,5%	0,0%
------	------	------	------	------	------

0,025	0,020	0,015	0,010	0,005	0,000
-------	-------	-------	-------	-------	-------

p (1 cola)

0,050	0,040	0,030	0,020	0,010	0,000
-------	-------	-------	-------	-------	-------

p (2 colas)

→
p = 0,002

En sentido estricto, la interpretación de un “valor de la 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 $p = P(\text{dif esperada} \geq \text{dif observada} \mid H_0)$, y se expresa así porque es una probabilidad condicionada a que H_0 sea cierta. Aunque nos resulte incómodo estamos obligados a formular la hipótesis nula (H_0) de que “no hay diferencias entre los promedios de los dos grupos que se comparan”, y la hipótesis alternativa (H_1) de que “sí hay diferencias entre los promedios de los dos grupos que se comparan”.

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 otro 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 (tabla 5):

Tabla 5: Ejemplos para observar cómo cambia el valor de p para la diferencia, sobre la simulación de los efectos (efectividades) de tres vacunas frente a placebo contra una infección, durante 1 año de seguimiento.

Tres ensayos clínicos simulados de tres vacunas, durante 1 año de seguimiento.	Vacunas de los supuestos 1, 2 y 3; n° eventos (%) n = 200	Placebo; n° eventos (%) n = 200	Cálculo por incidencias acumuladas <u>EN 1 AÑO</u>			
			RR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)	Potencia

Valor de p para la diferencia

Hoja información al usuario que no se maneja con los IC

Nº de pacientes con evento en 1 año por cada 100 tratados con:

Vacuna	Placebo
--------	---------

Incidencia de infección

Supuesto 1	40/200 (20%)	40/200 (20%)	1 (0,68-1,48)	0% (-7,85% a 7,85%)	-----	2,5%
Supuesto 2	36/200 (18%)	40/200 (20%)	0,9 (0,6-1,35)	2% (-5,69% a 9,71%)	50 (10 a -18)	7,35%
Supuesto 3	20/200 (10%)	40/200 (20%)	0,5 (0,3-0,82)	10% (3,01% a 17%)	10 (6 a 33)	79,97%

1,000
0,610
0,005

20%	20%
19%	19%
10%	20%

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 10%, por lo que tenemos muchos argumentos para rechazar la hipótesis nula, y decimos que la diferencia encontrada si es estadísticamente significativa.

A pesar de su utilidad, es importante señalar que **con el valor de p se puede estimar la significación estadística, pero no se puede estimar la significación práctica (relevancia clínica)**

Un mismo valor de p puede corresponder a una magnitud de efecto alta o baja. 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. A pesar de tener la misma significación estadística, la significación práctica (relevancia clínica) es muy distinta (tabla 6).

Tabla 6: Ejemplos simulados de dos vacunas que muestran el mismo valor de $p = 0,005$ (0,5%), con muy distinta relevancia clínica en términos de NNT.

Dos ensayos clínicos simulados de dos vacunas, durante 1 año de seguimiento.	Vacunas simuladas según se describen en su línea; nº eventos (%)	Placebo; nº eventos (%)	Cálculo por incidencias acumuladas <u>EN 1 AÑO</u>			
			RR (IC 95%)	RAR (IC 95%)	NNT (IC 95%)	Potencia

Valor de p para la diferencia

Hoja información al usuario que no se maneja con los IC

Nº de pacientes con evento en 1 año por cada 100 tratados con:

Vacuna	Placebo
--------	---------

Incidencia de infección

Vacuna A	20/200 (10%)	40/200 (20%)	0,5 (0,3-0,82)	10% (3,01% a 17%)	10 (6 a 33)	79,97%	0,005	10%	20%
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)	80,05%	0,005	5%	6%

BIBLIOGRAFÍA.

1. Pinker S. *Cómo funciona la mente*. Ferrán Meler-Orti (trad.). Barcelona. Ed. Destino. 2001. Cap 5: Buenas ideas. p. 436-7.
2. Pedersen TR, Kjeldsen S, 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*. José M^a Lebrón (trad.). Barcelona. Ed Gedisa. 1995. Cap 4: La doble provincia de los sentidos. p. 49-56.
9. Bauman Z. *Mortalidad, inmortalidad y otras estrategias de vida* (Del original: *Mortality, immortality and other life strategies*, 1992). Javier Eraso Ceballos (trad.). Madrid. Ed Sequitur 2014. Cap 5: La modernidad o deconstruyendo la mortalidad. p. 211.
10. Stolker JM, Spertus JA, Cohen DJ, Jones PG, Jain KK, Bamberger E, Lonergan BB, Chan PS. Rethinking composite end points in clinical trials: insights from patients and trialists. *Circulation*. 2014 Oct 7;130(15):1254-61.
11. Librero J, Peiró S. Comparaciones abstrusas en los resúmenes de ensayos clínicos en revistas médicas españolas. *Med Clín*. 2001;117:732-3.
12. Palomo L, Sánchez-Robles G. Extrapolating the NNT jeopardises biostatistics and leads to immortality. *Aten Primaria*. 2011;44(5):296-7.
13. Tversky A, Kahneman D. The framing of decisions and the rationality of choice. 1981;211(4481):453-8.
14. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science*. 1974;185(4157):1124-31.
15. 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.
16. 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.

17. 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.
18. Smith GD, Egger M. Who benefits from medical interventions? *BMJ*. 1994;308(6921):72-4.
19. Brett AS. Treating hypercholesterolemia. How should practicing physicians interpret the published data for patients? *N Engl J Med*. 1989;321(10):676-80.
20. 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.
21. 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.
22. Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet*. 1994;343(8907):1209-11.
23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. Perneger TV, Agoritsas T. Doctors and patients' susceptibility to framing bias: a randomized trial. *J Gen Intern Med*. 2011;26(12):1411-7.
33. 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.