

# ¿Cuándo optar por algo de peor calidad?

Víctor Dayan<sup>1</sup>

Diariamente, cuando salimos de compras, sin darnos cuenta realizamos estudios de no-inferioridad (NI). Imaginen cuando ustedes salen a comprar un nuevo equipo de microondas: evalúan distintas marcas, distintos precios, garantías del producto, y probablemente terminen sin elegir el mejor, más duradero y más caro (marca A), sino uno que cumpla su función de forma aceptable dentro de los parámetros definidos por ustedes. O sea, definimos de manera inconsciente un margen de *no inferioridad* respecto a la marca A y terminamos comprando el microondas que no sea peor a ese margen de inferioridad. En otras palabras, estamos dispuestos a comprar un microondas un poco peor, con tal de ahorrarnos un dinero.

En contraste con un ensayo clínico convencional, en el que lo habitual es que se intente demostrar la superioridad de un tratamiento A vs. un tratamiento B o placebo, en un estudio de NI el objetivo es demostrar que el tratamiento A es no inferior al control activo B. Aquí nos detendremos en dos conceptos: NI y control activo. Cuánto de lo *no inferior* es aceptable es en realidad un criterio clínico que habitualmente se desprende de la revisión sistemática de la evidencia disponible. Por otro lado, al referirnos a *control activo*, en los estudios de NI el brazo control es siempre el tratamiento estándar en el momento (por eso es *activo* y no *placebo*). Cuando un tratamiento es considerado de NI, quiere decir que el probable efecto “peor” en relación con el tratamiento estándar es clínicamente aceptable, considerando los otros beneficios que tiene. Estos otros beneficios habitualmente son un menor costo, una mayor disponibilidad o un tratamiento menos invasivo.

En la tabla 1, a continuación, se esquematizan las principales diferencias entre los estudios de superioridad (a los cuales estamos acostumbrados) y los de NI.

En un ensayo de NI, lo que se intenta demos-

trar es que el nuevo tratamiento es no inferior por determinado margen (el margen corresponde a una diferencia en el evento primario) al tratamiento estándar, mientras que en los estudios de superioridad se quiere demostrar si la diferencia entre ambos tratamientos es igual o no a 0. Por tal razón es que el error alfa se distribuye en dos colas (0,025 en cada una, 0,05 en total) para los estudios de superioridad y en una cola en los de NI (0,025). Un aspecto de importancia en la interpretación de estos resultados es cómo pueden verse afectados por la irregularidad en la conducción o la falta de seguimiento en estos ensayos. Como concepto, cuanto más irregular o más carente de cumplimiento de los protocolos (por alta tasa de cruzamientos, pérdidas en el seguimiento, poca adherencia al tratamiento), más difícil es encontrar diferencias entre los grupos. Por lo tanto, estas irregularidades afectan al nuevo tratamiento en los estudios de superioridad, pero lo favorecen en los estudios de NI. Si, por ejemplo, intentamos encontrar diferencias en la supervivencia entre el tratamiento A y B, pero encontramos que al año un 30% de los pacientes se perdieron, dicha pérdida de pacientes hará que sea más difícil encontrar diferencias entre ellos. Sin embargo, si la intención es demostrar la NI, será, por lo tanto, más fácil que así ocurra. Finalmente, mientras que el análisis de preferencia para los estudios de superioridad es por intención de tratar, para los de NI se realiza por protocolo. El análisis por intención de tratar tiende a diluir el efecto del tratamiento explorado (debido a cruzamientos, irregularidades durante el estudio); este es el análisis de elección en los estudios de superioridad, por ser más conservador. Si recordamos brevemente, en el análisis por intención de tratar se analizan dentro de cada grupo los pacientes desde el momento que fueron randomizados (independientemente de que hayan recibido el tratamiento o no). Puede ocurrir

1. Centro Cardiovascular Universitario.

Recibido Nov 9, 2021; aceptado Mar 14, 2022

**Tabla 1.** Comparación entre estudios de superioridad y no inferioridad.

	<i>Superioridad</i>	<i>No inferioridad</i>
Comparador	Tratamiento estándar o placebo	Tratamiento estándar
Margen de NI	No se aplica	Predeterminado
Hipótesis nula	2 tratamientos son iguales	El tratamiento nuevo es inferior por > margen
Hipótesis alternativa	Hay una diferencia entre tratamientos	El tratamiento nuevo es inferior por < margen
Error alfa	0,05 (dos colas)	0,025 (una cola)
Irregularidades durante el ensayo	Penaliza el nuevo tratamiento	Favorece el nuevo tratamiento
Análisis conservador	Intención por tratar	Por protocolo

Tomado de Kaul S<sup>(1)</sup>.

que el paciente randomizado para el tratamiento A no lo reciba, porque fallece antes o porque cambia de opinión o porque no puede realizarlo. En este caso, aunque reciba el tratamiento B igual sigue siendo analizado como parte del grupo A, determinando que el efecto real del tratamiento A “se diluya”. El análisis por protocolo, en cambio, analiza solo los pacientes que cumplieron con el protocolo, o sea, aquellos que recibieron el tratamiento tal cual está descrito, sin desvíos. De esta forma se logra evaluar específicamente el efecto del tratamiento. En los estudios de NI, el análisis por intención por tratar puede favorecer la NI (al “diluirse” la diferencia). Por dicha razón, la FDA sugiere que la NI sea demostrada también en el análisis por protocolo<sup>(1,2)</sup>.

### Interpretación de los estudios de no inferioridad

Dado lo complejo del tema, evitaremos adentrarnos en los pasos necesarios para definir el margen de NI para un estudio en particular. Tan solo diremos que para poder definir un margen adecuado deben cumplirse dos premisas: sentido común clínico y evidencia previa que avale dicho margen. Respecto al sentido común no creemos que sea necesaria su descripción. La evidencia previa se recolecta mediante una revisión sistemática y un metaanálisis del tema. Consideremos que, por ejemplo, queremos evaluar si el consumo de chocolate barato (B) es no-inferior al chocolate clase A para mejorar la mortalidad a largo plazo. La razón para conducir este estudio es que el chocolate B es un chocolate genérico más barato. El primer paso consiste en realizar una revisión sistemática y metaanálisis de todos los estudios que comparan el consumo de chocolate clase A vs. placebo y su relación con la mortalidad. En la figura 1 se describe el diagrama de bosque de dicho metaanálisis. El uso del chocolate clase A se asocia con una disminución en la

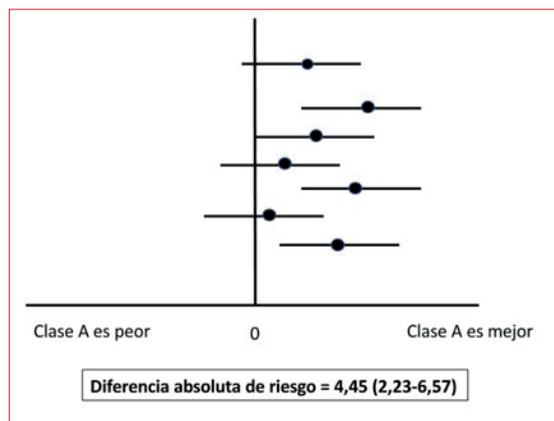
mortalidad de 4,45%, con un intervalo de confianza que va de 2,23% a 6,57%. En otras palabras, el menor efecto positivo del chocolate clase A en comparación con el placebo es de 2,23%.

Esto quiere decir que el margen de NI al evaluar el chocolate B no puede ser mayor que 2,23, dado que, de lo contrario, existe el riesgo de que el chocolate B sea considerado no inferior al de clase A, pero que en realidad sea peor o igual al placebo. A partir de este valor, el margen puede incluso hacerse menor, dependiendo del grado en que se pretenda preservar el efecto del control activo (clase A). Cuanto menor es dicho margen, más similar al efecto del control activo se exigirá para definir la NI. El extremo sería un margen de 0, en el que se pretende que no haya diferencias entre el chocolate clase B y clase A para definir NI.

Quizás sea básico, pero considero oportuno en este momento recordar lo que significa el intervalo de confianza (IC). El IC comúnmente aceptado del 95% es al que nos referiremos. Si les digo que la probabilidad de muerte es de 4,5% con un IC del 95% que va de 3,5% a 5%, ¿qué significa? ¡No quiere decir que el 95% de los pacientes tendrán una mortalidad entre 3,5 y 5! Quiere decir que, si repetimos el estudio 100 veces, en el 95% de las veces la mortalidad se ubicará entre esos valores.

### Tipos de margen de no inferioridad

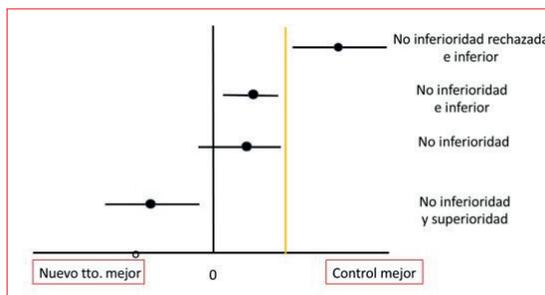
Los márgenes de NI pueden ser de dos tipos: riesgo absoluto o riesgo relativo. Los márgenes de riesgo absoluto se definen como la diferencia máxima esperada entre ambos tratamientos en la unidad de la variable resultado evaluada. Para los márgenes de riesgo relativo, lo que se utiliza es la proporción máxima esperada en el riesgo con el tratamiento nuevo. Se prefiere utilizar márgenes relativos dado que estos son fijos y no cambian según la diferencia entre el resultado esperado y el observado. Veamos unos ejemplos.



**Figura 1.** Diagrama de bosque resultante del metaanálisis de estudios que evalúan la diferencia en la mortalidad a largo plazo en pacientes con consumo de chocolate clase A vs. placebo.

### Estudio PRECOMBAT

El ensayo clínico PRECOMBAT fue un estudio de NI que evaluó como resultado primario la incidencia de muerte, *stroke*, revascularización e infarto al año en pacientes que son sometidos a cirugía de revascularización miocárdica (CRM) vs. angioplastia percutánea (AP) en lesión de tronco principal izquierdo no protegida<sup>(3)</sup>. En su construcción, los investigadores asumieron una incidencia del evento primario en la rama de CRM del 13%, y con base en esto definieron un margen de NI absoluto de 7%. ¿Eso qué quiere decir? Quiere decir que para que la AP sea considerada no inferior a la CRM, el límite superior del IC para la diferencia del evento primario entre ambos tratamientos no debe superar el 7%. ¿Y qué fue lo que ocurrió? El evento primario en la rama de AP fue de 8,7% y en la rama de CRM fue de 6,7%. Por lo tanto, la diferencia en el riesgo absoluto fue de 2%, y su IC 95% iba de -1,6% a 5,6%. El límite superior del IC del 95% fue de 5,6% y, dado que está por debajo de lo preespecificado (7%), el estudio concluye que es no inferior. Sin embargo, recordemos que cuando se planificó el estudio, el margen de NI se definió con base en una incidencia esperada en la rama de CRM de 13%, y no con base en la incidencia real observada (6,7%). Considerando la incidencia esperada (13%), lo que se esperaba era que, para definir la NI, la incidencia de eventos en AP fuera menor a un 54% ( $13\% + 7\%/13\%$ ) por encima de la incidencia en CRM. En otras palabras, usando un margen de riesgo relativo, se espera un margen de 1,54. Cuando la incidencia de eventos observada es menor que la que se usó para definir el margen absoluto, lo que ocurre es que se “infla” el margen relativo. Veamos qué pasó con PRECOMBAT. Como vimos, el margen relativo que corresponde



**Figura 2.** Esquema ilustrativo de las conclusiones derivadas de un estudio de NI de acuerdo con los extremos de los intervalos de confianza para un margen de riesgo absoluto. En naranja se representa el margen de no inferioridad.

al absoluto de 7% (para un esperado de 13%) era de 1,54. Ahora, con una incidencia observada real de 6,7%, al mantener el mismo margen absoluto el margen relativo trepa a 2,04 ( $6,7\% + 7\%/6,7\%$ ). O sea que, mientras que originariamente se rechazaba la NI si la incidencia del evento primario en AP era mayor a un 54% que en CRM, ahora se rechaza si es mayor a un 104%. Así pues, el margen relativo se “infló” de 1,54 a 2,04, favoreciendo la NI. Para que esto no ocurra, se prefiere usar siempre márgenes relativos desde el inicio, dado que no se modifican según la incidencia de eventos<sup>(1)</sup>.

### ABSORB III<sup>(4)</sup>

Este ensayo de NI evaluó el uso de *stents* bioabsorbibles vs. el control activo (*stent* con everolimus) en la falla de la lesión tratada. Con base en una incidencia estimada del evento en el control activo de 7%, el margen de NI establecido para el evento primario fue de 4,5%. La incidencia observada del evento primario en el control activo fue de 6,1%. La diferencia entre ambos fue de 1,7% (-0,5% a 3,9%). ¿Se logró demostrar la NI tanto mediante el uso de margen absoluto como relativo? La respuesta vendrá en el próximo número.

### Resultados probables luego de un estudio de no inferioridad

En la figura 2 se ilustran las distintas posibilidades que pueden resultar del estudio de NI. Como concepto general, debe identificarse los extremos del IC del 95% y ver cómo cae cada uno respecto al margen de NI y al 0 o 1, según se considere un margen de riesgo absoluto o relativo. En el caso de la figura, cada vez que el extremo de la derecha del IC atraviesa el margen de riesgo absoluto (línea naranja), se dice que se rechaza la NI. Luego, hay que ver qué pasa con los extremos respecto al

0 (en caso de diferencias absolutas) o el 1 (en caso de riesgo relativo). Si ambos extremos están del lado derecho (el control es mejor), el nuevo tratamiento es inferior; si ambos están del lado izquierdo (el tratamiento experimental es mejor), el nuevo tratamiento es superior. Por lo tanto, de un estudio de NI se desprende tanto si la NI se logró o rechazó como si además el estudio fue inferior o superior.

Recientemente se publicó el ensayo clínico de NI FAME 3, que evalúa la AP guiada por FFR vs. la CRM en pacientes con lesión severa de 3 vasos coronarios. El margen de NI relativo fue de 1,65. El resultado de FAME 3 mostró un riesgo relativo de 1,5 (IC 95% de 1,1-2,2). Eso implica que, por estar uno de los extremos del IC por encima de 1,65, se rechaza la NI. Concomitantemente, por estar ambos extremos por encima de 1 (riesgo relativo), se confirma la inferioridad del tratamiento experimental.

Como hemos visto, los estudios de NI son extremadamente débiles en el sentido que los criterios usados en definir la no inferioridad son subjetivos. Aun más, quienes lo definen pueden tener conflictos de interés que pueden interferir en la decisión objetiva de estos márgenes. Por tal razón, estamos obligados a analizar minuciosa y críticamente la metodología, cuestionándola cuando sea necesario. Solo así podremos brindar la mejor atención basada en evidencia a nuestros pacientes.

**Victor Dayan**, ORCID 0000-0002-5470-0585

Editor responsable: Dr. Federico Ferrando

### Bibliografía

1. Kaul S. Understanding the Merits and Drawbacks of Noninferiority Trials in Cardiovascular Medicine. *Can J Cardiol.* 2021; 37(9):1378-1393. doi: 10.1016/j.cjca.2021.05.017.
2. Non-inferiority clinical trials to establish. Guidance for industry [Internet]. Ciudad: U.S Department of Health and Human Services Foods and Drug Administration; 2021. Disponible en: <https://www.fda.gov/media/78504/download>. (consulta: 21 Nov 2021)
3. Park DW, Ahn JM, Park H, Yun SC, Kang DY, Lee PH, et.al. Ten-Year Outcomes after drug-eluting stents versus coronary artery bypass grafting for left main coronary disease: extended follow-up of the PRECOMBAT Trial. *Circulation.* 2020 ;141(18):1437-46. doi: 10.1161/CIRCULATIONAHA.120.046039.
4. Ellis SG, Kereiakes DJ, Metzger CD, Caputo RP, Rizik DG, et al. Everolimus-eluting bioresorbable scaffolds for coronary artery disease. *N Engl J Med.* 2015;373(20):1905-15. doi: 10.1056/NEJMoa1509038.

El autor elaboró en su totalidad el presente artículo.