

**COLABORACIÓN
 ESPECIAL**

Recibida: 18/5/2022
 Aceptada: 16/9/2022
 Publicada: 29/9/2022
 e202209074

e1-e13

Non-inferiority studies efficacy and safety: review of methodological and statistical aspects in clinical trials and non-interventional studies with medicinal products.

Los autores declaran que no existe ningún conflicto de intereses.

Las opiniones expresadas en esta publicación son responsabilidad exclusiva de los autores y no reflejan necesariamente las de los organismos en los que trabajan.

CORRESPONDENCIA

Diana González Bermejo
 Agencia Española de Medicamentos y Productos Sanitarios,
 C/ Campezo, 1, Edificio 8,
 CP 28022, Madrid, España.
 dgonzalezb@aemps.es

CITA SUGERIDA

González Bermejo D, López-Fando Santafé A, Fernández Dueñas A, Montero Corominas D, Huerta Álvarez C. Estudios de no inferioridad de eficacia y seguridad: revisión e interpretación de aspectos metodológicos y estadísticos en ensayos clínicos y estudios observacionales con medicamentos. Rev Esp Salud Pública. 2022; 96: 29 de septiembre e202209074.

Estudios de no inferioridad de eficacia y seguridad: revisión e interpretación de aspectos metodológicos y estadísticos en ensayos clínicos y estudios observacionales con medicamentos

AUTORES

Diana González Bermejo (1)
 Almudena López-Fando Santafé (2)
 Ana Fernández Dueñas (3)
 Dolores Montero Corominas (1)
 Consuelo Huerta Álvarez (4)

CONTRIBUCIONES DE AUTORÍA

Todos los autores han contribuido a:
 · La concepción y el diseño, adquisición de datos o su análisis e interpretación.
 · Redactar el artículo o hacer una revisión crítica de su contenido intelectual.
 · Dar la aprobación final a la versión a publicar.
 · Acceder a asumir responsabilidades sobre todos los aspectos del artículo y a investigar y resolver cualquier cuestión relacionada con la exactitud y veracidad de cualquier parte del trabajo.

FILIACIONES

- (1) División de Farmacoepidemiología y Farmacovigilancia, Agencia Española de Medicamentos Sanitarios y Productos Sanitarios. Madrid, España.
- (2) Ministerio de Asuntos Exteriores, Unión Europea y Cooperación. Madrid, España.
- (3) Área de Programas de Vacunación, Dirección General de Salud Pública, Ministerio de Sanidad. Madrid, España.
- (4) Departamento de Salud Pública, Universidad Complutense de Madrid. Madrid, España.

RESUMEN

Los estudios de no inferioridad son cada vez más frecuentes para introducir nuevos medicamentos en el mercado. Aunque existen situaciones en las que su uso está justificado, no existe un enfoque analítico único y conservador. Para arrojar resultados fiables y de calidad, deben seguir una estricta metodología, tanto en la ejecución como en la difusión de los resultados, la cual permita, tanto a las agencias reguladoras como a los clínicos, establecer conclusiones válidas y decisiones que repercutan en beneficio de la práctica clínica. La mayor parte de las revisiones publicadas se centran en los ensayos clínicos de no inferioridad de eficacia. En esta revisión se contemplan, además, los diseños observacionales y los aspectos específicos de los estudios de seguridad. Todo ello desde un punto de vista práctico y sencillo, con un contenido matemático mínimo.

PALABRAS CLAVE // Metodología; Revisión; Ensayos clínicos; Estudios observacionales.

ABSTRACT

Non-inferiority studies are increasingly more common for introducing new medicines in the market. Despite being situations where the use of this study design is justified, there is not a common analytical approach on how to conduct them. Pursuing a rigorous methodology, both in the study conduction and in its disseminations, is critical to ensure robust results to enable regulatory agencies and clinicians to reach valid conclusions and decisions which ultimately will benefit clinical practice. Most of the published reviews focus on the efficacy outcomes of non-inferiority clinical trials. We are unaware of other reviews that goes beyond and includes specific aspects for non-interventional designs and for studies focused on safety. Moreover, this review provides a simple and practical perspective with a minimum mathematical content on this complex type of studies.

KEYWORDS // Methodology; Review; Clinical trials; Observational studies.



EL PRINCIPAL OBJETIVO DE LOS ENSAYOS clínicos es demostrar la superioridad de un tratamiento respecto a otra estrategia terapéutica o placebo. Los ensayos frente a placebo permiten estimar el efecto neto del tratamiento experimental, mientras que si se utiliza un control activo se estimará el efecto comparado, el cual dependerá de la magnitud del efecto del control. Ambos comparadores presentan ciertas ventajas y desventajas. El uso de placebo está limitado por los aspectos éticos y prácticos, mientras que el uso de control activo dificulta la demostración de superioridad cuando este es muy bueno, por lo que cuando no se espera que el tratamiento experimental proporcione una ventaja importante de eficacia se recurre a estudios de no inferioridad (NI) (1,2). También aplica en otras circunstancias en las que no se esperan ventajas en la eficacia, pero sí en otras características, como mayor facilidad en la administración o menor frecuencia de efectos adversos (3).

Los estudios de NI se consideran un caso especial de los estudios de equivalencia. Estos últimos pretenden determinar si la diferencia en los efectos entre dos terapias está comprendida entre un pequeño intervalo previamente especificado. La finalidad de los estudios de NI es establecer que el efecto del tratamiento experimental, cuando es comparado con el control activo, no está por debajo de cierto margen de NI (MNI). Este MNI sería la diferencia de eficacia que se acepta entre el tratamiento experimental y el control activo, que no es clínicamente relevante, y mientras no se sobrepase se considera que el nuevo tratamiento es no inferior.

Los estudios de NI con control activo han aumentado considerablemente en la última década, especialmente en aquellas patologías para las cuales se dispone de múltiples alternativas terapéuticas (cardiovasculares, oncológicas, inmunodeficiencia adquirida), que

hacen que los estudios con placebo resulten éticamente inaceptables (4,5).

Recientemente, esta metodología también se ha aplicado en ensayos clínicos para evaluar si un tratamiento efectivo es lo suficientemente seguro. Por ejemplo, para demostrar la NI de los efectos neuropsicológicos de la guanfacina en comparación con la atomoxetina en niños de entre 6 y 17 años con trastorno por déficit de atención e hiperactividad (6).

En estos casos, para establecer el MNI de seguridad es necesario haber demostrado previamente la eficacia de los tratamientos, ya que se ha de poner en contexto con el beneficio del medicamento. Si el nuevo tratamiento ha demostrado mucha eficacia, se permitirá un margen de seguridad más amplio (7).

Los estudios de NI también pueden adoptar un diseño observacional y así estimar el efecto en condiciones reales. La evaluación del riesgo de un medicamento en dichos estudios debe considerar los factores de confusión y otros sesgos de selección y, al igual que los ensayos clínicos, no permiten posicionar al medicamento para la toma de decisiones en la práctica clínica. No obstante, se han utilizado, por ejemplo, para demostrar la NI del rivaroxaban respecto a la warfarina en la prevención del ictus o el embolismo sistémico en pacientes con fibrilación auricular (8) y también para comparar el riesgo de determinados eventos cardíacos y cerebrovasculares en pacientes con EPOC tratados con umeclidinium en comparación con tiotropio (9).

Por tanto, dado el creciente número de estudios publicados en la literatura científica y que cada vez son más utilizados en el ámbito regulatorio para introducir nuevos medicamentos en el mercado, resulta crucial conocer sus entresijos metodológicos e interpretar adecuadamente los resultados que arrojan, para estimar con claridad el beneficio real de la nueva terapia.

Estudios de no inferioridad de eficacia y seguridad: revisión e interpretación de aspectos metodológicos y estadísticos en ensayos clínicos y estudios observacionales con medicamentos.

DIANA GONZÁLEZ BERMEJO *et al.*

LA HIPÓTESIS NULA Y ALTERNATIVA EN ESTUDIOS DE NO INFERIORIDAD

EL PRIMER PASO PARA REALIZAR UN ESTUDIO clínico, tanto observacional como intervencionista, es establecer dos hipótesis opuestas sobre la población, la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

En un estudio de superioridad:

- H_0 mantiene que no existen diferencias de efecto entre los tratamientos a comparar (los tratamientos son iguales, la diferencia es nula).
- H_1 mantiene que existen diferencias.

Con los datos de nuestro estudio y mediante pruebas de significación estadística se obtiene el valor p que indica la probabilidad de que las diferencias hayan ocurrido por azar si la H_0 es cierta. Por consenso se ha establecido un valor umbral de 0,05 para rechazar o aceptar la H_0 . Es decir, si la p es menor de 0,05 existe una probabilidad menor del 5% de que las diferencias observadas sean debidas al azar, si la H_0 fuera cierta. Las diferencias son estadísticamente significativas y, por tanto, se rechaza H_0 .

En los estudios de superioridad, las diferencias se demuestran en los dos sentidos, es decir, que el tratamiento experimental sea mejor que el control activo o viceversa. Por eso se trata de un contraste bilateral o prueba estadística a dos colas (10).

Sin embargo, en los estudios de NI se pretende demostrar que el nuevo tratamiento es al menos tan eficaz o tan seguro como el comparador (en general, la terapia estándar u otro tratamiento de referencia, pero siempre un control activo), por lo que se utilizan con-

trastes unilaterales o pruebas estadísticas a una cola.

Con las pruebas bilaterales se obtienen valores de p superiores que con las pruebas unilaterales (menores posibilidades de alcanzar la significación estadística). Si la p a dos colas es significativa, también lo es la p a una cola (10,11). De la misma forma que en los estudios de superioridad, si en un estudio de NI se obtiene una $p < 0,05$, se rechazará H_0 , pero las hipótesis se definen de forma diferente:

En un estudio de NI de eficacia:

- H_0 establece que la diferencia supera el MNI que hemos establecido, es decir, el nuevo tratamiento no es tan eficaz como el control activo porque la diferencia es mayor de la aceptada.
- H_1 mantiene NI del nuevo tratamiento respecto al control activo.

En un estudio de NI de seguridad:

- H_0 mantiene que el tratamiento experimental presenta mayor frecuencia o mayor riesgo de eventos adversos que el control activo.
- H_1 establece que el nuevo tratamiento presenta menor riesgo que el control, o, en caso de ser superior, la diferencia se sitúa dentro del MNI (10,12).

La **TABLA 1** presenta la forma más habitual de expresar las hipótesis en los estudios de superioridad y NI. Un aspecto importante a considerar es que el no poder demostrar NI, en cuanto a eficacia o seguridad del nuevo tratamiento respecto al control activo, no quiere decir que el control activo sea superior. La superioridad deberá ser demostrada con un estudio diseñado para este fin.



Definición de la hipótesis nula y alternativa en los estudios de superioridad y no inferioridad de eficacia y seguridad.

Estudios	Hipótesis nula (H_0)	Hipótesis alternativa (H_1)
Estudio de superioridad de eficacia	$C-E=0$ $C=E$ La diferencia de eficacia es nula. C y E presentan eficacia similar.	$C-E \neq 0$ $C \neq E$ La diferencia de eficacia no es nula. C y E presentan eficacia diferente.
	$E-C=0$ $E=C$ La diferencia de la frecuencia de eventos adversos es nula. C y E presentan frecuencia de eventos adversos similar.	$E-C \neq 0$ $E \neq C$ La diferencia de la frecuencia de eventos adversos no es nula. C y E presentan frecuencia de eventos adversos diferentes.
Estudio de superioridad de seguridad	$E/C=1$ $E=C$ La relación de la frecuencia de eventos adversos es nula. C y E presentan riesgo de eventos adversos similar.	$E/C \neq 1$ $E \neq C$ La relación de la frecuencia de eventos adversos no es nula. C y E presentan riesgo de eventos adversos diferentes.
	$C-E > MNI$ C presenta mayor eficacia. La diferencia es mayor de lo aceptado. E no es tan eficaz como C.	$C-E < MNI$ La diferencia en la eficacia es menor de lo aceptado. E es al menos tan eficaz.
Estudio de no inferioridad de seguridad	$E-C > MNI$ La frecuencia de eventos adversos de E es mayor de la aceptada.	$E-C < MNI$ E presenta menor frecuencia de eventos adversos que C o en caso de producir más eventos, el aumento se encuentra dentro del MNI.
	$E/C > MNI$ El riesgo de eventos adversos de E es mayor de lo aceptado.	$E/C < MNI$ E presenta menor riesgo de eventos adversos que C o en caso de producir mayor riesgo, el aumento se encuentra dentro del MNI.

Abreviaturas: E: tratamiento experimental; C: control activo; MNI: margen de no-inferioridad.

Estudios de no inferioridad de eficacia y seguridad: revisión e interpretación de aspectos metodológicos y estadísticos en ensayos clínicos y estudios observacionales con medicamentos.

DIANA GONZÁLEZ BERMEJO et al.

INCERTIDUMBRE Y TAMAÑO DE MUESTRA

LA BIOESTADÍSTICA NO ES UNA CIENCIA exacta, se basa en probabilidades, ya que se suele trabajar con muestras procedentes de una población más amplia. Al contrastar las hipótesis y rechazar o aceptar H_0 , existe una probabilidad de equivocarse:

- Error tipo I o riesgo α :

Se conoce como error tipo I o riesgo α la probabilidad de rechazar H_0 cuando es verdadera. Es la probabilidad de cometer falsos positivos (se establece que existen diferencias estadísticamente significativas entre los tratamientos que se comparan, cuando realmente no existen).

En estudios de superioridad se suele asumir un riesgo del 5%.

En estudios de NI se recomienda asumir un riesgo del 2,5% y se interpreta como la probabilidad de concluir falsamente NI del tratamiento experimental cuando en realidad el control activo es muy superior.

- Error tipo II o riesgo β :

Es la probabilidad de aceptar H_0 cuando es falsa, es decir, la probabilidad de cometer falsos negativos. Se recomienda asumir un riesgo β del 10%-20%. En estudios de NI es el riesgo de concluir que el nuevo tratamiento es inferior, cuando realmente es no inferior al control activo.

- Potencia estadística:

El complementario del riesgo β ($1-\beta$) es la potencia estadística, que es la probabilidad de detectar diferencias (en estudios de superioridad) o no inferioridad (en estudios de NI) cuando en realidad existe.

La magnitud del error tipo I y tipo II dispuestos a tolerar va a determinar el tamaño de muestra necesario para obtener resultados precisos y disponer de suficiente potencia estadística (13). Un MNI mínimo, una mayor variabilidad de los datos o una menor probabilidad de error tipo I o II implican un mayor número de pacientes (13,14).

Actualmente existen aplicaciones y programas estadísticos gratuitos en internet que permiten estimar el tamaño de muestra de forma fácil y rápida (15).

MULTIPLICIDAD DE ANÁLISIS Y PENALIZACIÓN ESTADÍSTICA

UN ASPECTO A CONSIDERAR CUANDO SE realizan múltiples análisis de NI sobre la misma muestra (por ejemplo, cuando existen varios objetivos, variables, subgrupos o análisis intermedios) es la penalización estadística, que consiste en controlar el riesgo α , operando cada vez con un margen de error más estrecho para que, de forma global, el estudio con sus múltiples pruebas estadísticas mantenga el margen de error requerido (16,17).

Otras situaciones en las que se deben realizar ajustes en el riesgo α se presentan cuando interesa comprobar superioridad después de haber verificado NI. Demostrar superioridad frente al control activo es más difícil que frente a placebo. En estos casos, se debe especificar en el protocolo antes de llevar a cabo el estudio e ir ajustando el riesgo α , a medida que se van realizando los diferentes análisis con cada una de las variables de interés. Se recomienda siempre comprobar primero NI y, a continuación, superioridad. En caso contrario, se debe especificar el MNI antes de llevar a cabo el estudio; si no, no se seleccionará de forma objetiva, ya que la diferencia de efectos entre el nuevo tratamiento y el control activo ya se conoce (1,18,19).



Los diferentes tipos de análisis y métodos de imputación de valores perdidos pueden resultar a favor o en contra de la H_0 , la cual difiere si se trata de un estudio de NI o superioridad (19). El protocolo del estudio debe diferenciar, de forma exhaustiva, las diferentes asunciones y análisis estadísticos para cada uno de los diseños.

Existen pruebas estadísticas que, además del ajuste del riesgo α , tienen en cuenta otros aspectos como el orden de importancia de las diferentes variables o el número de hipótesis nulas verdaderas que son rechazadas y así demostrar la NI de todas las variables en su conjunto (18-21).

ELECCIÓN DEL MARGEN DE NO INFERIORIDAD

LA ELECCIÓN DEL MNI SE PUEDE BASAR EN juicios clínicos y experiencias previas preguntando a médicos y pacientes el grado de eficacia o seguridad que estarían dispuestos a sacrificar por el beneficio del nuevo tratamiento.

Otra opción es estimarlo mediante métodos matemáticos. En este caso, es necesario cono-

cer la diferencia del efecto del control activo frente a placebo observada en ensayos clínicos previos y partir de este valor, estimando la máxima pérdida de eficacia o seguridad del nuevo tratamiento respecto al control activo clínicamente aceptable.

No obstante, se debe tener en cuenta la escala de medida de las variables: la magnitud de los resultados difiere si se trata de una diferencia de proporciones o de una relación.

Cuando se demuestra NI del nuevo tratamiento frente al control activo, indirectamente se está demostrando superioridad frente a placebo.

Para algunos medicamentos como trombolíticos, antirretrovirales y antiinfecciosos, existen valores estándar de MNI para estudios de eficacia.

En estudios de seguridad depende del beneficio del nuevo tratamiento, aunque es difícil aceptar un aumento en el umbral del MNI cuando se trata de mortalidad o morbilidad irreversible (22,23). La **Tabla 2** presenta un ejemplo de la importancia de elegir un adecuado MNI.

Tabla 2
Ejemplo de la importancia de elegir un adecuado margen de no inferioridad.

Caso	Control activo (C) n=150	Tratamiento experimental (E) n=150	Diferencia de efectos (C-E)	Intervalo de confianza (IC) al 95% de la diferencia de efectos	Margen de no inferioridad (MNI)
1	121 (81%)	112 (74%)	7%	(-2%-16%)	15%
2	118 (79%)	103 (69%)	10%	(1%-19%)	20%

En el caso 1, el límite superior del IC 95% de la diferencia de efectos supera el MNI, no se demuestra la no inferioridad. En el caso 2, se ha utilizado un MNI más amplio, el límite superior del IC 95% de la diferencia se encuentra dentro del margen establecido, se demuestra la no inferioridad, a pesar de que la diferencia de efectos es mayor en el caso 2.

Estudios de no inferioridad de eficacia y seguridad: revisión e interpretación de aspectos metodológicos y estadísticos en ensayos clínicos y estudios observacionales con medicamentos.

DIANA GONZÁLEZ BERMEJO et al.

EVALUACIÓN DEL EFECTO DEL CONTROL ACTIVO

ES FUNDAMENTAL QUE EL EFECTO HISTÓRICO del control activo tenga la suficiente constancia en los estudios previos disponibles, para constituir una base sólida sobre la que elegir el MNI. Puede ocurrir que la variable de interés en el estudio de NI no sea la variable primaria en los estudios previos. Los ensayos clínicos suelen estar diseñados para estudiar la eficacia, y los aspectos de seguridad son evaluados de forma exploratoria con falta de sensibilidad para detectar diferencias. Por esta razón, el ensayo clínico del que surgió una señal de un evento adverso no debe utilizarse para establecer el MNI.

Las variables compuestas en estudios de seguridad pueden enmascarar el efecto real de los componentes individuales y la variabilidad entre subgrupos. En estos casos, se recomienda realizar un análisis de sensibilidad para evaluar la contribución de los componentes a nivel individual y seleccionar diferentes MNI. Si se utilizan variables subrogadas, estas deben estar relacionadas directamente con el mecanismo que induce el efecto. La restricción del tamaño de muestra o la realización de análisis intermedios, cuando se cruza un umbral de eficacia o seguridad en ensayos clínicos adaptativos, puede sobrevalorar el efecto cuando el tratamiento es inefectivo. La selección de estudios previos si únicamente han mostrado diferencias estadísticamente significativas también puede constituir un sesgo en la elección del MNI (7,24).

ENSAYOS CLÍNICOS DE NO INFERIORIDAD DE TRES BRAZOS

EN OCASIONES PUEDE EXISTIR DEMASIADA heterogeneidad y falta de consistencia del efecto del control activo en los ensayos clínicos previos, o incluso puede que no existan ensayos clínicos en la indicación de interés, pero se sabe que es efectivo. En estos casos,

está justificada la realización de ensayos clínicos de NI con tres brazos. El objetivo es determinar la sensibilidad del ensayo para detectar la diferencia que esperamos encontrar y establecer la proporción del efecto de control activo que es retenida por el nuevo tratamiento. Si ninguno muestra efecto superior frente al placebo, el estudio no es sensible para detectar las diferencias por defectos en la metodología o en el diseño. En general, el control activo es el tratamiento de referencia actual, por lo que ha debido demostrar efecto sobre placebo en algún ensayo anterior.

Puede ocurrir que el nuevo tratamiento muestre efecto sobre placebo, pero que resulte insuficiente si lo comparamos con el efecto del control activo sobre placebo. Por esta razón se estima la fracción de retención (λ) como la relación de la diferencia del efecto del tratamiento experimental sobre placebo entre la diferencia del efecto del control activo sobre placebo ($\lambda = E-P/C-P$).

La principal limitación de este tipo de diseños es la obtención del consentimiento informado de los sujetos, ya que deben estar dispuestos a participar sabiendo que pueden recibir placebo, cuando en realidad existen dos tratamientos activos (25-27). En el estudio *PRECISION*, que evaluó la NI de celecoxib respecto a naproxeno en el desarrollo de eventos cardiovasculares, el tercer grupo no recibió placebo, sino ibuprofeno (28).

ANÁLISIS DE LOS RESULTADOS

EN UN ENSAYO CLÍNICO, EL ANÁLISIS IDEAL de los resultados comprendería todos los sujetos aleatorizados que han seguido todos los procedimientos del protocolo. En la realidad, pueden existir pérdidas de seguimiento por diferentes motivos.

Existen dos tipos de análisis de los datos:

- Análisis por intención de tratar

El análisis por intención de tratar (ITT) incluye a todos los pacientes que han sido inicialmente asignados a cada grupo de tratamiento, independientemente del tratamiento que en realidad reciban posteriormente o si se desviaron del protocolo. Se trata de que las condiciones del ensayo clínico se aproximen a las de la práctica diaria, en las que también existen pacientes que no cumplen el tratamiento.

Implica asignar a todos los pacientes perdidos del grupo experimental la peor de las respuestas, y a los del grupo control, la mejor. Puesto que aumenta la probabilidad de concluir erróneamente que no existen diferencias entre los tratamientos, no es la estrategia más recomendable en estudios de NI de eficacia.

En estudios de NI de seguridad no es apropiado cuando algunos sujetos reciben otro tratamiento para el que no fueron aleatorizados.

- Análisis por protocolo

En los análisis por protocolo (PP) se incluyen los participantes que han completado el tratamiento y que han seguido exactamente las instrucciones del protocolo del estudio. El análisis PP es más conservador en estudios de NI de eficacia que el ITT, pero resulta inapropiado para estudios de NI de seguridad, al excluir a los pacientes que no han finalizado el seguimiento o se han retirado por problemas de seguridad. Lo más apropiado sería incluir todos los sujetos tratados y analizarlos según el tratamiento que han recibido e incluir un período de seguimiento una vez finalizada la exposición, para evaluar eventos adversos de inicio retardado (19,29).

La interpretación de los análisis ITT y PP en estudios observacionales es similar. La única diferencia es que la asignación de los tratamientos se realiza según la práctica clínica

habitual e independientemente de la participación del paciente en el estudio.

INTERPRETACIÓN DE LOS RESULTADOS CON INTERVALOS DE CONFIANZA

ACTUALMENTE SE RECONOCE QUE LOS intervalos de confianza aventajan a las pruebas de hipótesis como instrumento analítico para muchos tipos de investigación. Un valor p no refleja la magnitud de la diferencia, sino la probabilidad de haber observado un efecto si en realidad no existe, sobre la base de un resultado estadísticamente significativo. Cuando una muestra es suficientemente grande, cualquier resultado puede cobrar significación estadística, razón de por sí suficiente para dudar del valor de estas pruebas. Además, la arbitrariedad de los valores α y la variabilidad de los resultados si el estudio se repitiese con distintas muestras, son aspectos que reducen aún más la fiabilidad del valor p , a favor de los intervalos de confianza (30,31).

En las FIGURAS 1, 2 y 3 se presentan varios ejemplos de posibles resultados obtenidos en un estudio de NI y su interpretación según el tipo de variable (eficacia o seguridad), la escala de medida (diferencia o relación de riesgo) y la magnitud del MNI.

LIMITACIONES DE ESTUDIOS DE NI

LA PRINCIPAL LIMITACIÓN DE LOS ESTUDIOS de NI es el *biocreep*, efecto de arrastre o fenómeno de deriva, por el cual el efecto del producto investigado puede decaer a medida que se compara con controles cada vez menos eficaces o seguros (32). En la TABLA 3 se presenta un ejemplo del efecto del *biocreep* sobre los resultados y conclusiones en un estudio de NI. Este efecto es particularmente importante cuando los estudios de NI son utilizados para la introducción en el mercado de los fármacos *me too*, aquellos que presentan pequeñas modificaciones respecto al innovador, con

Estudios de no inferioridad de eficacia y seguridad: revisión e interpretación de aspectos metodológicos y estadísticos en ensayos clínicos y estudios observacionales con medicamentos.

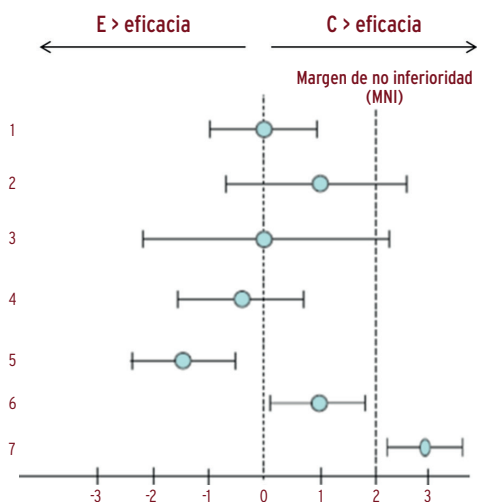
DIANA GONZÁLEZ BERMEJO et al.

Tabla 3
Ejemplo del efecto de *biocreep* sobre los resultados y conclusiones en un estudio de no inferioridad.

Tipo de comparación	Control activo (C) n=150	Tratamiento experimental (E) n=150	Diferencia de efectos (%)	Intervalo de confianza (IC) al 95% de la diferencia de efectos	Margen de no inferioridad (MNI)
1 (SUP)	Placebo	A 118 (79%)	-	-	-
2 (NI)	A 118 (79%)	B 103 (69%)	10%	0,9%-19,1%	20%
3 (NI)	B 103 (69%)	C 90 (60%)	9%	2,1%-19,5%	20%
4 (NI)	C 90 (60%)	D 78 (52%)	12%	3,2%-19,2%	20%
5 (NI)	D 78 (52%)	E 66 (44%)	14%	3,3%-19,3%	20%

Si en un primer estudio se demuestra la superioridad de A sobre placebo, el tratamiento A puede transformarse en estándar y ser utilizado como control activo en un segundo estudio de no inferioridad comparando B frente a A. Si se establece la no inferioridad de B respecto a A (control activo) y por alguna razón A deja de ser el tratamiento de referencia, por aparición de efectos adversos mortales o disminución de la eficacia con el tiempo (antibacterianos), entonces B podría ser el nuevo tratamiento disponible y por tanto el potencial control activo. En un tercer estudio C demuestra ser no inferior a B, pero su real efecto podría ser menor que el del placebo utilizado en el estudio original de superioridad de A frente a placebo.

Figura 1
Interpretación de los posibles resultados con intervalos de confianza de la diferencia de eficacia entre el control activo (C) y el tratamiento experimental (E) en estudios de no inferioridad.



1. $C-E=0$: sugiere igualdad de eficacia, el límite superior del intervalo de confianza no supera el MNI, por lo que se demuestra no inferioridad.
2. $C-E>0$: No se demuestra la no inferioridad. El límite superior del intervalo de confianza supera el MNI. No implica que C sea superior.
3. $C-E=0$: sugiere igualdad de eficacia, pero el límite superior del intervalo de confianza supera el MNI, por lo que no se demuestra la no inferioridad.
4. $C-E<0$: Se demuestra no inferioridad, pero no se podría reclamar la superioridad de E. El intervalo de confianza incluye 0.
5. $C-E<0$: Se demuestra no inferioridad, el intervalo de confianza no incluye 0. Se podría reclamar la superioridad de E con un diseño adecuado.
6. $C-E>0$: Podría concluirse no inferioridad, todos los valores del intervalo de confianza se encuentran por debajo del MNI. Sin embargo, E es inferior a C de forma significativa. Resultados muy improbables, porque requieren un gran tamaño de muestra y un MNI muy ancho.
7. E es inferior a C en cuanto a eficacia.

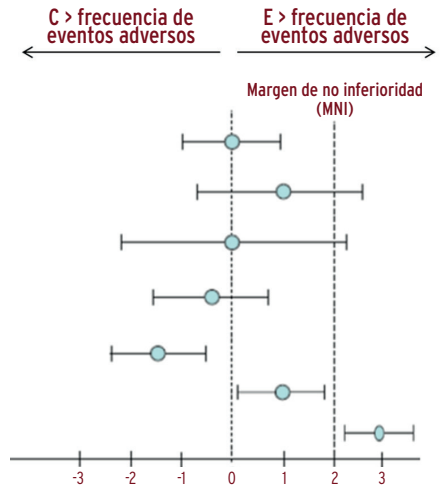
Estudios de no inferioridad de eficacia y seguridad: revisión e interpretación de aspectos metodológicos y estadísticos en ensayos clínicos y estudios observacionales con medicamentos.

DIANA GONZÁLEZ BERMEJO et al.

Rev Esp Salud Pública
Volumen 96
29/9/2022
e202209074

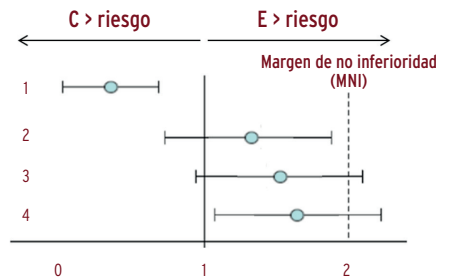
Interpretación de los posibles resultados con intervalos de confianza de la diferencia entre la frecuencia de eventos adversos del tratamiento experimental (E) y el control activo (C) en estudios de no inferioridad.

1. $E-C=0$: Se demuestra no inferioridad de E (E es tan seguro como C, no produce mayor frecuencia de eventos adversos, o de producir un aumento, se encuentra dentro del MNI).
2. $E-C>0$: No se demuestra no inferioridad (que E sea tan seguro como C). El intervalo de confianza de la diferencia E-C supera el MNI.
3. $E-C=0$: Sugiere igualdad de efectos, pero el límite superior del intervalo de confianza supera el MNI, por lo que no se demuestra la no inferioridad.
4. $E-C<0$: Se demuestra no inferioridad de E. (E es tan seguro como C, no produce mayor frecuencia de eventos adversos, o de producir un aumento se encuentra dentro del MNI).
5. $E-C<0$: Se demuestra no inferioridad. El intervalo de confianza no incluye 0, podría reclamarse superioridad en cuanto a seguridad de E (menor frecuencia de eventos adversos) con un diseño adecuado.
6. $E-C>0$: Podría concluirse no inferioridad, todos los valores del intervalo de confianza se encuentran dentro del margen establecido. Sin embargo, todos los valores del intervalo de confianza están a favor de C, E produce mayor frecuencia de eventos adversos de forma significativa. Resultados improbables, requieren un gran tamaño de muestra y un MNI muy ancho.
7. $E-C>0$: E presenta mayor frecuencia de eventos adversos, es menos seguro.



Interpretación de los posibles resultados con intervalos de confianza de la relación de riesgos (RR) del tratamiento experimental (E) respecto al control activo (C) en estudios de no inferioridad.

1. $E/C<1$: Se demuestra no inferioridad (E es tan seguro como C, no produce mayor riesgo de eventos adversos). El intervalo de confianza de la RR no incluye 1. Puede reclamarse superioridad en cuanto a seguridad de E (menor riesgo de eventos adversos) con un diseño adecuado.
2. $E/C>1$, el intervalo de confianza incluye 1 y el límite superior no supera el MNI. Se demuestra no inferioridad (E es tan seguro como C, no produce mayor riesgo de eventos adversos, o de producir un aumento, se encuentra dentro del MNI).
3. $E/C>1$, el intervalo de confianza incluye 1, pero el límite superior supera el MNI. No se demuestra no inferioridad (E es menos seguro).
4. $E/C>1$, el intervalo de confianza no incluye el 1 y el límite superior supera el MNI. No se demuestra la no inferioridad (E es menos seguro, el riesgo de eventos adversos es mayor que C).




Estudios de no inferioridad de eficacia y seguridad: revisión e interpretación de aspectos metodológicos y estadísticos en ensayos clínicos y estudios observacionales con medicamentos.

DIANA GONZÁLEZ BERMEJO et al.

alguna ventaja en la farmacocinética o en la administración (31).

En resumen, los estudios de NI son cada vez más frecuentes en la literatura científica y cada vez son más utilizados en el ámbito regulatorio para introducir nuevos medicamentos en el mercado. Por ello, es necesario disponer de una base sólida para la interpretación de los aspectos éticos y metodológicos particulares de este tipo de análisis. Aunque existen situaciones en las que su uso está justificado, no existe un enfoque analítico único y conservador y la elección arbitraria del MNI requiere de ardua justificación. Para arrojar resultados fiables y de calidad, deben seguir una estricta metodología, tanto en la ejecución como en la difusión de los resultados, que permita tanto a las agencias reguladoras como a los clínicos establecer conclusiones válidas y decisiones que repercutan en beneficio de la práctica clínica.

Es necesario recordar que un resultado negativo en un estudio de superioridad no implica la NI y debe prestarse especial atención a los MNI que suelen ser demasiado amplios para obtener diferencias de efectos clínicamente significativas. Además, resulta crucial que estos estudios incluyan los requisitos básicos de información, incluida la definición y justificación del MNI, el cálculo del tamaño muestral, la presentación de ambos tipos de análisis (por intención de tratar y por protocolo), así como los intervalos de confianza de los resultados. Sin esta información, se debe poner en cuestión el rigor científico de los datos presentados.

No obstante, esta revisión pretende proporcionar una visión global de los aspectos metodológicos y estadísticos de los estudios de NI de forma rápida y sencilla y con un contenido matemático mínimo. Para un mayor aprendizaje y un conocimiento más exhaustivo, se recomienda consultar la bibliografía referenciada. 

BIBLIOGRAFÍA



1. European Medicines Agency (EMA). *Points to consider on switching between superiority and non-inferiority*. Br J Clin Pharmacol. 2001; 52: 223-228.
2. Catalá-López F, González-Bermejo D, De la Fuente Honrubia C, Macías Saint-Gerons D. *Clinical equipoise and systematic reviews of randomized controlled trials*. Med Clin (Barc). 2015; 145 (11): 496-498.
3. Ebutt AF, Frith L. *Practical issues in equivalence trials*. Stat. Med. 1998; 17: 1691-1701.
4. Mauri L, D'Agostino RB Sr. *Challenges in the Design and Interpretation of Noninferiority Trials*. N Engl J Med. 2017; 377 (14): 1357-1367.
5. European Medicines Agency (EMA). *Reflection paper on the need for active control in therapeutic areas where use of placebo is deemed ethical and one or more established medicines available*. Disponible en <https://www.ema.europa.eu/en/documents/scientific-guide-line/reflection-paper-need-active-control-therapeutic-areas-where-use-placebo-deemed-ethical-one-more-en.pdf> [Citado septiembre de 2022].
6. Takeda (Shire). *A Phase 4, Multicenter, 2-part Study Composed of a 1-Year Randomized, Double-blind, Parallel-group, Placebo-controlled, Active-comparator, Dose-optimization Evaluation Followed by a 1-Year Open-label Evaluation to Assess the Safety and Efficacy of Guanfacine Hydrochloride Prolonged-release (SPD503) in Children and Adolescents Aged 6 to 17 Years With Attention-deficit/Hyperactivity Disorder*. Disponible en <https://clinicaltrials.gov/ct2/show/NCT04085172> [Citado septiembre de 2022].
7. Rothmann M, Lin N, Chen G, Chi GY, Temple R, Tsou HH. *Design and analysis of non-inferiority mortality trials in oncology*. Stat Med. 2003; 22 (2): 239-264.
8. Althunian TA, de Boer A, Groenwold RHH, Rengering KO, Souverein PC, Klungel OH. *Rivaroxaban was found to be noninferior to warfarin in routine clinical care: A retrospective noninferiority cohort replication*

9. GlaxoSmithKline. *Post-authorisation Safety (PAS) Observational Cohort Study to Quantify the Incidence and Comparative Safety of Selected Cardiovascular and Cerebrovascular Events in COPD Patients Using Inhaled UMEC/VI Combination or Inhaled UMEC versus Tiotropium (Study 201038)*. Disponible en <https://www.encepp.eu/encepp/viewResource.htm?id=45347> [Citado septiembre de 2022].

10. Scott IA. *Non-inferiority trials: determining whether alternative treatments are good enough*. *MJA.* 2009; 190 (6): 326-330.

11. Bland JM, Altman DG. *One or two sided tests of significant*. *BMJ.* 1994; 309 (6949): 248.

12. Department of Health and Human Services. Food and Drug Administration. *Non-inferiority clinical trials to establish effectiveness. Guidance for industry*. 2016. Disponible en: <https://www.fda.gov/media/78504/download> [Citado septiembre de 2022].

13. Estrada Pérez C, Jaimes Barragán F. *Ronda clínica y epidemiológica. Estudios de superioridad vs Estudios de no inferioridad*. *IATREIA.* 2014; 27 (1): 117-122.

14. Head SJ, Kaul S, Bogers AJC, Kappetein AP. *Non-inferiority study design: lessons to be learned from cardiovascular trials*. *Eur Hear. J.* 2012; 33 (11):1318-1324.

15. Negida A. *Sample Size Calculation Guide - Part 6: How to calculate the sample size for a non-inferiority or an equivalence clinical trial*. *Front Emerg Med.* 2019; 4(1): e15.

16. Molina Arias M. *El problema de las comparaciones múltiples*. *Rev Pediatr Aten Primaria.* 2014; 16 (64): 367-370.

17. McLaughlin MJ, Sainani KL. *Bonferroni, Holm and Hochberg corrections: fun names, serious changes to p values*. *PM R.* 2015; 6: 544-546.

18. Dmitrienko A, Wiens BL, Tamhane AC, Wang X. *Tree-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives*. *Stat*

Med. 2007; 26 (12): 2465-2478. doi: <http://www.doi.org/10.1002/sim.2716>. Erratum in: *Stat Med.* 2008; 27 (17): 3452.

19. Brittain E, Lin D. *A comparison of intent-to-treat and per protocol results in antibiotic non-inferiority trials*. *Stat Med.* 2005; 24: 1-10.

20. Westfall PH, Krishen A. *Optimally weighted, fixed sequence and gatekeeper multiple testing procedures*. *J Stat Plan Infer.* 2001; 99: 25-40.

21. Dmitrienko A, Offen WW, Westfall PH. *Gatekeeping strategies for clinical trials that do not require all primary endpoints to be significant*. *Stat Med.* 2003; 22: 2400-2403.

22. European Medicines Agency (EMA). *Guideline on the choice on the non-inferiority margin*. Disponible en: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-choice-non-inferiority-margin_en.pdf [Citado septiembre de 2022].

23. Treadwell J, Uhl S, Tipton K, Singh S, Santaguida L, Sun X *et al*. *Assessing equivalence and noninferiority*. 2012. Disponible en: https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/equivalence_research.pdf [Citado mayo de 2022].

24. D'Agostino RB, Massaro JM, Sullivan LM. *Non-inferiority trials: Design concepts and issues. The encounters of academic consultants in statistics*. *Stat. Med.* 2003; 22: 169-186.

25. Ochiai T, Hamasaki T, Evans SR, Asakura K, Ohno Y. *Group sequential three-arm noninferiority clinical trial designs*. *J Biopharm Stat.* 2017; 27(1): 1-24.

26. Kieser M, Friede T. *Planning and analysis of three-arm non-inferiority trials with binary endpoints*. *Stat Med.* 2007; 26 (2): 253-273.

27. Tang ML, Tang NS. *Tests of non-inferiority via rate difference for three arm clinical trials with placebo*. *J Biopharm Stat.* 2004; 14: 337-347.

28. Nissen SE, Yeomans ND, Solomon DH, Lüscher TF, Libby P, Husni ME *et al*. *PRECISION Trial Investi-*

Estudios de no inferioridad de eficacia y seguridad: revisión e interpretación de aspectos metodológicos y estadísticos en ensayos clínicos y estudios observacionales con medicamentos.

DIANA GONZÁLEZ BERMEJO *et al.*

gators. *Cardiovascular Safety of Celecoxib, Naproxen, or Ibuprofen for Arthritis*. N Engl J Med. 2016;375(26): 2519-2529.

29. Loli P. *Estudios de no-inferioridad en Neurología: tratamientos nuevos, ¿eficacia demostrada?* Neurol. Arg. 2010; 2 (1): 42-49.

30. Cumming G. *Replication and p intervals: p values predict the future only vaguely but confidence intervals do much better*. Perspect Psychol Sci. 2008; 3 (4): 286-300.

31. Poole C. *Low P-Values or Narrow Confidence Intervals: Which Are More Durable?* Epidemiology. 2001; 12 (3): 291-294.

32. Everson-Stewart S, Emerson SS. *Bio-creep in non-inferiority clinical trials*. Stat Med. 2010; 29 (27): 2769-2780.

DE
SD

Estudios de no inferioridad de eficacia y seguridad: revisión e interpretación de aspectos metodológicos y estadísticos en ensayos clínicos y estudios observacionales con medicamentos.

DIANA
GONZÁLEZ
BERMEJO
et al.

Rev Esp Salud Pública
Volumen 96
29/9/2022
e202209074

13