

# ¿Ensayos clínicos de superioridad, equivalencia o no inferioridad?

## ¿Clinical trials of superiority, non-inferiority or equivalence?

Carlos Manterola<sup>1,2,3</sup>, María José Hernández-Leal<sup>2,3</sup>, Tamara Otzen<sup>2,3</sup>, Juan Pablo Holguín<sup>2,3,4</sup>, Carla Salgado<sup>1,2,3</sup> y Luis Grande<sup>5,6</sup>

<sup>1</sup>Centro de Estudios Morfológicos y Quirúrgicos (CEMyQ), Universidad de La Frontera, Chile.

<sup>2</sup>Programa de Doctorado en Ciencias Médicas, Universidad de La Frontera, Chile.

<sup>3</sup>Núcleo Milenio de Sociomedicina.

<sup>4</sup>Universidad del Azuay, Ecuador.

<sup>5</sup>Universitat Autònoma de Barcelona, España.

<sup>6</sup>Servicio de Cirugía, Hospital del Mar, Barcelona, España.

Financiamiento: Parcialmente financiado por proyecto MEC 80190026.

Conflictos de interés: Ninguno.

Recibido: 9 de octubre de 2022 (segunda versión: 29 de marzo de 2023); Aceptado: 24 de noviembre de 2023

### Resumen

La eficacia de una nueva intervención se establece generalmente a través de ensayos clínicos (EC) con asignación aleatoria (AA). Sin embargo, entre otros tantos desafíos metodológicos, el especificar la hipótesis de un EC con AA, sigue siendo un problema complejo de resolver para los investigadores clínicos. En este manuscrito discutimos las características de tres variantes de los EC con AA: EC de superioridad (ECS), EC de no-inferioridad (ECNI), y EC de equivalencia (ECE). Estos tres tipos de EC tienen supuestos diferentes sobre los efectos de una intervención, por lo que plantear hipótesis y definir objetivos requiere conocer algunos supuestos subyacentes a estos EC, incluso hasta elementos relacionados con la estimación del tamaño de muestra para cada cual. El objetivo de este manuscrito fue describir las diferencias metodológicas entre ECS, ECNI y ECE.

**Palabras clave:** Ensayo clínico; ensayo clínico controlado; ensayo controlado aleatorio; ensayos de equivalencia; ensayos de no inferioridad; ensayos de superioridad; equivalencia terapéutica; metodología de investigación; diseños de investigación.

### Abstract

Efficacy and effectiveness of new interventions are generally established through randomized clinical trials (RCTs). However, among many other methodological challenges, specifying the hypothesis of a RCT remains complex problem for clinical researchers. In this manuscript we discuss the characteristics of three variants of RCTs: superiority RCT (SRCT), non-inferiority RCT (NIRCT), and equivalence RCT (ERCT). These three types of RCT have different assumptions about the effects of an intervention, so setting hypotheses and defining objectives requires knowing some assumptions underlying these RCTs, including elements related to the estimation of the sample size for each one. The aim of this manuscript was to describe methodological differences between SRCT, NIRCT and ERCT.

**Keywords:** “Clinical Trials as Topic”[Mesh]; “Controlled Clinical Trial” [Publication Type]; “Randomized Controlled Trials as Topic”[Mesh]; “Equivalence Trials as Topic”[Mesh]; Therapeutic equivalence; Non-inferiority trials; Superiority trials; Research Methodology; Research design.

### Introducción

Los ensayos clínicos (EC) con asignación aleatoria (ECAA), son el estándar de referencia para establecer eficacia y efectividad de una intervención de reciente desarrollo<sup>1,2</sup>; y conocer el objetivo de un ECAA, no sólo es relevante para comprender sus conclusiones, sino que porque a partir de este, surgen tres variantes de este diseño:

EC de superioridad (ECS), de equivalencia (ECE); y de no inferioridad (ECNI). Y, dado que estos estudios tienen objetivos diferentes, varían en algunos aspectos metodológicos<sup>2</sup>.

Un EC bien diseñado debe especificar el tipo de hipótesis que se probará y los procedimientos que se utilizarán en el análisis de los resultados primarios. Así, según el propósito del EC, se debe especificar si se pretende probar superioridad (A es más eficaz o más efectivo que

#### Correspondencia a:

Tamara Otzen Hernández  
tamara.otzen@ufrontera.cl

Carlos Manterola Delgado  
carlos.manterola@ufrontera.cl

B según sea el caso; ya sea porque reduce el impacto de la enfermedad, es más seguro, más conveniente, o más barato), que es la variedad con la que la mayor parte de los médicos clínicos están familiarizados<sup>3,4</sup>; la no inferioridad (el efecto de A no es menos eficaz o efectivo que el efecto de B)<sup>4,5</sup>; o la equivalencia (no demostrar superioridad de una intervención sobre otra, es decir, dos intervenciones -X y Z- no tienen una diferencia clínicamente significativa entre ellas)<sup>5,6</sup> (Figura 1).

De este modo, tanto en los ECNI como en los ECE, un resultado, verificado en los sujetos del brazo experimental, es comparado con el resultado observado en los sujetos del brazo comparador; por lo que, si ya existe una intervención bien establecida para una determinada condición clínica, no sería ético probar la eficacia (menos la efectividad), de la nueva intervención frente a un placebo, porque expondría a los sujetos asignados al brazo control a un riesgo inadmisible<sup>3</sup>.

Entonces, la decisión de optar por ECS, ECE o ECNI; depende del objetivo, los costes y recursos disponibles relacionados con tamaños de muestra de mayor o menor envergadura, así como por razones éticas. Por otra parte, en los ECNI cuando se logra demostrar no-inferioridad, se puede determinar la superioridad, lo que no se contempla en ECE<sup>3</sup>.

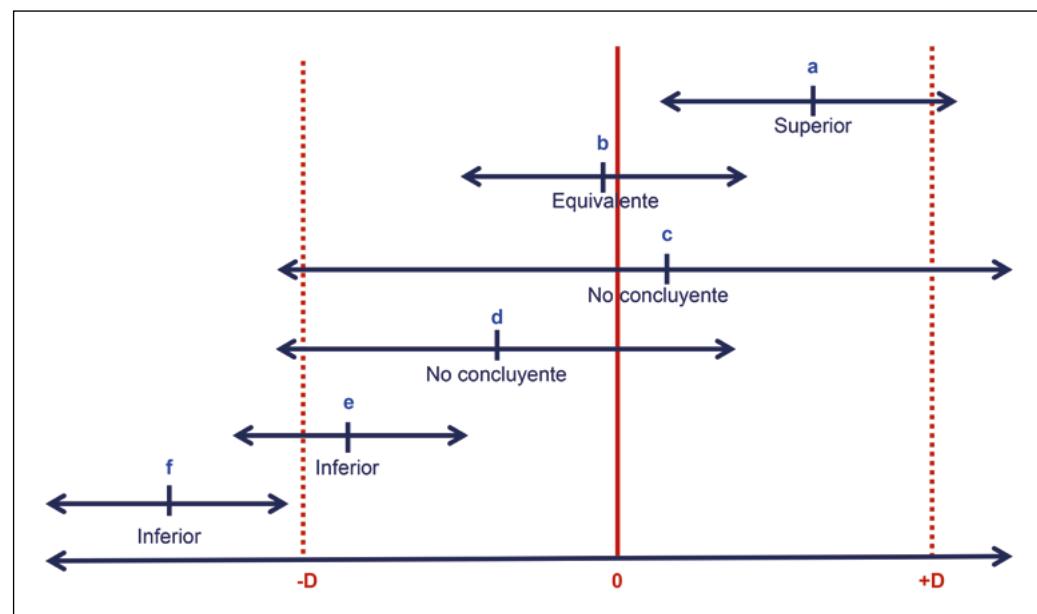
Por esto, se debe considerar que un resultado negativo de un ECS no permite probar que las intervenciones bajo investigación son equivalentes. Por su parte, los ECNI y los ECE, demandan altos estándares para proporcionar resultados válidos y confiables; por lo que se debe tener en consideración si los márgenes de equivalencia tienden a ser demasiado amplios para ser clínicamente significativos;

y que la afirmación de equivalencia puede ser falaz si el estudio no se ha ejecutado de forma rigurosa<sup>1,7</sup>. Por ello, se ha de exigir de todo EC, que reporte características básicas del diseño, como definición y justificación del margen de equivalencia, estimación del tamaño de la muestra considerando este margen, presentación de análisis por intención de tratar y por protocolo (ambos); y proporcionar los IC 95% para los resultados en evaluación<sup>8</sup>.

El objetivo de este manuscrito fue describir las diferencias metodológicas entre ECS, ECNI y ECE.

### **Ensayo Clínico de Superioridad**

Cuando el objetivo del ECAA es demostrar que una intervención es superior a otra, se emplea una estrategia que se denomina ECS; y habitualmente una prueba de superioridad no significativa, se interpreta erróneamente como que no hay diferencia entre las dos intervenciones. Probar que dos tratamientos tienen el mismo rendimiento es imposible con herramientas estadísticas; a lo sumo, se puede demostrar que son equivalentes<sup>3</sup>. Por ende, un ECS se diseña para determinar si la nueva intervención es mejor que la estándar o el placebo. Por ello, se debe rechazar la hipótesis que no hay diferencias estadísticamente significativas entre las intervenciones en evaluación; y así, mientras menor sea el valor de  $p$ , mayor probabilidad que esta sea por efecto de la nueva intervención y no por el azar. Posteriormente, se ha de verificar si la diferencia existente es clínicamente relevante (independiente del valor de  $p$ ); es decir, si el efecto (diferencia entre las intervenciones estudiadas), es igual o superior a un valor predeterminado como clínicamente relevante (que no necesariamente tiene relación con una diferencia estadís-



**Figura 1** Diferencia D encontrada entre intervenciones. Ejemplos de diferencias de intervenciones observadas (intervención control de nueva intervención), con IC 95%. Adaptada de Christensen<sup>1</sup> y de Lesaffre<sup>3</sup>. **a)** La nueva intervención es clínicamente mejor que el control. Sin embargo, la magnitud del efecto puede carecer de importancia clínica; **b)** Se puede considerar que las intervenciones tienen efectos equivalentes; **c) y d)** Resultado no concluyente; **e)** La nueva intervención es clínicamente peor que el control, pero la magnitud de la diferencia puede carecer de importancia clínica; **f)** La nueva intervención es clínicamente peor que el control.

ticamente significativa), con una precisión determinada por el IC 95% que no incluya el 0, porque en un ECS se pretende demostrar que la nueva intervención es superior que el estándar o un placebo<sup>2,7,9</sup>. Esto representa un desafío al momento de plantear  $H_0$  y  $H_1$ , entendiendo que en un ECS  $H_0: \mu_1 - \mu_0 \leq \delta$  vs.  $H_1: \mu_1 - \mu_0 > \delta$ , donde  $\delta \geq 0$ . De tal modo que mientras mayor sea  $\delta$ , más difícil será rechazar  $H_0$ , por lo que este margen es el umbral para plantear la superioridad de la nueva intervención<sup>7,10</sup>.

Por otro lado, es fundamental estimar un tamaño de muestra; y es común que la mayor parte de los estudios consideren igualdad numérica entre los grupos. Sin embargo, en ocasiones es posible que necesitemos asignar más sujetos a uno de ellos. Es así como en un estudio con una intervención y un placebo, puede ser deseable una muestra de mayor tamaño en el grupo placebo ( $n_1$ ), de modo de tener mayor poder de comparación con el grupo de intervención ( $n_2$ ). El valor de  $k$  va a estar representado por la razón de sujetos entre el grupo tratamiento y grupo control, por ejemplo, si hay 1 sujeto en el grupo de intervención por cada 2 sujetos en el grupo placebo, la razón será 1:2, por consiguiente el valor de  $k$  será 0,5<sup>3,12</sup>. Posterior a tener claridad sobre este concepto se debe emplear la fórmula según la variable resultado sea continua o dicotómica (Figura 2). De igual forma, tras el cálculo del tamaño de la muestra para ambos brazos del estudio, es necesario ajustarlo por las posibles pérdidas en el seguimiento. Por lo cual, se aplica la fórmula:  $N = n(1/(1-R))$ ; donde:  $N$  = número de participantes ajustados a las pérdidas;  $n$  = número de participantes obtenidos aplicando las fórmulas antes mencionadas; y  $R$  = proporción de pérdidas esperadas durante el seguimiento.

Ejemplo: “Combining baloxavir marboxil with standard-of-care neuraminidase inhibitor in patients hospitalised with severe influenza (FLAGSTONE): a randomised, parallel-group, double-blind, placebo-controlled, superiority trial”. Con el objetivo de probar si la combinación de baloxavir con inhibidores de la neuramidasa (NAI), tendría mejores resultados clínicos en comparación con la monoterapia con NAI en pacientes hospitalizados con influenza grave; se ejecutó un ECS controlado, aleatorizado, multicéntrico, de grupos paralelos, y con doble enmascaramiento. Se incluyeron pacientes de 12 o más años hospitalizados con influenza confirmada por RT-PCR o prueba rápida, y puntaje nacional de alerta temprana 2 (NEWS2) de 4 o más. Se reclutaron pacientes desde 124 centros de 25 países, con muestreo de bloques permutados, asignados aleatoriamente en una proporción 2:1 para recibir baloxavir+NAI (grupo baloxir), o placebo+NAI (grupo control). Baloxavir se administró vía oral los días 1 y 4 (40 mg para peso corporal < 80 kg, u 80 mg para ≥ 80 kg), y el día 7 si no se había producido mejoría clínica al día 5. Los NAI incluidos fueron oseltamivir, zanamivir y peramivir;

seleccionados y administrados según la práctica local estándar. El criterio de valoración fue el tiempo hasta la mejoría clínica (tiempo hasta un NEWS2 ≤ 2 durante 24 h o al alta hospitalaria, durante los 35 días de duración del estudio. Se enrolaron 366 pacientes (241 asignados al grupo baloxavir y 125 al control). La población analizada por intención de tratamiento fue de 322 pacientes (208 en el grupo baloxivir y 114 en el control). La mediana de tiempo hasta la mejoría clínica en el grupo de baloxavir fue 97,5 horas (IC 95%: 75,9; 117,2) vs. 100,2 horas (75,9; 144,4) en el grupo de control ( $p = 0,467$ ). Se verificaron eventos adversos graves en el 12% del grupo de baloxavir vs. 15% del grupo de control. La mortalidad fue 2% en el grupo de baloxavir vs. 6% en el grupo de control, pero ninguno de los casos se consideró relacionado con el tratamiento. Por lo que se concluyó que la combinación de baloxavir con NAI no produjo resultados clínicos superiores en comparación con los NAI solos<sup>13</sup>.

### **Ensayos Controlados de No-Inferioridad**

En un ECNI, el objetivo es demostrar que una intervención no es (mucho) peor, o menos efectiva que una estándar. Es decir, que los resultados obtenidos con la nueva intervención pueden ser más eficaces, o incluso similares; sin ser inferiores a la actual<sup>3</sup>. La principal indicación para conducir un ECNI es si se quiere evaluar la eficacia de un tratamiento experimental vs. un control activo, cuando la hipótesis es que el tratamiento experimental no es necesariamente más efectivo que el tratamiento probado, pero es estadísticamente “no inferior”, o que no es inaceptablemente menos efectivo<sup>14</sup>. Sus principales dificultades son:

- *Sensibilidad del ensayo.* Hay que evaluar la evidencia histórica que el control activo es superior al placebo y lo seguirá siendo en el contexto del ECNI en curso, por lo que, es fundamental asegurarse que exista un adecuado análisis estadístico de EC realizados en circunstancias similares a las del ECNI; lo que es complejo debido a la escasa cantidad de EC controlados por placebo metodológicamente adecuados.
- *Análisis de grupos en estudio.* Esto puede ser análisis según intención de tratar o, por protocolo. El primero considera los resultados del seguimiento completo de todos los pacientes aleatorizados en el estudio independiente de su cumplimiento con el tratamiento, lo que permite disminuir las diferencias de tratamiento idealizadas que pueden surgir bajo una premisa poco realista de 100% de cumplimiento; no obstante, ello, puede incrementar artificialmente la afirmación de no inferioridad al “diluir” verdaderas diferencias de tratamiento. Y el segundo, se enfoca solo en quienes cumplieron el tratamiento, o hasta que el cumplimiento cesó en cada sujeto, con la idea de evidenciar diferencias de tratamiento no deseadas; sin embargo, se ha

considerar, que el grupo que cumple el tratamiento completo puede ser un grupo selecto que podría sesgar favorablemente el resultado, ya que en ocasiones el no seguimiento puede deberse al simple hecho de la percepción del paciente que la intervención no funciona, por tanto, deja de usarla.

- *Escoger margen M (Δ)*: la relevancia de escoger un margen de no inferioridad adecuado (no muy amplio), se da principalmente en ECNI en los que se considera poco ético realizar un EC controlado por placebo. En estos casos, lo peor que puede ocurrir es que  $\Delta$  sea tan amplio que una nueva intervención no mejor que el placebo sea aceptado como no inferior; por lo cual, el margen  $\Delta$  que se escoja debe ser significativamente menor que el estimado.
- *Tamaño de muestra*. En los ECNI la muestra va estrechamente vinculada con el margen de no inferioridad y el intervalo de confianza que se escoja, ya que si hay no-inferioridad se debe estudiar una muestra suficiente para tener mayor probabilidad que el intervalo de confianza esté en el lado  $+ \Delta$ .

Por todas estas razones, una buena proporción de los ECNI son altamente cuestionados; porque no cumplen criterios metodológicos necesarios para tener adecuada validez interna, aumentando con ello la falsa aceptación de tratamientos inferiores como terapia alternativa<sup>15</sup>. Entonces, para establecer la “no inferioridad” de un tratamiento se deben cumplir tres criterios respecto de la intervención a considerar:

- que no puede mostrar inferioridad terapéutica respecto del control activo.
- que demostraría eficacia terapéutica en un EC controlado vs. placebo.
- que ofrece ventajas secundarias en costes, seguridad (efectos adversos), tolerancia (por ej.: vía de administración) y conveniencia<sup>16</sup>.

Así, la nueva intervención puede ser menos costosa, menos invasiva y tener menos efectos secundarios. Las hipótesis de los ECNI son  $\delta \geq 0$ , o margen de significación clínica que suele ser pequeño.  $H_0: \mu_0 - \mu_1 \leq d$  vs.  $H_1: \mu_0 - \mu_1 > d$ . La “no inferioridad” respecto al control activo, se puede comprender a partir de  $H_1$ . Si la diferencia media entre el grupo de tratamiento y control es mayor que  $d$ , el tratamiento no es inferior al control. A diferencia del ECS, no se requiere que la intervención en estudio sea mejor que el control activo. Es así como, si  $\delta > 0$ , la intervención en estudio puede ser “peor” que el control (es decir,  $\mu_1 - \mu_0 < 0$ ). Sin embargo, mientras  $\mu_1 - \mu_0 > -\delta$ , la intervención en estudio es la “no inferior”.

Por ende, es más fácil determinar “no inferioridad” que superioridad. Especialmente si comparamos fórmulas de tamaño de muestra en ambos escenarios (Figura 2).

Ejemplo: “Effect of 7 vs 14 days of antibiotic therapy

*on resolution of symptoms among afebrile men with urinary tract infection: a randomized clinical trial*”. Con el objetivo de determinar si 7 días de tratamiento no es inferior a 14 días al utilizar ciprofloxacina o cotrimoxazol como tratamiento de ITU en hombres afebriles; se realizó un ECNI controlado con placebo, en dos centros médicos de veteranos de E.U.A. (2014-2019). De 1.058 hombres elegibles, 272 fueron aleatorizados. Los participantes continuaron con el antimicrobiano prescrito por 7 días y se asignaron al azar para: continuar terapia antimicrobiana continua ( $n = 136$ ), o placebo ( $n = 136$ ) durante los días 8 a 14 de tratamiento. La variable resultado fue resolución de síntomas de ITU 14 días post-tratamiento. Se seleccionó un margen de no inferioridad de 10%. Para el análisis primario, se consideró a la población tratada (quienes tomaron  $\geq 26$  de 28 dosis y no olvidaron más de dos dosis consecutivas); y para el análisis secundario, se incluyó a todos los aleatorizados, independientemente de su adherencia al tratamiento. Otras variables de interés fueron recurrencia de síntomas de ITU y eventos adversos dentro de los 28 días posteriores a la suspensión del medicamento en estudio. De los 272 pacientes aleatorizados, todos completaron el ensayo y 93,4% se incluyeron en el análisis primario. La resolución de síntomas ocurrió en 93,1% del grupo de 7 días vs. 90,2% del grupo de 14 días (diferencia 2,9% [IC unilateral de 97,5%:-5,2% a  $\infty$ ]); cumpliendo el criterio de “no inferioridad”. En el análisis secundario, la resolución de síntomas ocurrió en 91,9% del grupo de 7 días vs. 90,4% del grupo de 14 días (diferencia: 1,5% [IC unilateral de 97,5%:-5,8% a  $\infty$ ]). La recurrencia de síntomas se produjo en 9,9% del grupo de 7 días vs. 12,9% del de 14 días (diferencia: -3,0% [IC 95%: -10,8%; 6,2%],  $p = 0,70$ ). Los eventos adversos ocurrieron en 20,6% del grupo de 7 días vs. 24,3% del de 14 días. Concluyéndose que en hombres afebriles con sospecha de ITU, el tratamiento con ciprofloxacina o cotrimoxazol durante 7 días no es inferior a 14 días de tratamiento, en relación con resolución de los síntomas de ITU 14 días después de la terapia con antimicrobianos; por lo que se respalda el uso de un ciclo de 7 días como alternativa al de 14 días<sup>17</sup>.

#### Ensayo Controlado de Equivalencia

Es fundamental comprender que equivalencia no significa igualdad, ni bioequivalencia. Cuando se dice que un tratamiento y el control son equivalentes, se refiere a que son similares.

En un ECE, las pruebas estadísticas tienen como objetivo demostrar que dos intervenciones no son muy diferentes entre sí en términos de eficacia o efectividad; donde el concepto “no muy diferentes” es clínico; por lo que se intenta demostrar equivalencia terapéutica para una misma enfermedad, generalmente un fármaco nuevo vs. otro ya existente (comparador activo o estándar). Generalmente se espera que el nuevo fármaco o dispositivo

	a) Cálculo para variable resultado continua	b) Cálculo para variable resultado dicotómica
ECS ECNI	$n_1 = kn_2; n_2 = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2 \left(1 + \frac{1}{k}\right)}{(d-\delta)^2}$	$n_1 = kn_2; n_2 = \frac{(Z_\alpha + Z_\beta)^2}{(d-\delta)^2} \left[ \frac{p_1(1-p_1)}{k} + p_2(1-p_2) \right]$
ECE	$n_1 = kn_2; n_2 = \frac{(Z_\alpha + Z_{\beta/2})^2 \sigma^2 \left(1 + \frac{1}{k}\right)}{(\delta -  d )^2}$	$n_1 = kn_2; n_2 = \frac{(Z_\alpha + Z_{\beta/2})^2}{(\delta -  d )^2} \left[ \frac{p_1(1-p_1)}{k} + p_2(1-p_2) \right]$

**Figura 2.** Fórmulas para estimar tamaño de muestra para diferentes tipos de EC para variables de resultado continuas (a) y dicotómicas (b). Adaptada de Flight<sup>6</sup> y de Wang<sup>10</sup>.  $n_1$  = Número de participantes necesarios en el grupo control.  $n_2$  = Número de participantes necesarios en el grupo de intervención.  $k$  = Razón de sujetos entre el grupo de control e intervención.  $Z_\alpha$  = Valor Z correspondiente al riesgo (Error tipo I).  $Z_\beta$  = Valor Z correspondiente al riesgo (Error tipo II).  $\sigma$  = Varianza esperada de la variable respuesta.  $d$  = Tamaño del efecto que se desea detectar.  $\delta$  = Diferencia clínicamente significativa.  $p_1$  = Proporción de la variable resultado en los participantes del grupo control.  $p_2$  = Proporción de la variable resultado en los participantes del grupo intervención.

tenga la misma eficacia o efectividad que el que se está utilizando, pero que ofrezca alguna ventaja, como menor coste o efectos colaterales<sup>5,7,18</sup>.

Lamentablemente, el diseño, conducción y análisis de los resultados provenientes de ECE no son bien conocidos ni comprendidos, por lo que a menudo este tipo de estudios incluyen pocos pacientes o presentan sesgos de selección y confusión, concluyendo que no hay diferencias cuando quizás si las había<sup>18</sup>.

Un ECE debe reflejar de forma fiel los ensayos exitosos anteriores del comparador activo; y se deben minimizar las pérdidas de pacientes y otras desviaciones del protocolo. Por otra parte, el análisis por “intención de tratar” no es una buena solución para abordar problemas inevitables, por lo que es más recomendable como estrategia de análisis de resultados el análisis “por protocolo” (excluye a pacientes con violaciones importantes del protocolo, o niveles inaceptablemente bajos de adherencia al tratamiento en estudio); aunque existe escepticismo con este método, porque perturbaría la integridad de la AA<sup>2</sup>. Además, la precisión del análisis debe basarse en el IC 95%<sup>6</sup> (Figura 2).

Dicho de otra forma, un ECE se diseña y conduce para demostrar la inexistencia de diferencias entre dos intervenciones; o sea, para acreditar que la intervención en estudio tiene efectos similares a la actual; para lo cual, se debe elegir un delta o margen de diferencia clínicamente aceptable, de modo de poder definir el máximo nivel de diferencia razonable desde el punto de vista clínico<sup>19</sup>. En consecuencia, la diferencia puede ser positiva (a favor)

o negativa (en contra). De esta forma, si dos intervenciones son equivalentes, el IC 95%, debería ubicarse en el intervalo entre  $-\Delta$  y  $+\Delta$ . Por ello, para estudios de bioequivalencia suelen emplearse IC 90%<sup>9</sup>. Sin embargo, para otros tipos de resultados se sigue utilizando un IC 95%<sup>10</sup>, como es el caso de uno de los estudios que se ejemplifica más adelante.

Ahora bien, para cuantificar similitud utilizando un rango de tolerancia, las hipótesis para un ECE se especifican como:  $H_0: |\mu_1 - \mu_0| \geq \delta$  vs.  $H_1: |\mu_1 - \mu_0| < \delta$ , donde  $\delta > 0$  será el margen de tolerancia predeterminado. Si se rechaza  $H_0$ , entonces la diferencia de medias de los dos grupos estará en el rango de tolerancia y por ende, el tratamiento y el control serán equivalentes. Por tanto, las hipótesis en un ECE son las mismas que  $H_0: \mu_1 - \mu_0 \leq -\delta$  y  $\mu_0 - \mu_1 \leq \delta$  vs.  $H_1: \mu_1 - \mu_0 > -\delta$  y  $\mu_0 - \mu_1 > \delta$ . Entonces, un ECE es la intersección de dos ECNI; porque intuitivamente, el tratamiento y el control son equivalentes, si y sólo si ninguno es inferior al otro<sup>7</sup>.

Se pueden realizar ECE de tres brazos (medicamento de prueba, de referencia y placebo); que consta de tres pruebas de hipótesis, con dos que demuestran la superioridad de los fármacos de prueba y de referencia frente al placebo, y otra que demuestra la equivalencia de los fármacos de prueba y referencia. Al diseñar un ECE de tres brazos es necesario minimizar la probabilidad que se pueda encontrar un medicamento de prueba que resulte equivalente al de referencia, sin ser superior al placebo; por esto, una de las estrategias metodológicas consiste en probar la equivalencia de los fármacos (no

placebo), a través de hipótesis basadas en la razón de las diferencias de las proporciones. El tamaño de muestra requerido para lograr la potencia deseada en el nivel de significación dado se puede obtener resolviendo la función de potencia<sup>20</sup>. La conducción de un ECE exige metodologías diferentes durante el diseño y análisis comparado con ECS. Específicamente, los cálculos estándar de tamaño de muestra no son adecuados y el ECE requiere mayor cantidad de pacientes comparado con un ECAA estándar.

Ejemplo: “*Simplified antibiotic regimens for treatment of clinical severe infection in the outpatient setting when referral is not possible for young infants in Pakistan (Simplified Antibiotic Therapy Trial): a randomised, open-label, equivalence trial*”. Con el objetivo de evaluar equivalencia de dos regímenes de antimicrobianos simplificados, con menos inyecciones y administración oral en lugar de parenteral vs. tratamiento de referencia para lactantes pequeños con infección clínica grave; se realizó un ECE de tres grupos en cinco comunidades de Pakistán (2010-2013). Se reclutaron lactantes de hasta 59 días de edad, con signos de infección clínica grave. Se asignó bencilpenicilina procaína intramuscular y gentamicina una vez al día durante 7 días (referencia); amoxicilina oral dos veces al día y gentamicina intramuscular una vez al día durante 7 días; o bencilpenicilina procaína intramuscular y gentamicina una vez al día durante 2 días seguidos de amoxicilina oral dos veces al día durante 5 días. La variable resultado fue fracaso terapéutico dentro de los 7 días posteriores al inicio. Se realizó análisis por protocolo. Se consideró que los tratamientos experimentales eran tan eficaces como el de referencia si el límite superior del IC 95% para la diferencia en el fracaso del tratamiento era < 5. Se estudiaron 2.453 lactantes, pero por seguimiento clínico inadecuado y baja adherencia, se incluyeron finalmente 2.251 en el análisis por protocolo. A 820 (747 por protocolo) se les asignó el tratamiento de referencia; a 816 (751 por protocolo), amoxicilina y gentamicina; y a 817 (753 por protocolo), bencilpenicilina procaína, gentamicina y amoxicilina. Se informó fracaso de tratamiento en los 7 días posteriores al enrolamiento en 12% del grupo de referencia; 10% de los de amoxicilina+gentamicina (diferencia de riesgo con la referencia -1,9. IC 95%: -5,1; 1,3); y 13% de los tratados con bencilpenicilina+gentamicina+amoxicilina (diferencia de riesgo con referencia 1,1. IC 95%: -2,3; 4,5). Se concluyó que dos regímenes de antimicrobianos simplificados que requieren menos inyecciones son equivalentes a un tratamiento de referencia para lactantes con signos de infección clínica grave<sup>21</sup>.

Finalmente, se puede señalar, que los médicos clínicos deben tener presente que un resultado negativo en un ECS nunca probaría que las intervenciones en estudio son equivalentes, sino que es posible que se haya producido

en este caso un error de tipo 2; por lo que se señalar que los tiempos en los que solo los ECS eran suficientes han terminado. Desde hace ya tiempo, se han propuesto una variedad de opciones, entre las que destacan los ECNI y los ECE. Ambos (no indicados en determinadas áreas), exigen altos estándares para proporcionar resultados de adecuada validez interna. Los ECNI, permiten probar nuevas intervenciones no superiores al estándar, pero que podrían ser atractivas para los pacientes por exhibir menos eventos adversos; sin embargo, la elección del margen de no-inferioridad es un potencial problema. Y los ECE, nos permiten asegurar que las intervenciones son esencialmente iguales desde el punto de vista de su eficacia, y sólo difieren en costes, facilidad de administración, etc.; no obstante lo cual, se ha de tener en consideración que los márgenes de equivalencia tienden a ser grandes para ser clínicamente significativos y que la afirmación de equivalencia puede ser engañosa.

### **Herramientas para evaluar ECNI**

Por una parte, existen guías de lectura crítica para artículos de ECNI<sup>22-24</sup>, en las que se consideran las siguientes preguntas.

#### *¿Son válidos los resultados?*

¿Comenzaron las intervenciones estándar y nueva con factores pronósticos similares?

¿Se mantuvo el equilibrio pronóstico entre los dos grupos a medida que avanzaba el EC?

¿Los investigadores analizaron a los pacientes según el tratamiento que recibieron, así como los grupos a los que fueron asignados?

¿Los investigadores informaron un margen de no inferioridad predefinido?

¿Los investigadores impulsaron el estudio para una prueba de no inferioridad?

#### *¿Cuáles son los resultados?*

¿Se consideraron todos los resultados importantes para los pacientes?

¿Se estimó la precisión de los resultados?

¿Se interpretó apropiadamente el concepto de no inferioridad?

#### *¿Cómo puedo aplicar los resultados en mi práctica clínica?*

¿Los pacientes del estudio eran similares a mi(s) paciente(s)?

¿Se consideraron todos los resultados importantes para los pacientes?

Las posibles ventajas del nuevo tratamiento ¿valen el daño y costos potenciales?

Por otra parte, la declaración CONSORT considera recomendaciones para ECNI y ECE<sup>23</sup>.

**Tabla 1. Hipótesis y errores para los tres tipos de EC**

Tipo EC	$H_0$	$H_1$	Error tipo I	Error tipo II
ECS	No hay diferencia significativa entre la nueva intervención y el placebo.	Hay una diferencia significativa entre la nueva intervención y el placebo.	Concluir erróneamente superioridad de la intervención en estudio, cuando no existe.	No concluir superioridad cuando sí existe.
ECE	La nueva intervención no es equivalente a la intervención de referencia.	La nueva intervención es equivalente a la intervención de referencia.	Conclusión errónea de equivalencia de las intervenciones cuando no hay equivalencia.	No concluir equivalencia cuando sí existe.
ECNI	La nueva intervención no es no-inferior a la intervención de referencia.	La nueva intervención no es inferior a la intervención de referencia.	Conclusión errónea de no-inferioridad de la nueva intervención cuando no hay no-inferioridad.	No concluir no-inferioridad cuando sí existe.

ECS: Ensayo clínico de superioridad; ECE: Ensayo clínico de equivalencia; ECNI: Ensayo clínico de no inferioridad.

**Tabla 2. Resumen de las diferencias entre los tres tipos de EC.**

Características	ECS	ECNI	ECE
Condición para su aplicación	Comparación de nueva intervención con intervención NO estandarizada	Comparación de nueva intervención con intervención estándar	Comparación de nueva intervención con estándar (genérico)
Intervención	Nueva intervención	Nueva intervención	Nueva intervención
Control	Placebo o intervención no estandarizada	Intervención estándar	Intervención estándar
Contexto y razones para su utilización	Buscando una intervención más eficaz.	Demostrar eficacia comparable. Que la nueva intervención no sea peor que el estándar, dentro de un margen, pero que tenga menos complicaciones, o complicaciones menos graves, o administración más simple, etc.	Que las intervenciones sean sustancialmente iguales; como los estudios de bioequivalencia farmacéutica (márgenes inferior y superior respecto de concentraciones ineficaces y tóxicas); pero con costes menores.
Objetivos	Demostrar que una intervención es superior a un placebo u a otra intervención; es decir, que el tratamiento experimental proporciona un beneficio para los pacientes.	Demostrar que la intervención nueva no es menos eficaz que el estándar, pero que esta diferencia se compensa con otras propiedades de la nueva intervención.	Demostrar que las intervenciones son esencialmente iguales, pero con menores costes.
Hipótesis	1 cola	1 cola	2 colas
Clave para generar hipótesis	D	d y Δ	d y Δ
Recomendación de tipo de análisis	APIT	APIT	APIT y APP
Tipo de reporte	Valor $p$ e IC 95%	IC 95%	IC 95%

d: Tamaño del efecto. Δ: diferencia clínicamente significativa. IC 95%: Intervalo de confianza del 95%. APIT: Análisis por intención de tratar. APP: Análisis por protocolo.

## Conclusión

ECS, ECE como ECNI, son ampliamente utilizados. Las hipótesis y errores de ECS, ECE y ECNI, así como las diferencias entre ellos, se resumen en las Tablas 1 y 2. La decisión sobre utilizar uno sobre otro dependerá especialmente de las hipótesis y objetivos a desarrollar. No obstante ello, se puede señalar que:

1. Es más fácil establecer superioridad que “no inferioridad”.
2. Superioridad es un caso especial de “no inferioridad”.
3. Equivalencia es la combinación de dos ECNI.

*Agradecimientos:* Se agradece a ANID - MILENIO - NCS2021\_013 y ANID + subvención a instalación en la academia convocatoria año 2022 + Folio (85220114).

## Referencias bibliográficas

- 1.- Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *J Hepatol* 2007; 46: 947-54. <https://doi.org/10.1016/j.jhep.2007.02.015>.
- 2.- Dunn DT, Copas AJ, Brocklehurst P. Superiority and non-inferiority: two sides of the same coin? *Trials* 2018; 19: 499. <https://doi.org/10.1186/s13063-018-2885-z>.
- 3.- Lesaffre E. Noninferiority Clinical Trials: The good, the bad, and the ugly. *Semin Liver Dis* 2018; 38: 97-102. <https://doi.org/10.1055/s-0038-1655777>.
- 4.- Angeli F, Verdecchia P, Vaudo G, Masnaghetti S, Reboldi G. Optimal use of the non-inferiority trial Design. *Pharmaceut Med.* 2020; 34: 159-65. <https://doi.org/10.1007/s40290-020-00334-z>.
- 5.- Herr M, Descatha A, Aegester P. Non-inferiority and equivalence trials: Key methodological issues. *Rev Med Interne* 2018; 39: 352-9. <https://doi.org/10.1016/j.revmed.2017.06.010>.
- 6.- Flight L, Julious SA. Practical guide to sample size calculations: non-inferiority and equivalence trials. *Pharm. Stat.*, 2016; 15: 80-9. <https://doi.org/10.1002/pst.1716>.
- 7.- Stefanos R, Graziella D', Giovanni T. Methodological aspects of superiority, equivalence, and non-inferiority trials. *Intern Emerg Med* 2020; 15: 1085-91. <https://doi.org/10.1007/s11739-020-02450-9>.
- 8.- Manterola C, Otzen T. Estudios Experimentales 1<sup>a</sup> Parte. El Ensayo Clínico. *Int J Morphol* 2015; 33: 342-9. <https://doi.org/10.4067/S0717-95022015000100054>.
- 9.- Kishore, K. & Mahajan, R. Understanding superiority, noninferiority, and equivalence for clinical trials. *Indian Dermatol Online J* 2020; 11: 890-4. [https://doi.org/10.4103/idoj.IDOJ\\_130\\_20](https://doi.org/10.4103/idoj.IDOJ_130_20).
- 10.- Wang B, Wang H, Tu XM, Feng C. Comparisons of superiority, non-inferiority, and equivalence trials. *Shanghai Arch. Psychiatry* 2017; 29: 385-8. <https://doi.org/10.11919/j.issn.1002-0829.217163>.
- 11.- Sackett DL. Superiority Trial. Wiley StatsRef: Statistics Reference Online, 2014.
- 12.- Chow SH, Liu JP. Design and analysis of clinical trials: Concepts and Methodologies (3<sup>rd</sup> ed.), 2013. Wiley Series in Probability and Statistics.
- 13.- Kumar D, Ison MG, Mira JP, Welte T, Hwan Ha J, Hui DS, et al. Combining baloxavir marboxil with standard-of-care neuraminidase inhibitor in patients hospitalised with severe influenza (FLAGSTONE): a randomised, parallel-group, double-blind, placebo-controlled, superiority trial. *Lancet Infect Dis.* 2022; 22: 718-30. [https://doi.org/10.1016/S1473-3099\(21\)00469-2](https://doi.org/10.1016/S1473-3099(21)00469-2).
- 14.- Garattini S, Bertele' V. Noninferiority trials. *N Engl J Med.* 2018; 378: 303. <https://doi.org/10.1056/NEJMcl1714668>.
- 15.- Kaji AH, Lewis RJ. Noninferiority trials: is a new treatment almost as effective as another? *JAMA*. 2015; 313: 2371-2. <https://doi.org/10.1001/jama.2015.6645>.
- 16.- Ruste M, Fellahi JL, Jacquet-Lagrèze M. Noninferiority trials: a specific design with a particular methodology. *Ann Intensive Care.* 2022; 12: 45. <https://doi.org/10.1186/s13613-022-01020-w>.
- 17.- Drekonja DM, Trautner B, Amundson C, Kuskowski M, Johnson JR. Effect of 7 vs 14 days of antibiotic therapy on resolution of symptoms among afebrile men with urinary tract infection: A randomized clinical trial. *JAMA*. 2021; 326: 324-31. <https://doi.org/10.1001/jama.2021.9899>.
- 18.- Ranganathan P, Pramesh CS, Aggarwal R. Equivalence trials. *Perspect Clin Res.* 2022; 13: 114-7. [https://doi.org/10.4103/picr.picr\\_55\\_22](https://doi.org/10.4103/picr.picr_55_22).
- 19.- Rosas-Peralta M, Santos-Martínez LE, Magaña-Serrano JA, Valencia-Sánchez JS, Garrido-Garduño M, Pérez-Rodríguez G. Methodology for superiority versus equivalence and non-inferior clinical studies. A practical review. *Rev Med Inst Mex Seguro Soc.* 2016; 54: 344-53. <https://ichgcp.net/pl/cClinical-trials-registry/publications/115047-methodology-for-superiority-versus-equivalence-and-non-inferior-clinical-studies-a-practical-review>.
- 20.- Tseng YK, Hsu KN. Study design for a three-arm equivalence clinical trial with binomially distributed outcomes. *J Biopharm Stat.* 2021; 31: 736-44. <https://doi.org/10.1080/10543406.2021.1968891>.
- 21.- Mir F, Nisar I, Tikmani SS, Baloch B, Shakoor S, Jehan F, et al. Simplified antibiotic regimens for treatment of clinical severe infection in the outpatient setting when referral is not possible for young infants in Pakistan (Simplified Antibiotic Therapy Trial [SATT]): a randomized, open-label, equivalence trial. *Lancet Glob Health.* 2017; 5: e177-e185. [https://doi.org/10.1016/S2214-109X\(16\)30335-7](https://doi.org/10.1016/S2214-109X(16)30335-7).
- 22.- Mulla SM, Scott IA, Jackevicius CA, You JJ, Guyatt GH. How to use a noninferiority trial: users' guides to the medical literature. *JAMA* 2012; 308: 2605-11. <https://doi.org/10.1001/2012.jama.11235>.
- 23.- Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA.* 2012; 308: 2594-604. <https://doi.org/10.1001/jama.2012.87802>.
- 24.- Thoma A, Farrokhyar F, Walther D, Braga LH, Sprague S, Goldsmith CH. Users' guide to the surgical literature: how to assess a noninferiority trial. *Can J Surg* 2017; 60: 426-32. <https://doi.org/10.1503/cjs.000317>.