

Artículo Especial: Ensayos aleatorizados de no inferioridad y de equivalencia

Non inferiority and equivalence randomized trials

Agustín Ciapponi *

Resumen

En este artículo explica los conceptos básicos de los ensayos aleatorizados de no inferioridad y de equivalencia y presenta una ampliación y una puesta al día de la declaración CONSORT 2010 para el reporte de ensayos clínicos¹.

Abstract

This article explains the basics of noninferiority and equivalence randomized trials and presents an extension and an update of the CONSORT statement (2010) for clinical trials reporting¹.

Palabras clave: ensayo clínico, bioequivalencia, no inferioridad. **Key words:** clinical trial, bioequivalence, non inferiority.

Ciapponi A. Ensayos aleatorizados de no inferioridad y de equivalencia. Evid Act Pract Ambul. Jul-Sep 2013;16(3):92-94.

Importancia de los diseños de no inferioridad o de equivalencia

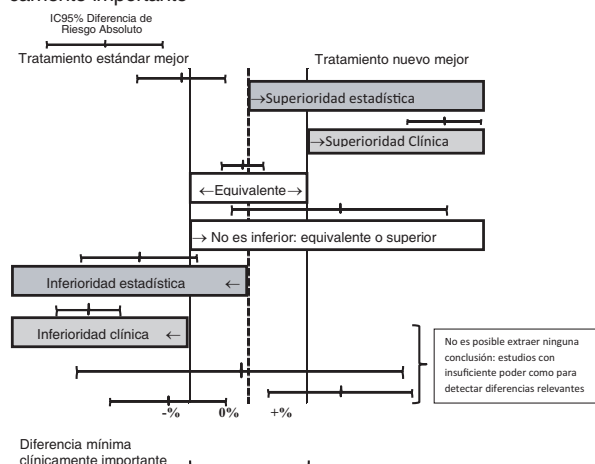
La mayoría de los ensayos tienen como objetivo determinar si una intervención es superior a otra. Estos son los ensayos de superioridad, que intentan rechazar la hipótesis nula de que ambas alternativas son iguales. En cambio, los ensayos de equivalencia o de no inferioridad se emplean para demostrar que ante un mejor perfil de costos, seguridad, disponibilidad, facilidad de administración o de otras ventajas, un nuevo tratamiento tiene al menos igual eficacia que el tratamiento estándar.

Que no se haya logrado constatar una diferencia estadísticamente significativa no implica que se haya demostrado la igualdad. Dicho de otro modo, en un test de hipótesis, la hipótesis nula (H0) sostiene que no hay diferencias. Por lo tanto se la puede rechazar de forma concluyente si se detectan diferencias que tienen baja probabilidad de haber ocurrido por azar (estadísticamente significativas; $p < 0,05$), pero nunca se podrá decir que se ha demostrado que H0 sea cierta porque no pudieron documentarse estas diferencias, ya que la significancia estadística depende mucho del tamaño muestral. La falta de significancia estadística sólo demuestra que no hay datos suficientes como para hacer inferencias o bien que las diferencias halladas tienen una probabilidad de haber sido por azar mayor a nuestro umbral de aceptación (en general 5% o una posibilidad en 20).

La incapacidad de demostrar la verdad de H0 a partir de resultados no significativos crea un problema cuando lo que se persigue es, justamente, demostrar igualdad. La solución es pensar que los ensayos se hacen para demostrar que hay diferencias clínicamente importantes, y no simplemente cualquier diferencia².

Se habla de la diferencia mínima clínicamente importante (DMCI), que es definida mediante criterios prácticos no relacionados con la estadística, sino con el efecto que tendrá esa diferencia en la salud de los pacientes. Ver figura 1.

Figura 1. Resultados de estudios según la diferencia mínima clínicamente importante



Por ejemplo, si se asume que clínicamente no tendrá importancia aumentar en medio centímetro el perímetro de la cintura (0,5 cm) ya que no se asociará a resultados clínicamente relevantes (aumento de la mortalidad o de la incidencia de alguna forma de enfermedad cardiovascular), pero a partir de un aumento de un centímetro 1 cm, el cambio comienza a ser importante, se podría convenir que una DMCI es de 1 cm. Esto implica que diferencias inferiores a 1 cm no serán consideradas importantes.

Los ensayos de equivalencia valoran si dos terapias son iguales y se basan en este margen de DMCI ("delta", Δ) que se fija a priori, durante el diseño de la investigación. Se considera que este es el límite a partir del cual una diferencia tendrá importancia desde el punto de vista práctico o clínico. El objetivo de este tipo de trabajo es demostrar que la diferencia entre la nueva terapia y la convencional estará completamente dentro de ese margen. Se calculará un intervalo de confianza del 95% (IC95%) para la diferencia entre ambos grupos y se demostrará la igualdad, siempre y cuando ese intervalo de confianza de la diferencia entre los tratamientos comparados esté totalmente incluido dentro del margen "delta", que suele ser muy pequeño. Ver figura 1.

Por esto mismo, los ensayos de equivalencia requieren mayor tamaño muestral, pues a mayor tamaño muestral más estrecho será el intervalo de confianza de cualquier diferencia que pueda obtenerse. Mientras que los ensayos de equivalencia intentan probar que una nueva terapia obtiene un resultado similar al tratamiento estándar, es decir, que el efecto del nuevo tratamiento está entre $-\Delta$ y $+\Delta$; los de no inferioridad intentan probar que una nueva terapia no es peor que el tratamiento habitual, o que la diferencia entre ambos sería muy pequeña e inferior a Δ . De este modo, la pregunta en los ensayos de no inferioridad no es simétrica dado que una eventual superioridad del nuevo tratamiento para el resultado primario sería un beneficio adicional no contemplado en el diseño. Por eso, un ensayo de no inferioridad, se suele valorar con pruebas de significación estadística a una cola.

Algo distintos a los ensayos de equivalencia y a los de no inferioridad ya mencionados, son los ensayos de bio-equivalencia, destinado solamente a valorar si el principio activo de un medicamento genérico tiene las mismas características fármaco-dinámicas que el fármaco estándar de referencia. Algunos ensayos de no inferioridad han sido criticados por ser meros estudios de un nuevo producto comercializable (medicamentos llamados "yo también" o "me too") que no ofrecería ventajas sobre otros productos ya existentes en el mercado. Quienes sostienen una visión crítica sobre este tipo de investigación, lo hacen considerando que utilizan preguntas clínicamente irrelevantes y por lo tanto son investigaciones sin sustento ético³. Sin embargo, esta visión no es compartida por otros autores⁴.

Una reciente revisión de ensayos de no inferioridad de 583 tratamientos farmacológicos publicados entre 1989 y 2009 mostró una tendencia creciente en la publicación de este tipo de investigación, con sólo un ensayo publicado antes de 1999 y más de 100 publicados anualmente a partir de 2007⁵. Un tercio de ellos pertenecían al campo de las enfermedades infecciosas o la cardiología.

Calidad de reporte de los ensayos de no inferioridad o de equivalencia

Las primeras revisiones de la calidad de los ensayos que afirman ser de equivalencia encontraron con frecuencia falencias importantes. Se argumentaba indebidamente equivalencia en el 67% de los 88 estudios publicados desde 1992 hasta 1996 sobre la base de pruebas no estadísticamente significativas de superioridad⁶. La mitad de los estudios declaró la equivalencia como un objetivo, pero sólo el 23% reportó que estos habían sido diseñados con un margen preestablecido de equivalencia. Revisiones más recientes han encontrado que la calidad de los informes de ensayos de no inferioridad y/o de equivalencia sigue siendo pobre. Una revisión que abarca desde 1990 a 2000 documentó que sólo alrededor de una quinta parte de los ensayos de no inferioridad y de equivalencia había proporcionado una justificación adecuada para el margen de no inferioridad⁷. En otra revisión que incluyó los ensayos de no inferioridad indexados en PubMed desde febrero de 2009, los autores constataron que casi todos de estudios publicados sobre ensayos de equivalencia de medicamentos (y/o de no inferioridad) habían especificado el margen de no inferioridad, pero sólo 24% explicaba cómo había sido calculado^{8,9}.

Extensión de la declaración CONSORT 2010 para ensayos de no inferioridad o de equivalencia

La Declaración CONSORT (Consolidated Standards of Reporting Trials), que incluye una lista de cotejo y un diagrama

de flujo, es una guía desarrollada para ayudar a los autores a mejorar el reporte de los resultados de los ensayos controlados aleatorizados (ECA)¹⁰. Se actualizó por última vez en 2010. Su enfoque principal es en los ECA individuales con dos grupos paralelos para evaluar la posible superioridad de un tratamiento en comparación con otro.

La Declaración CONSORT se ha extendido a otros diseños de ensayos tales como los ensayos con aleatorización de conglomerados (grupos o "conglomerados" de individuos) o los ensayos de no-inferioridad y de equivalencia.

Presentamos lista de cotejo actualizada para reportar ensayos de no inferioridad y de equivalencia, sobre la base de la versión de 2010 de la Declaración CONSORT, que permite a los lectores a evaluar la fiabilidad de sus resultados y conclusiones.

La lista de cotejo para ensayos de no inferioridad, actualizada en consonancia con la Declaración CONSORT 2010, se presenta en la tabla 1. La misma se muestra en dos columnas.

La primera comprende la declaración CONSORT 2010 y la segunda, la extensión para ensayos de no inferioridad y de equivalencia, brindándose ejemplos como notas al pie de los ítems considerados necesarios.

Abarca los siguientes aspectos principales: 1) la justificación para la adopción de un diseño de no inferioridad, 2) cómo las hipótesis del estudio fueron incorporadas en el diseño, 3) la elección de los participantes, las intervenciones (especialmente del tratamiento de referencia) y los resultados, 4) los métodos estadísticos utilizados, incluyendo el cálculo del tamaño muestral, 5) cómo el diseño afecta la interpretación y las conclusiones.

Tabla 1: información que se debe incluir al informar un ensayo aleatorizado de no inferioridad o de equivalencia (extensión de la lista de cotejo CONSORT 2010).

Sección/tema	Ítem N°	Ítem de la lista de comprobación	Extensión para ensayos de no inferioridad	
Título Resumen [§]	1a	Identificado como un ensayo aleatorizado en el título	Identificado como un ensayo aleatorizado de no inferioridad en el título	
	1b	Resumen estructurado del diseño, métodos, resultados y conclusiones del ensayo (para una orientación específica, véase CONSORT for abstracts)	Se deben reflejar además las hipótesis de no inferioridad, especificando el margen de no inferioridad para cada resultado de no inferioridad, considerando en la conclusión las hipótesis de no inferioridad y superioridad.	
Introducción	Antecedentes**	2a	Antecedentes científicos y justificación	Justificación para el diseño de no inferioridad
	Objetivos††	2b	Objetivos específicos o hipótesis	Hipótesis sobre no inferioridad, especificando el margen de no inferioridad y su justificación
Métodos	Diseño del ensayo	3a	Descripción del diseño del ensayo (por ejemplo, paralelo, factorial), incluida la razón de asignación	
		3b	Cambios importantes en los métodos después de iniciar el ensayo (por ejemplo, criterios de selección) y su justificación	
	Participantes	4a	Criterios de selección de los participantes	Reportar si los participantes en el ensayo de no inferioridad son similares a los de cualquier ensayo(s) de eficacia del tratamiento de referencia
		4b	Procedencia (centros e instituciones) en que se registraron los datos	Reportar si el tratamiento de referencia en el ensayo de no inferioridad es idéntico (o muy similar) que en cualquier ensayo(s) que estableció eficacia
Intervenciones	5	Descripción de las intervenciones para cada grupo con detalles suficientes como para permitir la replicación, incluidos cómo y cuándo se administraron realmente	Especificar el resultado(s) de no inferioridad y si la hipótesis para el resultado(s) principal y secundario son de no inferioridad o superioridad; así como si los resultados del estudio de no inferioridad son idénticas (o muy similares) a los de cualquier ensayo(s) que estableció la eficacia del tratamiento de referencia.	
	Resultados	6a	Especificación a priori de las variables de respuesta (o desenlace) principal(es) y secundarias incluidas y cómo y cuándo fueron evaluadas	
		6b	Cualquier cambio en las variables de respuesta tras el inicio del ensayo, junto con los motivos de la(s) modificación(es)	
Tamaño muestral†††	7a	Cómo se determinó el tamaño muestral	Si fue calculado usando un criterio de no inferioridad y de ser así, cuál fue el margen de no inferioridad	
	7b	Si corresponde, explicar cualquier análisis intermedio y las reglas de interrupción	A qué resultado(s) aplican y si están relacionados con una hipótesis de no inferioridad	
Aleatorización	Generación de la secuencia	8a	Método utilizado para generar la secuencia de asignación aleatoria	
		8b	Tipo de aleatorización; detalles de cualquier restricción (como bloques y tamaño de los bloques)	
	Mecanismo de ocultación de la asignación	9	Mecanismo utilizado para implementar la secuencia de asignación aleatoria (como contenedores numerados de modo secuencial), describiendo los pasos realizados para ocultar la secuencia hasta que se asignaron las intervenciones	
	Implementación	10	Quién generó la secuencia de asignación aleatoria, quién seleccionó a los participantes y quién asignó los participantes a las intervenciones	



Enmascaramiento	11a	Si se realizó, a quién se mantuvo cegado después de asignar las intervenciones (por ejemplo, participantes, cuidadores, evaluadores del resultado) y de qué modo	
	11b	Si es relevante, descripción de la similitud de las intervenciones	
Métodos estadísticos	12a	Descripción de métodos estadísticos utilizados para comparar los grupos en cuanto a la variable de respuesta principal y las secundarias; y si fue utilizado un abordaje de una o dos colas del intervalo de confianza	
	12b	Métodos de análisis adicionales como análisis de subgrupos y análisis ajustados	
Resultados	13a	Para cada grupo, el número de participantes asignado aleatoriamente, que recibieron el tratamiento propuesto y que fueron incluidos en el análisis principal	
	13b	Para cada grupo, pérdidas y exclusiones después de la aleatorización, junto con los motivos	
Reclutamiento	14a	Fechas que definen los períodos de reclutamiento y de seguimiento	
	14b	Causa de la finalización o de la interrupción del ensayo	
Datos basales	15	Una tabla que muestre las características basales demográficas y clínicas para cada grupo	
Números analizados	16	Para cada grupo, número de participantes (denominador) incluidos en cada análisis y si el análisis se basó en los grupos inicialmente asignados	
Resultados y estimación***	17a	Para cada respuesta o resultado final principal y secundario, los resultados para cada grupo, el tamaño del efecto estimado y su precisión (como intervalo de confianza del 95%)	Para el resultado (s) para los que existe una hipótesis de no inferioridad, recomiendan incluir una figura que muestre los intervalos de confianza y el margen de no inferioridad
	17b	Para las respuestas dicotómicas, se recomienda la presentación de los tamaños del efecto tanto absoluto como relativo	
Análisis secundarios	18	Resultados de cualquier otro análisis realizado, incluido el análisis de subgrupos y los análisis ajustados, diferenciando entre los especificados a priori y los exploratorios	
Discusión	Daños (perjuicios)	19	Todos los daños (perjuicios) o efectos no intencionados en cada grupo (para una orientación específica, véase CONSORT for harms)
	Limitaciones	20	Limitaciones del estudio, abordando las fuentes de posibles sesgos, las de imprecisión y, si es pertinente, la multiplicidad de análisis
	Generalización	21	Posibilidad de generalización (validez externa, aplicabilidad) de los hallazgos del ensayo
	Interpretación†††	22	Interpretación consistente con los resultados, con balance de beneficios y daños, y considerando otras evidencias relevantes
Otra información	Registro	23	Número de registro y nombre del registro de ensayos.
	Protocolo	24	Dónde puede accederse al protocolo completo del ensayo, si está disponible.
	Financiación	25	Fuentes de financiación y otras ayudas (como suministro de medicamentos) y papel de los financiadores.

⁹ Métodos-Objetivo: "...para determinar si la duloxetina no es inferior a (tan bueno como) pregabalina en el tratamiento del dolor asociado con la neuropatía diabética periférica. ... se estableció no inferioridad si la mejoría en la media semanal de la puntuación de dolor diario con duloxetina no era peor que la mejoría media con pregabalina, dentro de la variabilidad estadística, por un margen de -0,8 unidades..."

Resultados: "...el límite inferior del intervalo de confianza del 97,5% correspondía a una diferencia de medias -0,05, para establecer la no inferioridad..."

Conclusiones: "...la duloxetina no fue inferior a la pregabalina para el tratamiento del dolor en pacientes con neuropatía diabética periférica con respuesta analgésica insuficiente a la gabapentina..."

*** "...las guías alemanas consideran al fluorouracilo como un adyuvante del tratamiento estándar para el cáncer rectal localmente avanzado. La capecitabina es un derivado de fluoropirimidina oral que fue tan eficaz como el fluorouracilo y el ácido folínico para el tratamiento adyuvante del cáncer de colon en estadio III. Tampoco fue inferior al fluorouracilo en combinación con oxaliplatino como tratamiento de primera línea del cáncer colorrectal metastásico...ningún ensayo aleatorizado ha comparado capecitabina con fluorouracilo perioperatorio en la enfermedad localmente avanzada. Nuestra elección de un diseño de ensayo de no inferioridad se basa en la expectativa de que la no inferioridad de la capecitabina, administrada por vía oral en forma ambulatoria, sería suficiente para inclinar la relación riesgo-beneficio en su favor..."

†† "...un análisis secuencial para la comparación de antiplaquetarios fue desarrollado y planificado para probar por primera vez la no inferioridad del uso combinado de aspirina más dipiridamol de liberación prolongada en comparación con el clopidogrel. Si esta condición se cumpliera, entonces la superioridad de la aspirina más dipiridamol de liberación prolongada versus el clopidogrel podría ser evaluada en una segunda prueba de hipótesis nula convencional de ausencia de diferencia entre los dos tratamientos..."

†††Ejemplo 1 (no inferioridad): "...con 1.715 accidentes cerebrovasculares recurrentes, tendríamos un poder estadístico de 82% para rechazar la hipótesis nula de inferioridad, suponiendo una reducción del riesgo relativo del 6,5% con aspirina más dipiridamol de liberación prolongada en comparación con clopidogrel..."

Ejemplo 2 (equivalencia): "...el margen de equivalencia (Δ) fue del 5% y el rango de -5% a 5% fue predefinido como aceptable de las tasas de terminación [de aborto médico] entre los dos tipos de proveedores. El margen se basa en diferencias clínica y estadísticamente importantes, así como en criterios éticos, de costos y de viabilidad. El tamaño de la muestra de 1.086 mujeres se estimó como suficiente (con un IC95% y una potencia de 80%) para establecer la equivalencia. El cálculo del tamaño de la muestra permitió 10% de pérdidas durante el seguimiento..."

⁹⁹ Para la variable principal de eficacia, la no inferioridad de lumiracoxib con indometacina podría afirmarse si el límite inferior del IC parara la diferencia en el cambio medio de la intensidad del dolor evaluado en una escala Likert de 5 puntos fuera mayor que -0,5. Esta prueba de no inferioridad sólo se realizó para la variable principal de eficacia..."

*** "...el RR no ajustado que compara la supervivencia global entre el grupo DGLC [discección de ganglio linfático centinela] independiente y el grupo ALND [discección de los ganglios linfáticos axilares] fue 0,79 (IC90%, 0,56 a 1,10), y no cruzó el límite especificado de 1,3. El RR ajustado fue de 0,87 (IC del 90% 0,62 a 1,23)...". Ver figura 1.

††† "...la diferencia de riesgo de aborto completo fue de 1,24% (IC95% -0,53 a 3,02, y cae dentro del rango de equivalencia predefinido (-5% a 5%), lo que nos permite concluir que la provisión de abortos con medicamentos hasta la novena semana de gestación por proveedores de nivel medio y los médicos fue de similar eficacia..."

Recibido el 28/06/13 y aceptado el 19/07/13.

Referencias

- Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG, Group C. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA. Dec 26 2012;308(24):2594-2604.
- Mayo N. Randomised trials and other parallel comparisons of treatment. In: Bailar JC III Ed, ed. Medical uses of statistics, 3ª ed. N. Jersey: John Wiley & Sons; 2009:51-90.
- Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. Lancet. Dec 1 2007;370(9602):1875-1877.
- Nunn AJ, Meredith SK, Spiegelman MK, Ginsberg AM, Gillespie SH. The ethics of non-inferiority trials. Lancet. Mar 15 2008;371(9616):895; author reply 896-897.
- Suda KJ, Hurley AM, McKibbin T, Motl Moroney SE. Publication of noninferiority clinical trials: changes over a 20-year interval. Pharmacotherapy. Sep 2011;31(9):833-839.
- Greene WL, Concato J, Feinstein AR. Claims of equivalence in medical research: are they supported by the evidence? Ann Intern Med. May 2 2000;132(9):715-722.
- Lange S, Freitag G. Choice of delta: requirements and reality--results of a systematic review. Biometrical journal. Biometrische Zeitschrift. Feb 2005;47(1):12-27; discussion 99-107.
- Wangge G, Klungel OH, Roes KC, de Boer A, Hoes AW, Knol MJ. Interpretation and inference in noninferiority randomized controlled trials in drug research. Clinical pharmacology and therapeutics. Sep 2010;88(3):420-423.
- Wangge G, Klungel OH, Roes KC, de Boer A, Hoes AW, Knol MJ. Room for improvement in conducting and reporting non-inferiority randomized controlled trials on drugs: a systematic review. PLoS One. 2010;5(10):e13550.
- Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet. Apr 14 2001;357(9263):1191-1194.