

# INFAC METODOLOGÍA: ENSAYOS DE NO INFERIORIDAD, ANÁLISIS DE SUBGRUPOS

## SUMARIO

- ▶ INTRODUCCIÓN
- ▶ ENSAYOS CLÍNICOS DE NO INFERIORIDAD
  - Margen de no inferioridad o valor  $\delta$  o  $\Delta$
  - Consideraciones metodológicas
  - Interpretación de los resultados
- ▶ ANÁLISIS DE SUBGRUPOS
  - Cuestiones preliminares
  - Criterios para valorar la posible diferencia encontrada entre los subgrupos
  - Conclusiones

## INTRODUCCIÓN

La investigación clínica con fármacos permite generar conocimiento de calidad. De entre las diferentes formas de investigación clínica, el ensayo clínico aleatorizado y controlado (ECA) es el que genera una evidencia científica más robusta<sup>1</sup>.

Clásicamente, los ensayos clínicos se han diseñado para demostrar que un nuevo tratamiento es superior al tratamiento que se utiliza como comparador o a un placebo. Durante los últimos años ha aumentado el número de ensayos clínicos de no inferioridad<sup>1</sup> cuyo objetivo es demostrar que el tratamiento en estudio no es peor que el comparador tomando como referencia unos valores preespecificados de la variable estudiada<sup>2</sup>. Este incremento se explica, probablemente, por la dificultad de desarrollar fármacos que aporten un valor terapéutico añadido en términos de eficacia respecto a los ya disponibles y por los criterios de autorización de nuevos medicamentos por parte de las agencias reguladoras, que exigen demostrar un balance beneficio-riesgo positivo independientemente de si aportan o no un beneficio incremental ante otras opciones terapéuticas ya autorizadas<sup>1</sup>. Sin embargo, los ensayos de no inferioridad han sido también criticados y, en opinión de algunos autores, la exposición de los pacientes a un tratamiento que no va a ser mejor que otros ya existentes no tiene justificación ética. Si se supone que el fármaco presenta ventajas en tolerancia y/o sencillez de uso, para demostrarlo, debería realizarse un estudio de superioridad ya que cualquier supuesta ventaja debería traducirse en un mejor cumplimiento y en un resultado superior en lugar de no inferior<sup>3</sup>.

Por otro lado, cada vez es más frecuente que los ensayos clínicos y metaanálisis incluyan el análisis de subgrupos. Este análisis de subgrupos se realiza cuando interesa evaluar los resultados de algún grupo especial de sujetos del ECA, habitualmente en función de su edad, sexo, gravedad, comorbilidad, etc. En estos análisis, la probabilidad de encontrar un resultado significativo debido al azar aumenta con el número de análisis que realicemos y su uso puede ser aún más controvertido, cuando el tratamiento no demuestra los efectos esperados, ya que podría tratarse de «encontrar a toda costa» algún grupo de pacientes a los que el tratamiento beneficie<sup>4</sup>.

El objetivo de este INFAC es ofrecer a los lectores herramientas para realizar una lectura crítica de los ensayos de no inferioridad y de análisis de subgrupos.

## ENSAYOS CLÍNICOS DE NO INFERIORIDAD

Un ensayo de no inferioridad está diseñado para demostrar que el tratamiento experimental es no inferior al tratamiento comparador por una diferencia clínica aceptable denominada margen de no inferioridad o valor delta. Por el contrario, los ensayos clínicos clásicos están diseñados para mostrar que un tratamiento experimental es superior a placebo o al tratamiento comparador<sup>5</sup>. En la tabla 1 se pueden ver las diferencias de ambos tipos de ensayos.

Tabla 1. Comparación entre los estudios de superioridad y no inferioridad (modificado de 5)

Variable*	Superioridad	No inferioridad
Comparador	Tratamiento estándar o placebo	Tratamiento estándar (placebo en estudios de seguridad)
Margen de no inferioridad	No aplica	Predeterminado
Hipótesis nula	Los dos tratamientos son iguales	El nuevo tratamiento es peor porque supera un margen preespecificado
Hipótesis alternativa	Hay diferencia entre los tratamientos	No hay diferencia entre los tratamientos
Error alfa	5% (dos lados)	2,5% (un lado)
Error beta	< 0,20	< 0,10
Error delta	Utilizado para estimar el tamaño de la muestra	Utilizado para estimar el tamaño de la muestra y para el análisis estadístico
Análisis	Por Intención de Tratar	Por Protocolo y por Intención de Tratar

\*estos conceptos estadísticos se pueden consultar, entre otras, en las siguientes direcciones web:

- Interpretando la literatura médica ¿qué necesito saber? Parte I y parte II. INFAC. 2006;14(7,8). Disponible en: <https://www.euskadi.eus/informacion/boletin-infac/web01-a2cevime/es/>
- Estudiantes por la mejor evidencia (ExME): <https://s4be.cochrane.org/blog/tag/keyconcepts/>
- Salazar C, del Castillo S. Fundamentos básicos de estadística. 1.ª ed. Disponible en: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKewjAhKfyzMT-AhXCiFwKHdhDRYQF-noECA0QAQ&url=http%3A%2F%2Fup.ac.pa%3A8080%2Fxmlui%2Fbitstream%2Fhandle%2F123456789%2F1371%2Festad%25C3%25ADsticaFundamentos%2520B%25C3%25A1sicos%2520de%2520-Libro.pdf%3Fsequence%3D1%26isAllowed%3Dy&usq=AOvVaw2owUFMfp8yeHVL-ZyRbj5d>
- Cobo E, González JA; Muñoz P, Bigorra J, Corchero C, Miras F, et al. Bioestadística para no estadísticos. Barcelona: Elsevier;2007. Disponible en: <https://blog.utp.edu.co/estadistica/files/2017/02/Bioestad%C3%ADstica-Para-no-Estad%C3%ADsticos-Cobo-1ed.pdf>

Estos ensayos podrían tener interés en los siguientes casos<sup>1,2</sup>:

- el desarrollo de nuevas formas farmacéuticas o formas galénicas de fármacos ya disponibles cuando no es posible realizar estudios de bioequivalencia
- nuevos fármacos para patologías para las que no es ético el uso de placebo como comparador
- nuevos fármacos que no se espera que sean más eficaces pero que podrían presentar potenciales ventajas en seguridad, pauta de tratamiento o costes
- ensayos clínicos en los que el objetivo principal es demostrar seguridad

Este tipo de ensayos suelen tener dos brazos, y habitualmente se compara un tratamiento en estudio con otro que actúa de comparador, pero pueden incluir un tercer brazo con placebo. Este último tipo de diseño hace posible proporcionar una demostración directa de la superioridad del tratamiento en estudio y del comparador frente a placebo<sup>2,5</sup> y es apropiado en los casos en los que sea posible desde el punto de vista ético, y especialmente para patologías con tendencia a la mejora espontánea o cuando la medida del efecto es poco sensible por ejemplo, en el dolor o la rinitis<sup>1</sup>.

El uso de placebo como comparador apenas se usa en este tipo de ensayos, si se usa en ensayos diseñados específicamente para examinar la seguridad de nuevos tratamientos<sup>5</sup> como en los ensayos de seguridad cardiovascular de los nuevos hipoglucemiantes<sup>1</sup>.

## MARGEN DE NO INFERIORIDAD O VALOR DELTA $\delta$ O $\Delta$

**La selección del margen de no inferioridad es clave** y representa uno de los principales retos de los ensayos de no inferioridad. Cuantifica cuánto de peor se permite que sea el tratamiento experimental considerando sus ventajas adicionales<sup>5</sup>, determina el tamaño de la muestra del estudio (cuanto mayor es el margen de no inferioridad, menor es el tamaño de la muestra necesaria)<sup>1,5,6</sup> y se usa como valor de referencia para las conclusiones de no inferioridad<sup>2,6</sup>. Debe especificarse para cada contexto clínico en base a una combinación entre el análisis estadístico y el criterio clínico que puede variar según patología y tipo de fármaco<sup>1,5</sup>. Tanto la [FDA](#) como la [EMA](#) tienen guías para la elección de este margen que orientan sobre cómo se debe elegir, aunque en general no indican valores concretos.

En primer lugar, se debe revisar la evidencia disponible sobre la eficacia del tratamiento comparador frente a placebo en la patología que se quiere estudiar y posteriormente, se decidirá en base al criterio clínico, qué margen de esa eficacia se debe conservar para poder considerar al nuevo tratamiento como no inferior<sup>1,2,5,7</sup>. No puede exceder el efecto de magnitud más pequeño que sería plausible que tuviera el tratamiento comparador con respecto a placebo<sup>7,8</sup>. Cuanto mayor es el margen, más fácil es concluir la no inferioridad y más riesgo hay de considerar como similar un tratamiento que realmente es peor que el comparador<sup>1,5</sup>.

En general, los márgenes de no inferioridad deberían estar fijados en diferencias relativas (riesgo relativo, odds ratio, hazard ratio (HR)) ya que éstas se ven menos afectadas por la variabilidad en la incidencia de eventos observados y estimados. Tampoco deberían utilizar resultados combinados de eficacia y seguridad ya que un tratamiento experimental poco eficaz pero más seguro podría parecer tan bueno o incluso mejor que el tratamiento comparador<sup>5</sup>.

Dos ejemplos de ensayos de no inferioridad con medicamentos utilizados en atención primaria son los estudios de seguridad cardiovascular de los nuevos antidiabéticos<sup>1,9</sup> y los de los nuevos anticoagulantes frente a warfarina, donde se buscaba demostrar que éstos eran al menos tan eficaces como la warfarina en la reducción de eventos tromboembólicos<sup>7,8</sup>:

- En la mayoría de los estudios de seguridad cardiovascular de antidiabéticos se utiliza como margen de no inferioridad el valor 1,3 para el límite superior del intervalo de confianza de la HR. Esto se interpreta como que se acepta que con el nuevo tratamiento se incremente hasta un 30% el riesgo de sufrir un acontecimiento respecto a placebo para ser considerado no inferior a éste. En algunos casos la FDA también acepta un margen de HR = 1,8 (como en el de semaglutida). Se ha cuestionado si este margen no es demasiado laxo<sup>1,9</sup>.
- En el caso de los nuevos anticoagulantes, el margen de no inferioridad se fijó en 1,46. Este margen se calculó de la siguiente manera: primero se cuantificó el efecto de la warfarina en la prevención de eventos tromboembólicos respecto a placebo o ausencia de tratamiento a partir de los datos de un metaanálisis y se obtuvo un RR = 0,38 (IC95% 0,28-0,52). Lo siguiente que se hizo fue cambiar la categoría de referencia, es decir, calcular el efecto del placebo o de la ausencia de tratamiento respecto a la warfarina. Dicho efecto sería el inverso de 0,38, que corresponde a RR 2,63 (IC95% 1,92-3,57). El margen inferior del IC, 1,92, podría considerarse el dintel mínimo de no inferioridad, pero las agencias reguladoras fueron más exigentes y determinaron que para calcular el margen de no inferioridad había que considerar que la warfarina tuviera el 50% del efecto que realmente tiene, con lo que el margen de no inferioridad quedó fijado en 1,46<sup>8</sup>.

## CONSIDERACIONES METODOLÓGICAS

En los ensayos de no inferioridad es preciso tener en cuenta las siguientes consideraciones<sup>1,5,7</sup>:

- Deben existir datos de eficacia previos del tratamiento comparador utilizado y, en caso de que esos datos se hayan obtenido frente a placebo, asegurarse de que hayan sido de superioridad.
- Se debe elegir un margen de no inferioridad correcto.
- El diseño debe ser lo más similar posible a los estudios en los que el tratamiento comparador mostró la superioridad.
- Así como en los análisis de superioridad se recomienda el análisis por intención de tratar (AIT) (se deben considerar todos los participantes del estudio, independientemente de que los datos no sean completos, haya pérdidas en el seguimiento, etc.), en los ensayos de no-inferioridad puede ser preferible realizar un análisis por protocolo (APP), que incluye únicamente los pacientes que han completado el plan de tratamiento y han cumplido el protocolo, para asegurar que se ha realizado el ensayo de forma óptima con el objetivo de evidenciar las diferencias entre los tratamientos. Esto hace que se incluyan menos pacientes y que pueda no haber una representación real de la población. Por ello, las autoridades regulatorias recomiendan la realización de ambos análisis<sup>1,2,5</sup>. La no inferioridad debería concluirse solo si ambos tipos de análisis la apoyan<sup>5</sup>.

## INTERPRETACIÓN DE LOS RESULTADOS

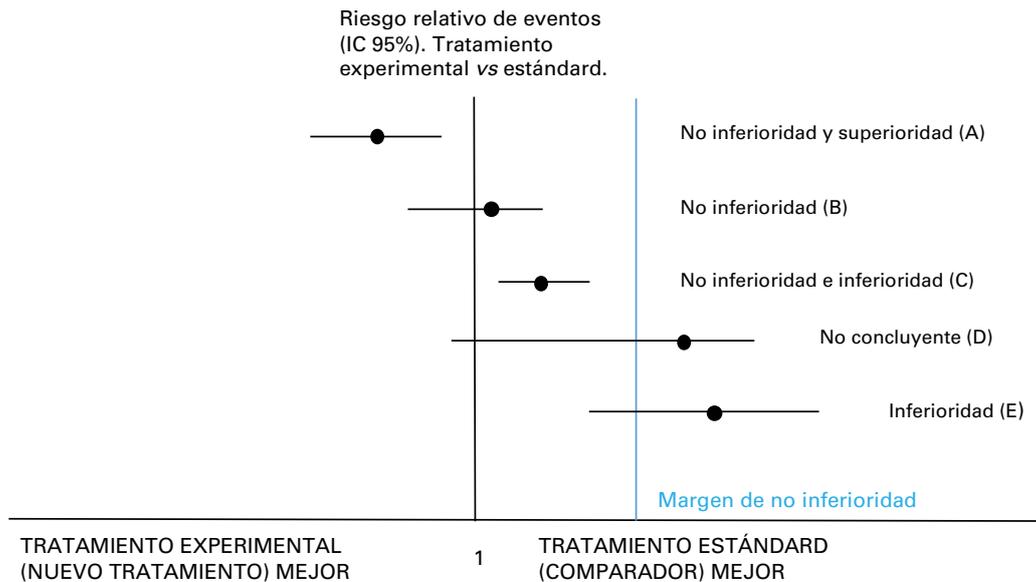
Para determinar si un tratamiento es no inferior a otro se utilizan los límites del intervalo de confianza (IC) del estimador central utilizado y se comparan con los valores inferior y superior del margen de no inferioridad preestablecido<sup>6</sup>. Si el IC excluye el margen de no inferioridad preespecificado, se concluye que el tratamiento en estudio es no inferior al tratamiento control<sup>6,7</sup>.

Hay que tener en cuenta la amplitud de esos IC. A veces son tan amplios, que lo que se considera no inferior desde un punto de vista estadístico, puede ser cuestionable desde un punto de vista clínico. Cuando la diferencia entre el tratamiento comparador y el placebo es pequeña, dependiendo del límite de no inferioridad, el efecto del fármaco supuestamente no inferior en realidad podría estar cerca del placebo<sup>7</sup>.

En los ensayos de no inferioridad, en función de los límites del IC, se pueden obtener cinco resultados diferentes (ver figura 1)<sup>1,7</sup>.

Una vez establecida la no inferioridad, se puede intentar determinar la superioridad del tratamiento experimental, para ello este análisis debe estar predefinido en el protocolo y se debe realizar en la población por intención de tratar<sup>1,6</sup>.

Figura 1. Posibles resultados de los ensayos de no inferioridad utilizando de una medida de riesgo relativo<sup>1</sup>



A: el tratamiento experimental es no inferior al tratamiento comparador porque el límite superior del IC no excede el margen de no inferioridad. Se demuestra también la superioridad desde un punto de vista estadístico porque todo el intervalo es menor que 1.

B: el tratamiento experimental es no inferior al tratamiento comparador porque el límite superior del IC no excede el margen de no inferioridad.

C: situación paradójica en la cual se demuestran la no inferioridad y la inferioridad dado que el límite superior del IC no excede el margen de no inferioridad, pero todo el intervalo es mayor que 1.

D: resultados no concluyentes porque el IC incluye el margen de no inferioridad y el 1. Suele sugerir una carencia de poder estadístico.

E: el tratamiento experimental no demuestra la no inferioridad respecto al tratamiento comparador porque el límite superior del IC excede el margen de no inferioridad. Además, es inferior porque todo el intervalo es mayor que 1.

En resumen, a la hora de evaluar un ensayo de no inferioridad se debe considerar lo siguiente (modificado de 5):

- El tratamiento experimental debe ofrecer ventajas en seguridad, coste o conveniencia frente al comparador.
- El tratamiento comparador debe ser el más adecuado y haber demostrado tener un efecto grande, fiable y consistente en ensayos clínicos controlados con placebo.
- El margen de no inferioridad debe de estar definido a priori y estar basado en criterios estadísticos y en el juicio clínico. Es preferible que esté expresado en una escala relativa de riesgo (riesgo relativo, odds ratio o hazard ratio) que en una diferencia absoluta del riesgo.
- La población y protocolo del ensayo de no inferioridad debería ser idéntica a la de los ensayos frente a placebo del comparador.
- El IC no debe exceder el margen de no inferioridad.
- La no inferioridad debe estar validada por el análisis por intención de tratar y por protocolo.
- Si se quiere declarar superioridad una vez que se ha demostrado la no inferioridad, este análisis debe estar preestablecido en el protocolo y estar demostrado a partir del análisis por intención de tratar.

## ANÁLISIS DE SUBGRUPOS

Cada vez son más habituales los ensayos clínicos, metanálisis y revisiones sistemáticas que analizan subgrupos de pacientes tratando de identificar las poblaciones específicas que pueden beneficiarse más (o menos) de un determinado tratamiento. Aunque los análisis de subgrupos pueden ayudar a tomar decisiones individuales sobre tratamientos, también pueden ser engañosos<sup>10</sup>.

Un ejemplo clásico que ilustra el riesgo de aceptar acríticamente los resultados de los análisis de subgrupos, es el del análisis de subgrupos del ensayo Second International Study of Infarct Survival (ISIS-2), estudio realizado para conocer si la aspirina reducía la mortalidad vascular en pacientes que habían sufrido un infarto

(ver tabla 2). Se observó una diferente reducción en la mortalidad cardiovascular dependiendo del signo del zodiaco de los pacientes: los géminis y libra no experimentaron una reducción de la mortalidad cardiovascular atribuible a la aspirina a diferencia de los pacientes de otros signos del zodiaco. Aunque la interacción alcanzó significación estadística ( $p = 0,003$ ), este resultado no tiene sentido, ya que no hay ninguna hipótesis plausible, basada en el conocimiento científico o biológico existente, que permita explicar este hallazgo<sup>10</sup>. La insensatez de mirar solamente la estadística.

Tabla 2. Análisis de Subgrupos del Segundo Estudio Internacional de Supervivencia al Infarto (ISIS-2)

	Nº/Total (%) de Pacientes		Riesgo Relativo (IC 95%)	Reducción o incremento del Riesgo Relativo
	Aspirina	Placebo		
Mortalidad vascular en todos los pacientes	804/8587 (9,4)	1016/8600 (11,8)	0,79 (0,73-0,87)	Reducción 20,7%
Géminis o Libra	150/1357 (11,1)	147/1442 (10,2)	1,08 (0,87-1,34)	Incremento 8,4%
Otros signos del zodiaco	654/7228 (9,0)	868/7187 (12,1)	0,75 (0,68-0,82)	Reducción 25,4%

Los análisis de subgrupos, por su naturaleza exploratoria, permiten generar hipótesis que siempre deben ser confirmadas mediante otros ensayos clínicos. La realidad es que rara vez se realizan estos ensayos (algunas veces de forma interesada) y, partiendo de esos resultados, se pueden tomar decisiones clínicas con alto grado de incertidumbre<sup>11</sup>. Así, hay ejemplos de análisis de subgrupos que han mostrado una diferencia clínicamente importante en el efecto de un tratamiento que posteriormente se ha comprobado que no es cierta como: que la aspirina no es efectiva en la prevención secundaria del ictus en mujeres, que el tratamiento antihipertensivo no es efectivo o puede ser perjudicial en personas ancianas, etc<sup>12</sup>.

Sin embargo, existen situaciones en las que los análisis por subgrupos no pueden ser ignorados. Este análisis es importante cuando existen diferencias potencialmente grandes: entre grupos en el riesgo de un mal resultado con o sin tratamiento, si existe heterogeneidad potencial del efecto del tratamiento en relación con la fisiopatología, si hay cuestiones prácticas sobre cuándo tratar, o si hay dudas sobre el beneficio en grupos específicos, como los ancianos, lo que conduce a un infratratamiento potencialmente inapropiado<sup>12</sup>.

Los análisis de subgrupos están sujetos a los mismos errores que cualquier test estadístico. La realización de múltiples pruebas estadísticas, una por cada característica que define el subgrupo (sexo, edad...), aumenta la posibilidad de encontrar diferencias estadísticamente significativas cuando realmente no las hay (error alfa). Esto puede ser debido al azar o a una distribución desequilibrada en el subgrupo de otras características. Por otra parte, la disminución del tamaño de muestra, al dividirla en subgrupos, aumenta la posibilidad de concluir que no existen diferencias cuando realmente sí las hay (error beta)<sup>11</sup>.

Para ayudar a diferenciar entre los que pueden aportar información relevante y los que pueden generar errores e inducir a tomar decisiones que potencialmente podrían perjudicar a los pacientes se han desarrollado diferentes guías<sup>10,11</sup>. El Grupo de Evaluación de Novedades, Estandarización e Investigación en Selección de medicamentos (GENESIS) de la Sociedad Española de Farmacia Hospitalaria, ha desarrollado un [check list](#) para intentar mejorar la realización e interpretación del análisis de subgrupos. Consta de dos partes: la primera contiene tres cuestiones preliminares para valorar si el análisis reúne las condiciones básicas para tener impacto en la toma de decisiones en la práctica clínica, y la segunda incluye siete criterios para valorar la aplicabilidad de la posible diferencia encontrada<sup>11</sup>.

## CUESTIONES PRELIMINARES<sup>11</sup>

### 1. Nivel de evidencia del estudio

1.1. El estudio sobre el que se realiza el análisis de subgrupos debe presentar el mayor nivel de evidencia y aplicabilidad disponible para justificar un posicionamiento terapéutico. En caso contrario, no se debería tener en cuenta el resultado del análisis de subgrupos, independientemente de que exista o no interacción estadística.

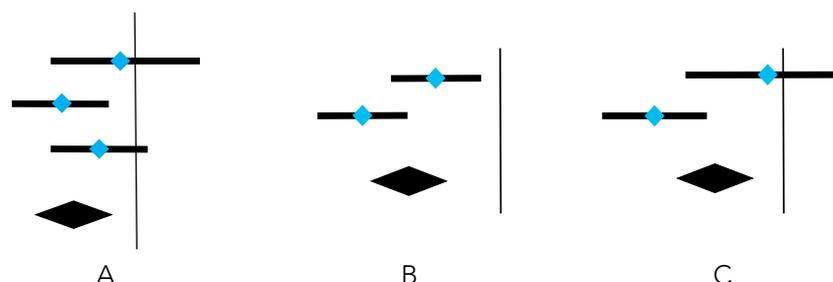
1.2. La variable de eficacia o seguridad considerada debe tener una clara relevancia clínica o, en su defecto, ser una variable subclínica o intermedia, medida como variable principal en el estudio. Un análisis de subgru-

pos en una variable de poca relevancia clínica carece de interés para la selección de pacientes o para la toma de decisiones.

## 2. Seleccionar los factores que determinan subgrupos con diferencias de efecto entre sí

Se puede hacer visualmente comprobando el grado de solapamiento (si es menor o mayor del 50%) entre los subgrupos en el **gráfico Forest plot**. Si el valor extremo de cada intervalo de confianza no llega al valor central del otro, significa que el solapamiento podría ser inferior al 50%, por lo que podría existir interacción y que podrían existir diferencias entre los subgrupos.

Ejemplos:



\*Figuras cedidas por Silvia Fénix (Farmacéutica del Hospital Puerto del Real, Cádiz)

- Los intervalos se solapan más de un 50%. No hay diferencias entre los subgrupos y coinciden con el resultado global del estudio.
- El solapamiento es menor del 50%. Hay diferencias entre los subgrupos, pero en ambos existe beneficio estadísticamente significativo y son consistentes con el resultado global del estudio, aunque el beneficio es más probable en un subgrupo que en otro. Esta situación podría ser interesante en escenarios donde el fármaco fuera muy tóxico o claramente ineficiente, y fuera necesario seleccionar la población de mayor beneficio.
- El solapamiento entre intervalos es menor del 50%, existe interacción, y uno de los subgrupos no presenta un resultado estadísticamente significativo. En esta situación, se observa que en uno de los subgrupos existe beneficio y en el otro no.

Esto que sugiere el análisis visual del Forest Plot, se debe comprobar estadísticamente calculando la *p* de interacción (significación estadística), que nos indica la probabilidad de que las diferencias entre los subgrupos se deban al azar. Una *p* de interacción significativa ( $< 0,1$  o menor de un valor razonable preespecificado por los autores) indica que ese factor debe ser valorado. Por desgracia, el 71,8% de las publicaciones no aportan la *p* de interacción<sup>11</sup>.

Para calcular la *p* de interacción existen diferentes calculadoras disponibles<sup>11</sup>:

- Para dos subgrupos: la calculadora de Joaquim Primo disponible en [https://gruposdetrabajo.sefh.es/genesis/genesis/Enlaces/Interaccion\\_2subgrupos\\_Joaquin\\_Primo.xls](https://gruposdetrabajo.sefh.es/genesis/genesis/Enlaces/Interaccion_2subgrupos_Joaquin_Primo.xls)
- Para más de 2 subgrupos, la «*p*» de heterogeneidad de metaanálisis

**3. Considerar la secuencia temporal**, asegurándose de que los pacientes presentaban el factor que establece los subgrupos antes de recibir la intervención.

## CRITERIOS PARA VALORAR LA POSIBLE DIFERENCIA ENCONTRADA ENTRE LOS SUBGRUPOS<sup>11</sup>

Una vez se encuentra una *p* de interacción significativa se pueden sospechar diferencias entre subgrupos. Pero para descartar que el hallazgo sea falso, se atenderá también a los otros criterios que apoyan una relación causal.

- Asociación estadística:** se comprueba por la *p* de interacción, que indica la probabilidad de que dichas diferencias se deban al azar. Una vez se encuentra una *p* de interacción significativa se pueden sospechar diferencias entre subgrupos. Se considera significativa una  $p < 0,05$ . Pero, como al dividir la población en subgrupos la potencia estadística baja, se puede considerar indicativo de diferencia una  $p < 0,1$ , o incluso  $p < 0,2$  en el caso de la regresión multivariante.

2. **Pre-especificación de los subgrupos:** el análisis de los subgrupos estudiados deberá haber sido preespecificado en el protocolo del ensayo. Hay que tener en cuenta que, para encontrar alguna diferencia, bastaría hacer un número suficiente de análisis. Un análisis *post-hoc* puede esconder lo que en el argot evaluador se llama «ir de pesca», sobre todo cuando se trata de salvar un estudio con resultado global fallido o irrelevante.
3. **La población (n) de los subgrupos:** una «n» demasiado reducida facilita la descompensación de las características basales en las ramas estudiadas.
4. **El número de factores estudiados:** a mayor número de factores estudiados, mayor es la probabilidad de encontrar una interacción debida al azar. Se debe disponer de un número reducido y justificado de factores pre-especificados analizados (< 10 factores).
5. **Resultado global del estudio:** cuando, según el diseño planificado del estudio, no se ha podido demostrar el beneficio, el análisis de subgrupos por sí solo tiene una fiabilidad limitada.
6. **Plausibilidad biológica:** las diferencias entre los subgrupos deben ser explicables mediante una hipótesis aceptable, acorde al conocimiento actual de la enfermedad.
7. **Consistencia:** existencia de estudios similares en los que se debería verificar la misma hipótesis, los resultados en los grupos de pacientes estudiados deben ser compatibles con el hallazgo que se analiza. La consistencia puede valorarse con otras variables del mismo estudio, otras indicaciones del mismo fármaco o con otros fármacos del mismo mecanismo de acción.

“ Si existe interacción y los criterios se cumplen de forma razonable, es posible considerar los resultados por subgrupos y valorar su aplicabilidad en la práctica clínica. ”

## CONCLUSIONES

Los análisis de subgrupos tratan de identificar las poblaciones específicas que pueden beneficiarse más (o menos) de un determinado tratamiento. Son exploratorios y permiten generar hipótesis que siempre deben ser confirmadas mediante otros ensayos clínicos. Para que los resultados de estos análisis sean relevantes, se deben cumplir las siguientes características:

- El análisis se debe restringir a un número limitado de subgrupos previamente definidos y justificados. En general, los análisis *post-hoc* deben ser tratados con escepticismo independientemente de su trascendencia
- Debe existir interacción estadísticamente significativa ( $p$  de interacción)
- Debe tener plausibilidad biológica
- Debe existir consistencia con los resultados de estudios similares

## AGRADECIMIENTOS

Agradecemos a Emilio Alegre y Silvia Fénix, farmacéuticos del Hospital Universitario de Puerto Real, Cádiz, la revisión del texto sobre «Análisis de Subgrupos», así como sus acertados comentarios y sugerencias

## BIBLIOGRAFÍA

1. Molina Nadal A. *Ensayos clínicos de no inferioridad: qué son y cuándo se deberían utilizar*. BIT. 2020;31(8):49-56.
2. European Medicines Agency. *Guideline on the choice of the non-inferiority margin*. Committee for medicinal products for human use (CHMP). London, 2005.
3. Garattini S, Bertele V. Non inferiority trials are unethical because they disregard patients' interest. *Lancet*. 2007;370:1875-7.
4. *Interpretando la literatura médica ¿qué necesito saber? Parte II*. INFAC.2006;14(8):23-6.
5. Kaul S. Understanding the merits and drawbacks of noninferiority trials in cardiovascular medicine. *Can J Cardiol*. 2021;37:1371-95. DOI: 10.1016/j.cjca.2021.05.017
6. Roumeliotis S, D'Arrigo G, Tripepi G. Methodological aspects of superiority, equivalence and non-inferiority trials. *Int Emer Med*. 2020;15:1085-91.
7. Mauri L, D'Agostino RB Sr. *Challenges in the Design and Interpretation of Noninferiority Trials*. *N Engl J Med*. 2017;377(14):1357-67.
8. Ferreira-González I. Bases para la interpretación de los estudios de no inferioridad: a propósito de los estudios ROCKET-AF, RE-LY y ARISTOTLE. *Rev Esp Cardiol*. 2014;67(6):432-5.

«El boletín INFAC es una publicación electrónica que se distribuye gratuitamente a las y los profesionales sanitarios de la CAPV. El objetivo de este boletín es la promoción del uso racional del medicamento para obtener un mejor estado de salud de la población».

- Husain M, Birkenfeld A, Donsmark M, Dungan K, Eliashewitz FG, Franco DR, Jeppesen OK, et al. *Oral Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes*. N Engl J Med. 2019;381:841-51. DOI: 10.1056/NEJMoa1901118
- Sun X, Ioannidis JPA, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. JAMA. 2014;311:405-11.
- Gil Sierra MD, Fénix Caballero S, Abdel kader Martín L, Fraga Fuentes MD, Sánchez Hidalgo M, Alarcón de La Lastra Romero C, Alegre Del Rey EJ. *Checklist for clinical applicability of subgroup analysis*. J Clin Pharm Ther.2020;45(3):530-8.
- Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications and interpretation. Lancet. 2005;365:176-86.

Fecha de revisión bibliográfica: septiembre 2022

Este boletín debe citarse: Metodología: ensayos de no inferioridad, análisis de subgrupos. INFAC. 2023;31(1):1-9.

Es de gran importancia que se notifiquen a la Unidad de Farmacovigilancia del País Vasco las sospechas de reacción adversa a los nuevos medicamentos. La notificación se puede realizar a través de OSABIDE, del formulario de notificación on line de la intranet de Osakidetza, rellenando la tarjeta amarilla o a través de la nueva web de la AEMPS: <https://www.notificaRAM.es>

Para consultas, sugerencias y aportaciones dirigirse a: al farmacéutico de atención primaria de su OSI o a CEVIME o al e-mail: [infac@osakidetza.eus](mailto:infac@osakidetza.eus)

**Consejo de Redacción:** Consejo de Redacción: Iñigo Aizpuru<sup>1</sup>, Miren Albizuri<sup>2</sup>, Iciar Alfonso<sup>3</sup>, María Armendáriz<sup>4</sup>, Sergio Barrondo<sup>5</sup>, Saioa Domingo<sup>4</sup>, Maitane Elola<sup>6</sup>, Arritxu Etxebarria<sup>4</sup>, Julia Fernández<sup>4</sup>, Isabel Fontán<sup>4</sup>, Leire Gil<sup>4</sup>, Ana Isabel Giménez<sup>4</sup>, Naroa Gómez<sup>4</sup>, Juan José Iglesias<sup>2</sup>, Josune Iribar<sup>4</sup>, Nekane Jaio<sup>4</sup>, Itxasne Lekue<sup>4</sup>, M.ª José López<sup>1</sup>, Javier Martínez<sup>4</sup>, Lorena Martínez<sup>7</sup>, Mónica Martínez<sup>4</sup>, Carmela Mozo<sup>4</sup>, Elena Olloquegi<sup>4</sup>, Elena Ruiz de Velasco<sup>4</sup>, Rita Sainz de Rozas<sup>4</sup>, Inés San José<sup>4</sup>, Elena Valverde<sup>4</sup>, Miren Zubillaga<sup>2</sup>.

1. Farmacéutico/a CEVIME; 2. Médico/a de familia; 3. Farmacóloga clínica; 4. Farmacéutico/a de atención primaria; 5. Profesor/a de farmacología (UPV); 6. Médica de medicina interna; 7. Pediatra de atención primaria.



<http://www.euskadi.eus/informacion/boletin-infac/web01-a2cevine/es/>



Queda totalmente prohibido el uso de este documento con fines promocionales

Eusko Jaurlaritzaren Argitalpen Zerbitzu Nagusia

Servicio Central de Publicaciones del Gobierno Vasco

