

**TEMA 4: DISTINTOS  
CONTRASTES EN  
ENSAYOS CLÍNICOS**

**1 – Introducción:** contrastes de hipótesis, p-valores e intervalos de confianza, significación estadística y clínicamente relevante

**2 – Ensayos Clínicos Aleatorizados:** definición de ECA, poblaciones de análisis

**3 – Tipos de diseño:** definición de hipótesis, posibles sesgos, interpretación de resultados

**4 – Intercambio de diseños**

**5- As Good as or Better Trials**

**6 - Conclusiones**

# 1-Introducción

## Concepto de Intervalo de Confianza

Un **intervalo de confianza** es un rango de valores, construido a partir de la muestra, donde con una gran seguridad, el valor verdadero de la población se encuentra en él.

Se dice que el intervalo (a,b) es un **intervalo de confianza** para P con un nivel de confianza  $1-\alpha$  ( $\alpha=0.05$ ,  $\alpha=0.01$ ) si,

$$P\{P \in (a,b)\} \geq 1 - \alpha \quad \alpha = 0.05, \alpha = 0.01$$

**“La prevalencia de SM según los criterios de la OMS ha sido 71.5% [I.C. 95%:68.6%-74.2%]”**

# 1-Introducción

## Problema de contraste de hipótesis

Se entiende por **contrastar una hipótesis** comprobar la veracidad o falsedad de cierta afirmación o hipótesis referida a alguna característica de la población.

### **Hipótesis nula: $H_0$**

Afirmación que se cuestiona. Se mantendrá como cierta a no ser que los datos evidencien lo contrario.

### **Hipótesis alternativa: $H_1$**

Afirmación que se aceptará como verdadera cuando se demuestre la falsedad de  $H_0$ .

**¿Cómo debe plantearse la hipótesis nula?**

**Debería recoger la afirmación que se pretende comprobar que es falsa.**

**¿Cómo debe plantearse la hipótesis alternativa?**

**Debería recoger la afirmación que se desea probar fuera de toda duda**

# 1-Introducción

---

## El valor $p$

---

- ✚ Se calcula una vez observados los datos de la muestra
- ✚ Es una medida de la discrepancia que presentan los datos observados con la hipótesis nula.

**¿Cómo medimos la discrepancia de los datos de nuestra muestra con la hipótesis?**

# El valor $p$

- ✚ Si los datos son muy discrepantes con la hipótesis nula, sería **poco probable** obtener otros más discrepantes

$p$ -valor  
pequeño  $< 0.05 = \alpha$

- ✚ Si los datos son poco discrepantes con la hipótesis nula, sería **bastante probable** obtener otros más discrepantes

$p$ -valor  
grande  $> 0.05 = \alpha$

**$p$ -valor: probabilidad de obtener el valor muestral u otro más discrepante o contradictorio con la  $H_0$ , si  $H_0$  fuera cierta**

Planteamiento de una hipótesis  $H_0$  (y de su alternativa  $H_1$ )

Contesta el p-valor

¿Hay suficiente evidencia en contra de esta hipótesis?

$p\text{-valor} > 0.05$

Continuamos con la hipótesis

$p\text{-valor} < 0.05$

La hipótesis alternativa queda probada

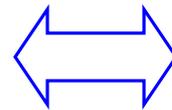
# Relación entre el valor $p$ y los intervalos de confianza

Existe una dualidad entre los contrastes bilaterales y los intervalos de confianza

$H_0: \mu_1 = \mu_2$     $H_1: \mu_1 \neq \mu_2$

Intervalo de confianza  $(1-\alpha)$   
para  $\mu_2 - \mu_1 : (a, b)$

La hipótesis nula se rechaza con un  $p < \alpha$



El valor  $\theta$  no pertenece al intervalo de confianza  $(a, b)$

# Problema del $p$ -valor: depende del tamaño de la muestra

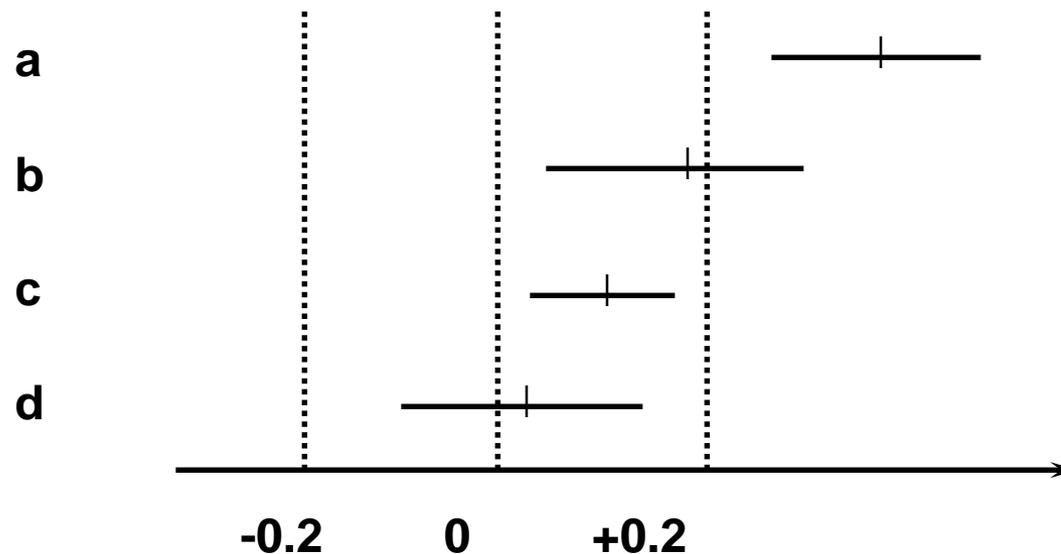
- ✚ Si hay pocos datos, podría ocurrir que no se detectaran diferencias significativas ( $p > 0.05$ ) aunque éstas existan (alto valor de  $\beta$ )
- ✚ Si hay demasiados datos, podría ocurrir que se detecten diferencias significativas ( $p < 0.05$ ) cuando realmente no son relevantes clínicamente

Importante la elección de un tamaño muestral adecuado

La valoración acerca de si una diferencia es o no clínicamente relevante debe hacerse independientemente del valor de  $p$ , utilizando intervalos de confianza

# Significación estadística y significación clínicamente relevante

**Diferencia clínicamente relevante: 20%**



**Intervalos de confianza al 95% para la diferencia entre placebo y tratamiento activo**

- a – Diferencia significativa ( $p < 0.05$ ) y clínicamente relevante
- b – Diferencia significativa ( $p < 0.05$ ) pero no clínicamente concluyente
- c – Diferencia significativa ( $p < 0.05$ ) pero no clínicamente relevante
- d – Diferencia no significativa ( $p > 0.05$ ) ni clínicamente relevante

# 2-Ensayos Clínicos Aleatorizados

## Tipos de estudios

¿Hay intervención?

No

Sí

Estudio observacional

Estudio experimental

-Administración fármaco,  
Momento de administración,  
duración.

-Aleatorización en dos o más  
grupos (Siempre un grupo  
Control).

# 2-Ensayos Clínicos Aleatorizados

**Estudios experimentales con humanos**



**ECA: Ensayo Clínico Aleatorizado**

- Administración fármaco, momento de administración, duración.
- Aleatorización en dos o más grupos (Siempre un grupo Control).
- Enmascaramiento: simple ciego, doble ciego, triple ciego.
- Código ético.
- Fases de un ECA.

# ENMASCARAMIENTO

Proceso por el que se oculta el tipo de tratamiento que el participante recibe en el ensayo

---

## Enmascaramiento

---

	Abierto	Simple ciego	Doble ciego	Triple ciego
Paciente	●	●	●	●
Investigador	●	●	●	●
Analista de datos	●	●	●	●

---

# FASES DE UN EC:

## - ENSAYOS CLÍNICOS FASE I:

- Sujetos: Voluntarios sanos.
- Objetivo:Comprobar la seguridad y tolerabilidad del compuesto.

## - ENSAYOS CLÍNICOS FASE II:

- Sujetos: Pacientes con la patología a estudio
- Objetivo: Búsqueda la dosis terapéutica, eficacia y seguridad en el tratamiento de pacientes con la patología de estudio. Completar datos de seguridad

## - ENSAYOS CLÍNICOS FASE III:

- Sujetos: Pacientes con la patología a estudio.
- Objetivo: Seguridad y eficacia clínica en gran número de pacientes y con la dosis terapéutica bien definida.Nuevas Indicaciones.

## - ENSAYO CLÍNICO FASE IV:

- Sujetos: Muchos pacientes (Condiciones reales de uso)
- Objetivo: Seguridad en condiciones reales de uso en poblaciones con gran número de sujetos. Nuevas indicaciones, presentaciones.

# **Ensayo aleatorio sobre basiliximab frente a placebo en el control del rechazo celular agudo en los receptores de un aloinjerto renal**

Nashan et al. Lancet Vol 22, nº1, 1998.

**Ensayo clínico fase III.**

**Aleatorizado**

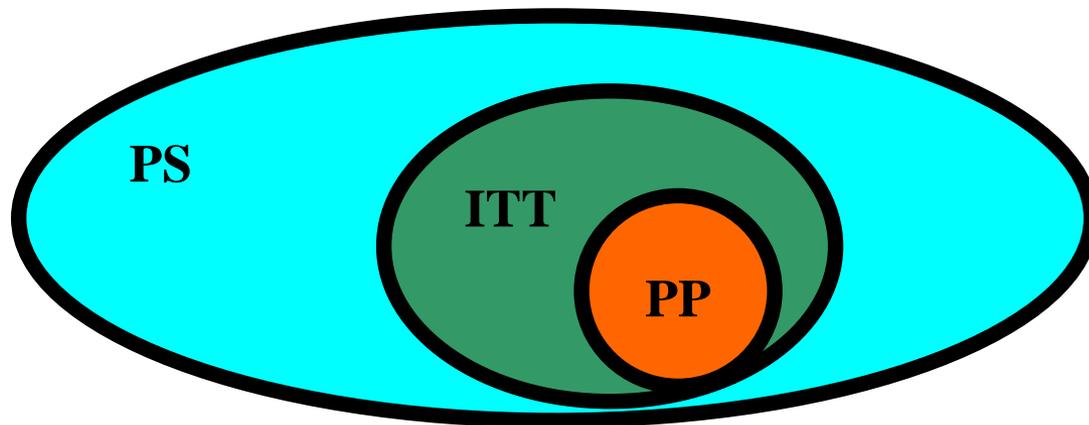
**Doble ciego**

**Controlado con placebo**

**Multicéntrico: con 21 centros europeos/canadienses.**

# Poblaciones de análisis en un ECA

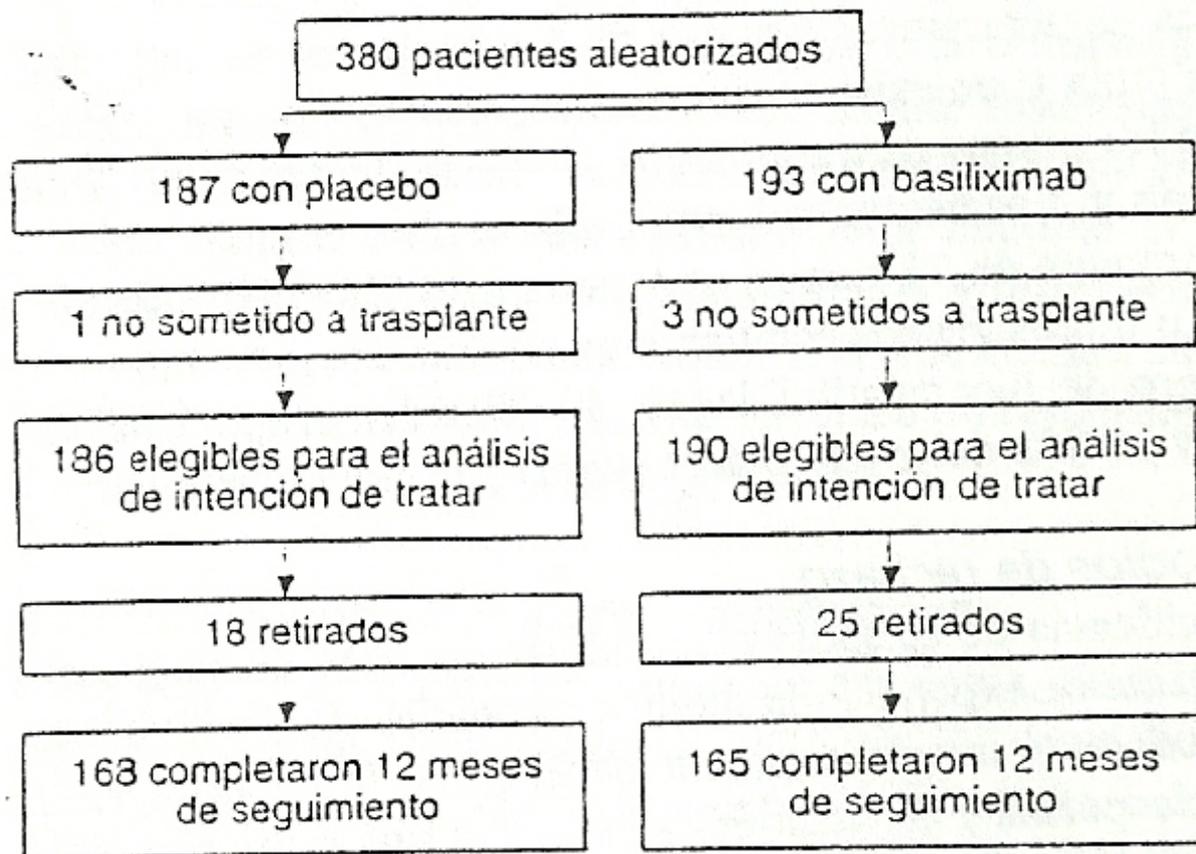
- **Población de seguridad (PS)**: pacientes asignados a un tratamiento y que al menos lo reciben una vez
- **Población por Intención de Tratar (ITT)**: pacientes de la población de seguridad y de los cuales se tiene al menos una medida de eficacia
- **Población por Protocolo (PP)**: pacientes de la ITT que no presentan violaciones del protocolo.



Abuso lenguaje:  
**Muestra**  
**Población**

Se definieron dos poblaciones para el análisis: una población de pacientes según intención de tratar que fueron asignados aleatoriamente al tratamiento, recibieron al menos una dosis de la medicación del estudio y se les realizó el trasplante, y toda la población tratada de pacientes que fueron aleatorizados y recibieron al menos una dosis de la medicación del estudio (incluyendo los que no fueron sometidos al trasplante).

Se realizaron los análisis de supervivencia principalmente con



# 3. Tipos de diseño según comparación

---

## ✚ Superioridad

✓ Demostrar diferencias significativas a favor de el tratamiento experimental o a favor del tratamiento control

## ✚ No inferioridad

✓ Demostrar que el tratamiento experimental no es menos efectivo que el control

## ✚ Equivalencia

✓ Demostrar la ausencia de diferencias relevantes entre tratamientos

# Ensayo aleatorio sobre basiliximab frente a placebo en el control del rechazo celular agudo en los receptores de un aloinjerto renal

Nashan et al. Lancet Vol 22, nº1, 1998.

## **Objetivo principal:**

*Valorar los efectos del tratamiento profiláctico con 40mg de baxilisimab sobre la incidencia de rechazo agudo en los primeros 6 meses después del trasplanrte*

**Variable principal:** experimentar o no al menos un episodio de rechazo agudo en los primeros 6 meses después del trasplante.

## **Objetivos secundarios:**

*Seguridad y tolerancia, Efectos adversos*

**Variables secundarias:** Gravedad histológica del rechazo, incidencia de episodios de rechazo agudo confirmados por biopsia y resistentes a los corticoides 6 meses después del trasplante, pérdida del injerto. Muertes y complicaciones infecciosas relacionadas con el fármaco del estudio. Efectos adversos

# 3.1 Estudios de superioridad

---

**Demostrar diferencias significativas a favor del tratamiento experimental o a favor del tratamiento control**

**Hipótesis nula:  $H_0$**

No hay diferencia entre  
tratamientos

**Hipótesis alternativa:  $H_1$**

Un tratamiento es mejor  
que el otro

# **Estudio: “Eficacia de Olanzapina frente a Risperidona”**

Objetivo: Evaluar la efectividad de la Olanzapina en los síntomas negativos de esquizofrenia, en comparación con la Risperidona, tras un año de tratamiento

## **Hipótesis nula: $H_0$**

La eficacia de la Olanzapina es la misma que la de la Risperidona

## **Hipótesis alternativa: $H_1$**

La eficacia de la Olanzapina no es la misma que la de la Risperidona

¿Qué variable se va a utilizar para valorar la eficacia?

Variable principal X: el cambio basal-final en la puntuación de la escala de evaluación de síntomas negativos (SANS).

Variable numérica

**Hipótesis nula:  $H_0$**

$$\mu_{\text{olanzapina}} = \mu_{\text{risperidona}}$$

**Hipótesis alternativa:  $H_1$**

$$\mu_{\text{olanzapina}} \neq \mu_{\text{risperidona}}$$

$\mu_{\text{olanzapina}}$ : El cambio medio basal-final en la escala SANS en los pacientes con Olanzapina

$\mu_{\text{risperidona}}$ : El cambio medio basal-final en la escala SANS en los pacientes con Risperidona

**H<sub>0</sub>** :No hay diferencia  
entre tratamientos

**H<sub>1</sub>** :Un tratamiento es  
mejor que el otro

¿Hay suficiente evidencia para probar que un  
tratamiento es mejor que otro?

**p-valor > 0.05**

**0 ∈ Int. confianza**

**p-valor < 0.05**

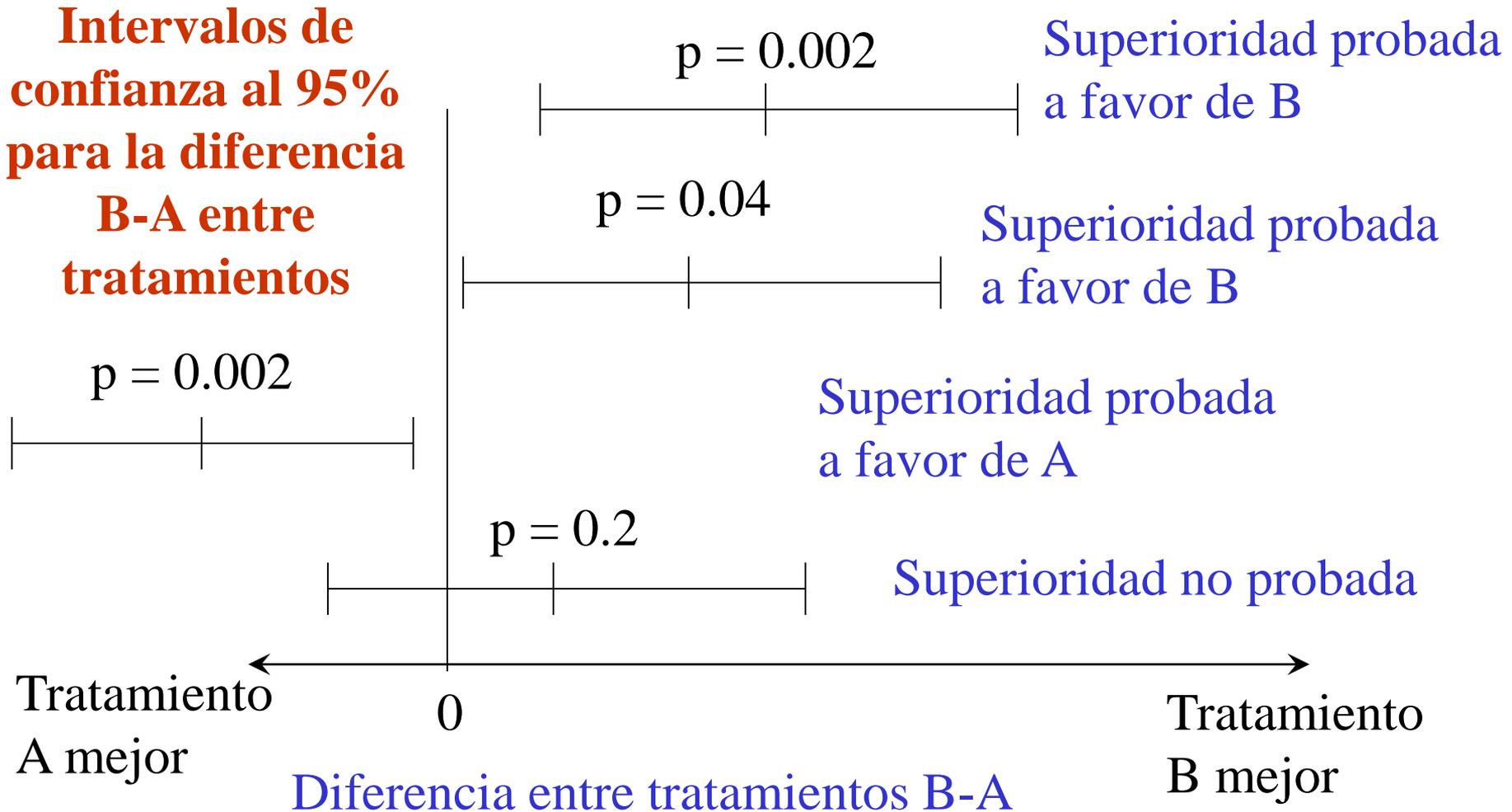
**0 ∉ Int. confianza**

No hay suficiente  
evidencia para probar  
diferencias  
significativas entre  
tratamientos.

Existen diferencias  
estadísticamente  
significativas entre  
tratamientos

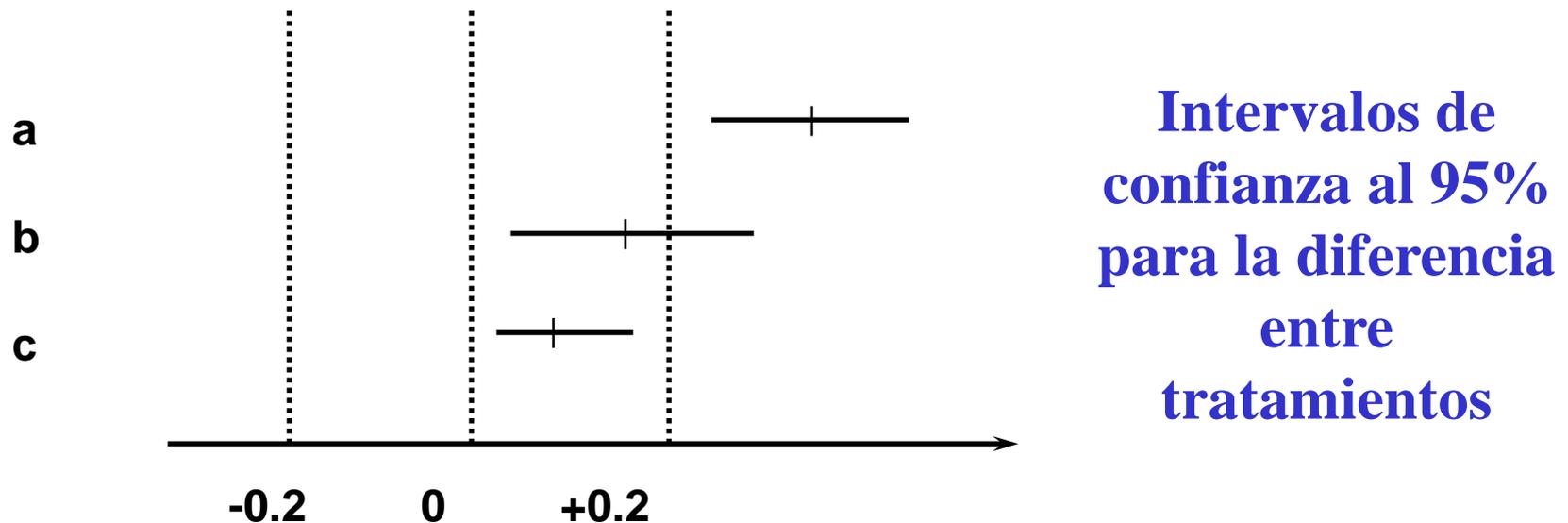
# Estudios de superioridad

**Intervalos de confianza al 95% para la diferencia B-A entre tratamientos**



# $p < 0.05$ . ¿Es la significación estadística, clínicamente relevante?

**Diferencia clínicamente relevante: 20%**



- a – Diferencia significativa ( $p < 0.05$ ) y clínicamente relevante
- b – Diferencia significativa ( $p < 0.05$ ) pero no clínicamente concluyente
- c – Diferencia significativa ( $p < 0.05$ ) pero no clínicamente relevante

# **$p > 0.05$ . ¿Por qué no se ha probado que existan diferencias entre tratamientos?**

---

1. Porque realmente no sean distintos los tratamientos
2. Porque el diseño no consigue detectar diferencias entre tratamientos, aunque las haya, es decir el diseño tiene poca **SENSIBILIDAD** (Capacidad de detectar que un tratamiento sea mejor que otro cuando realmente lo es). Importancia de una elección adecuada del tamaño muestral

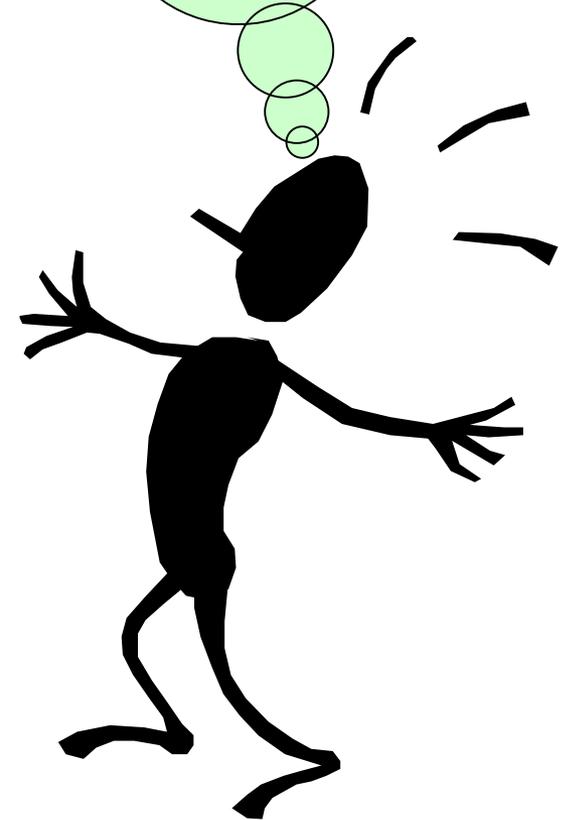
**Y, si no existen diferencias significativas...¿he demostrado que es equivalente o no inferior?**

**!!Rotundamente NO!!**



## **CONCLUSIÓN**

**El no demostrar diferencias significativas en un estudio de superioridad, no implica demostrar equivalencia o no inferioridad**



## 3.2 Estudios de no inferioridad

**Demostrar que el tratamiento experimental no es menos efectivo que uno ya existente (control) (puede ser más efectivo o tener efectos similares)**

**Hipótesis nula:  $H_0$**

El tratamiento B es inferior que el control A

**Hipótesis alternativa:  $H_1$**

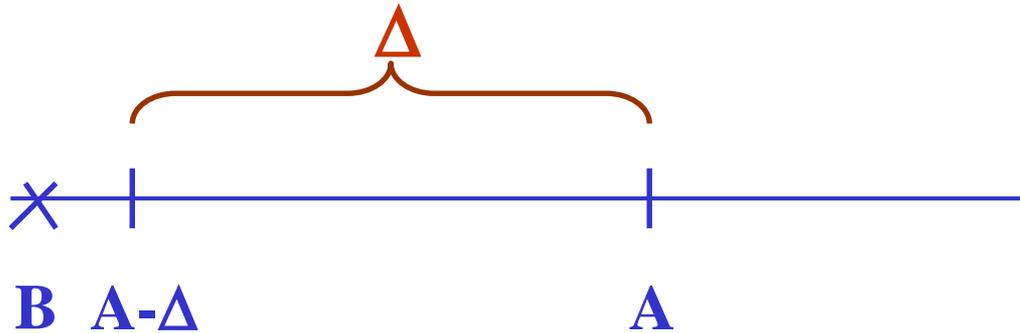
El tratamiento B no es inferior que el control A

**Pero...¿cuánto de inferior?**



**A: efecto medio del tratamiento Control**

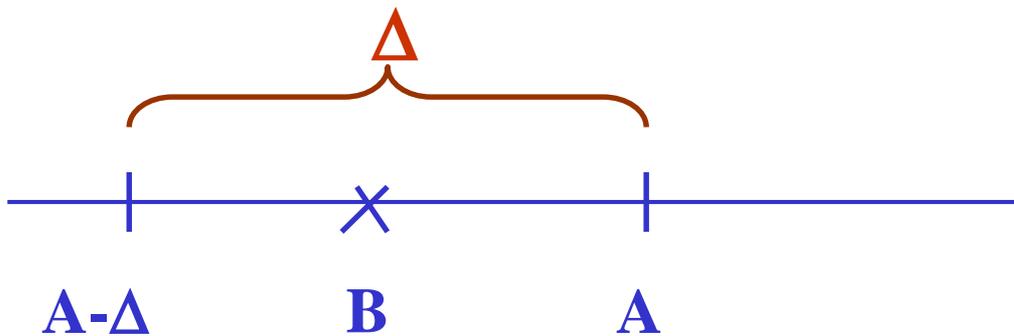
**B: efecto medio del nuevo Tratamiento**



**H<sub>0</sub>: Nuevo  
Tratamiento  
Inferior al Control**

$$\mathbf{H_0: B \leq A - \Delta ;}$$

$$\mathbf{H_0: B - A \leq -\Delta}$$



**H<sub>1</sub>: Nuevo  
Tratamiento No  
Inferior al Control**

$$\mathbf{H_1: B > A - \Delta ;}$$

$$\mathbf{H_1: B - A > -\Delta}$$

## **Estudio : “Duloxetine Versus Placebo and Paroxetine in the Treatment of Major Depression”**

Objetivo: Evaluar la no inferioridad de Duloxetina 60 mg BID en los síntomas depresivos a través del HAMD17, en comparación con Paroxetina 20 mg/día, tras nueve semanas de tratamiento

### **Hipótesis nula: $H_0$**

La eficacia de Duloxetina es inferior que la de Paroxetina

### **Hipótesis alternativa: $H_1$**

La eficacia de Duloxetina es igual o mejor (no inferior) que la de Paroxetina

**Margen de no inferioridad ( $\Delta$ ) = -2.2**

**Tratamiento control activo (A): Paroxetina**

**Tratamiento experimental (B): Duloxetina**

**Hipótesis nula:  $H_0$**

$$\mu_{\text{Duloxetina}} - \mu_{\text{Paroxetina}} \leq -2.2$$

**Hipótesis alternativa:  $H_1$**

$$\mu_{\text{Duloxetina}} - \mu_{\text{Paroxetina}} > -2.2$$

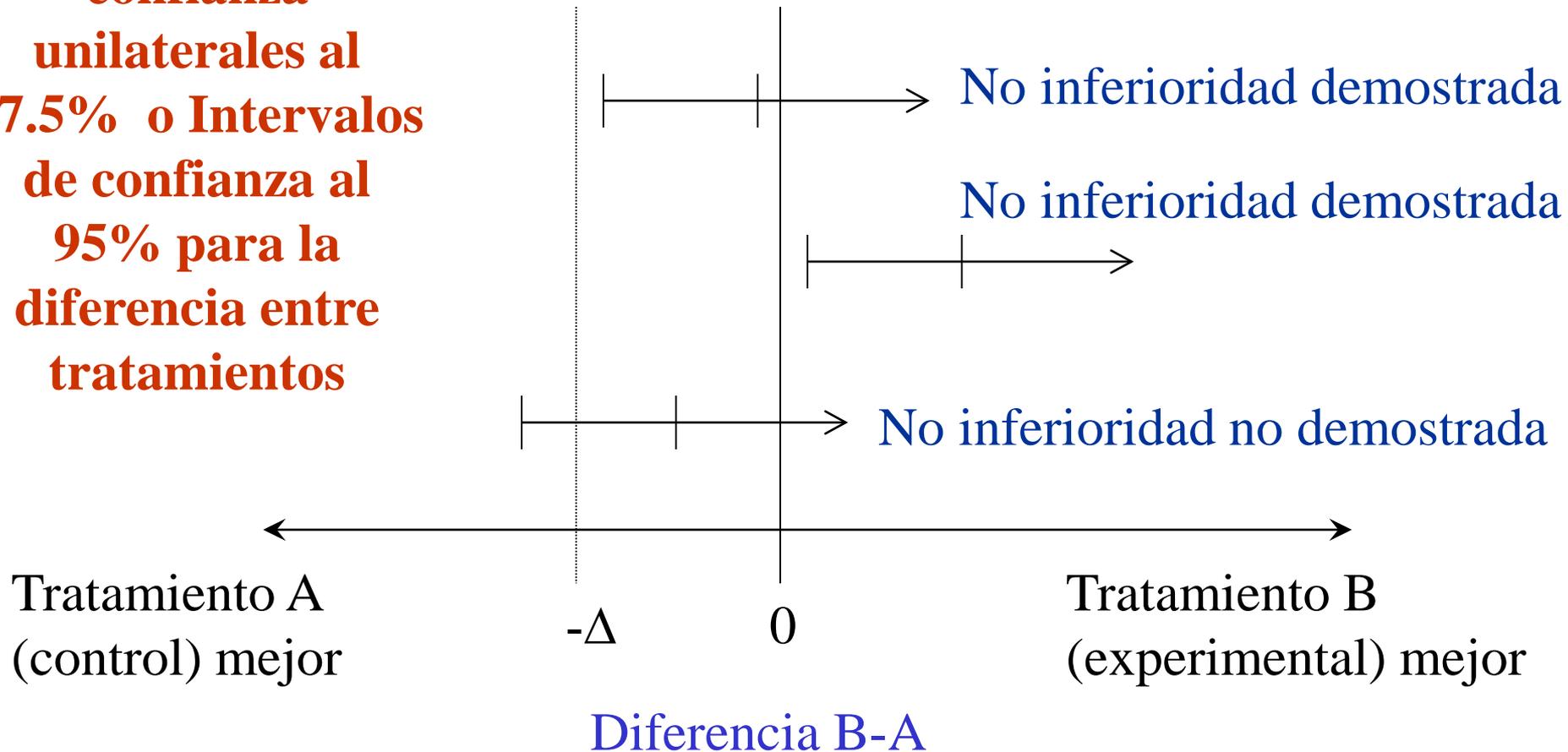
$\mu_{\text{Duloxetina}}$ : El cambio medio basal-final en la escala  
HAMD17 en los pacientes con Duloxetina

$\mu_{\text{paroxetina}}$ : El cambio medio basal-final en la escala  
HAMD17 en los pacientes con Paroxetina

**SE RESUELVE UTILIZANDO INTERVALO DE  
CONFIANZA PARA LA DIFERENCIA DE  
TRATAMIENTO B-A**

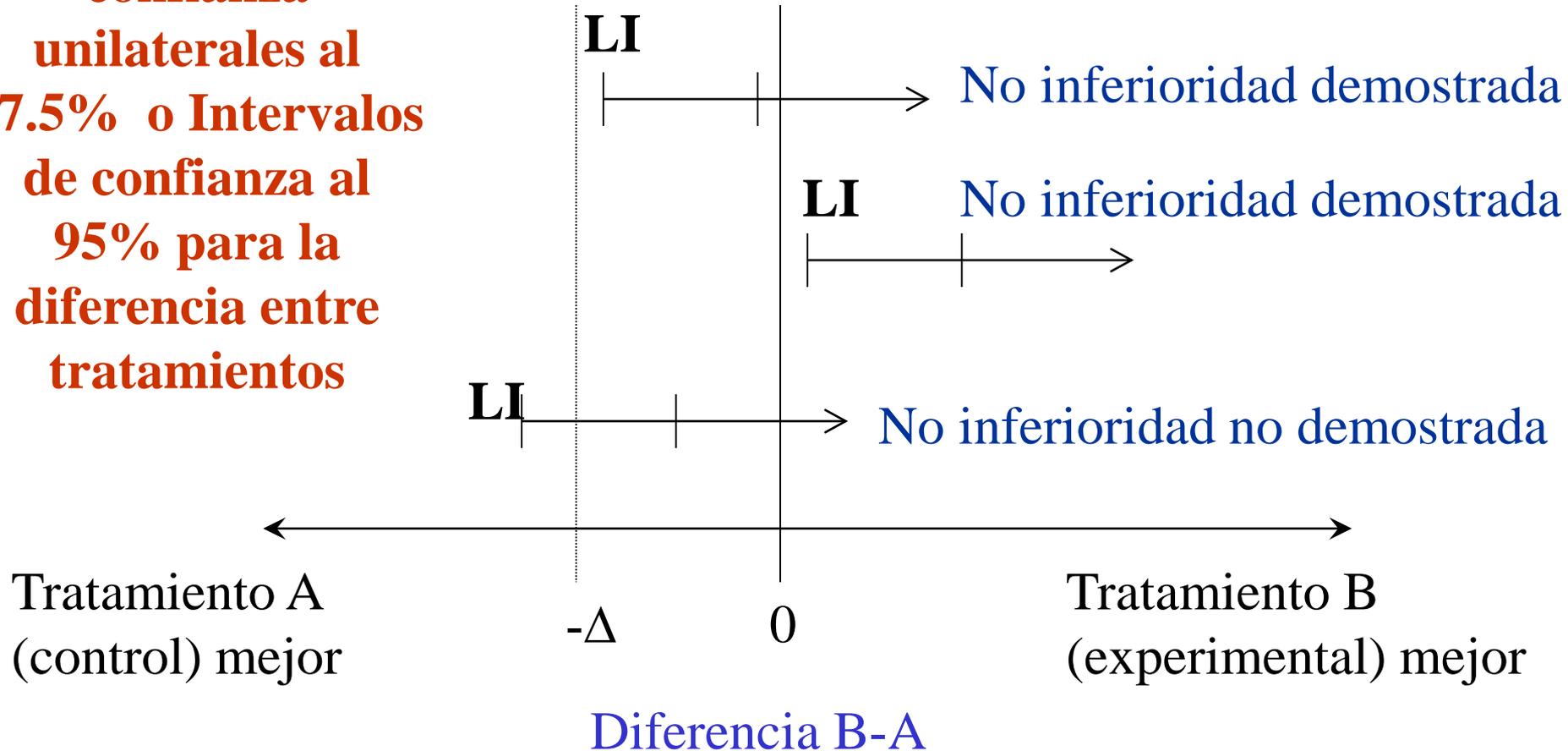
# Estudios de no inferioridad

**Intervalos de confianza unilaterales al 97.5% o Intervalos de confianza al 95% para la diferencia entre tratamientos**



# Estudios de no inferioridad

**Intervalos de confianza unilaterales al 97.5% o Intervalos de confianza al 95% para la diferencia entre tratamientos**



**$-\Delta < LI$ . ¿Se puede concluir que B es no inferior al control A?**

---

**Todavía NO**

Es necesario comprobar que el diseño tiene suficiente capacidad de detectar que B es inferior al control A, si realmente lo es  $\Rightarrow$

**SENSIBILIDAD** del diseño.

**¿En qué consiste la sensibilidad?**

# **$-\Delta < LI$ . ¿Se puede concluir que B es no inferior al control A?**

---

## **¿En qué consiste la sensibilidad?**

En demostrar que el diseño ha detectado que el control activo es realmente eficaz

## **¿Cómo?**

- Comprobando que el efecto del control obtenido en el estudio es similar al obtenido en experiencias históricas de estudios parecidos
- Introduciendo placebo, si es posible (tres ramas)

# **- $\Delta > LI$ . ¿Por qué no se ha probado que B no es inferior al control A?**

---

- 1. Porque realmente el tratamiento B es inferior al tratamiento control A**
- 2. Porque, a pesar de ser B no inferior a A, no tenemos suficientes datos para poderlo demostrar. Importancia de la elección del tamaño muestral.**

## 3.3 Estudios de Equivalencia

Se diseña para demostrar que no existe diferencia entre dos tratamientos, esto es, la equivalencia clínica entre tratamientos.

### Hipótesis nula: $H_0$

Un tratamiento es mejor que el otro (no son equivalentes)

### Hipótesis alternativa: $H_1$

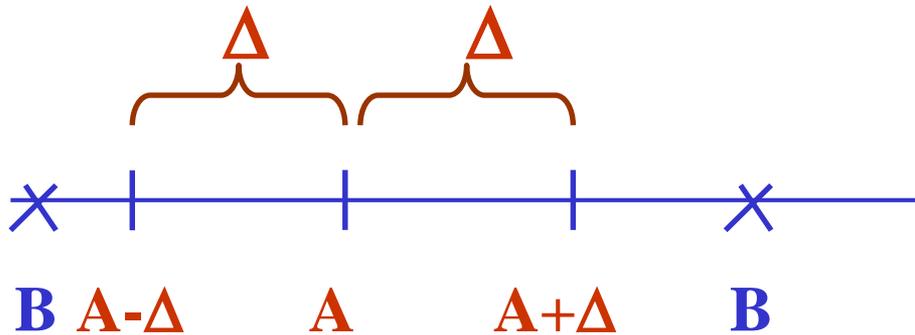
No hay diferencia clínica entre tratamientos (son equivalentes)

Pero...¿cuánto de equivalentes?

$+\Delta$   $0$   $-\Delta$

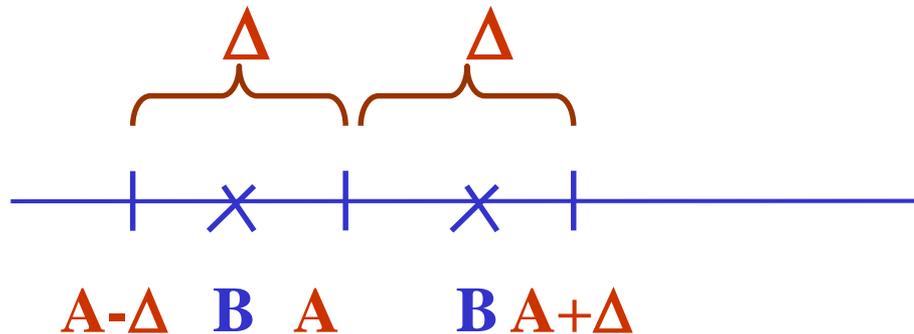
**A: efecto medio del tratamiento Control**

**B: efecto medio del nuevo Tratamiento**



**H<sub>0</sub>: Nuevo Tratamiento no es equivalente al Control**

$$H_0: |B-A| \geq \Delta$$

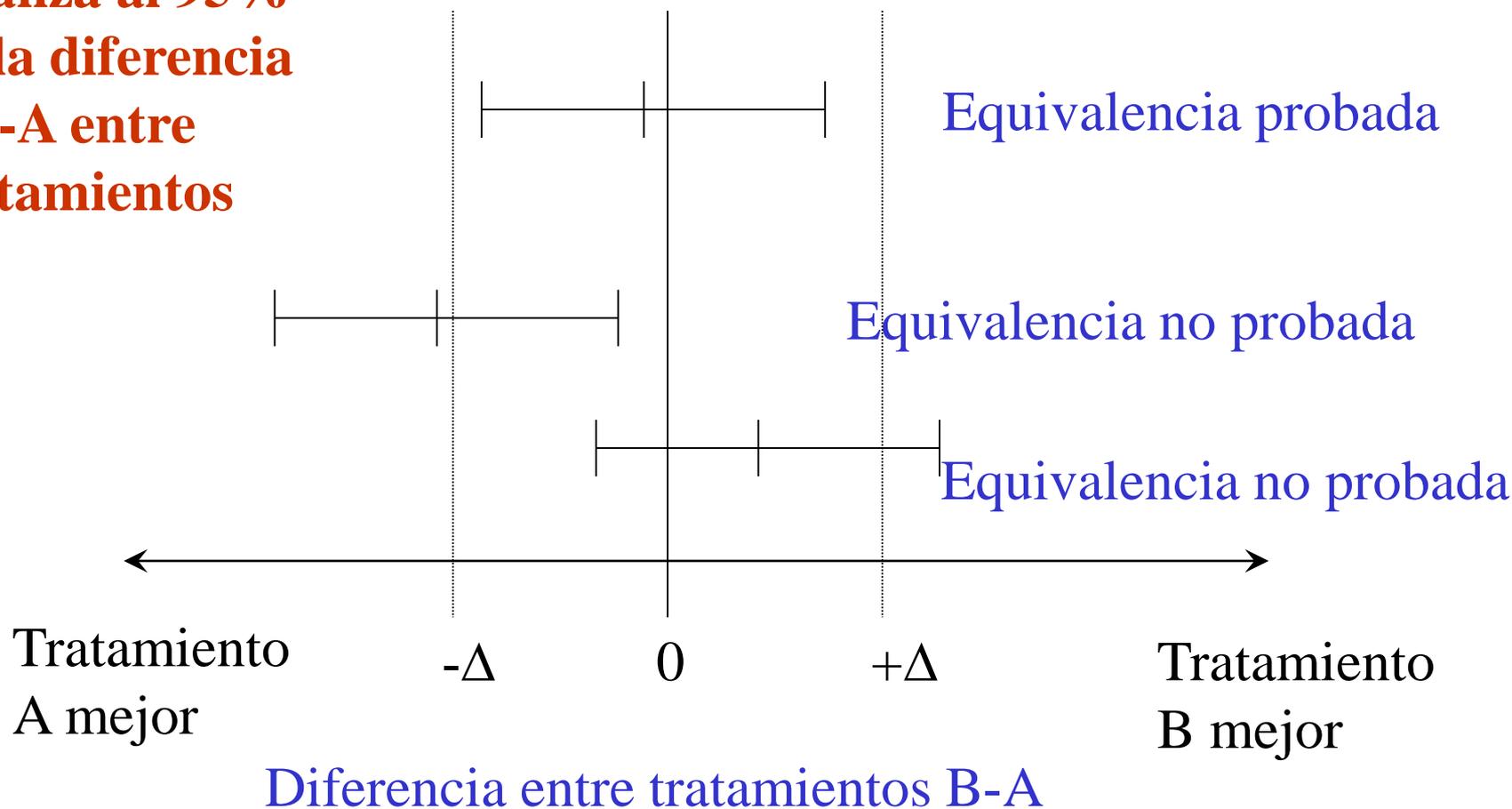


**H<sub>1</sub>: Nuevo Tratamiento es equivalente al Control**

$$H_1: |B-A| < \Delta$$

# Estudios de Equivalencia

**Intervalos de confianza al 95% para la diferencia B-A entre tratamientos**



**En el caso de concluir equivalencia**  
sería necesario, de igual manera que en  
los estudios de no inferioridad, el  
análisis de la **sensibilidad** del estudio

**En el caso de no concluir equivalencia,**  
si el tamaño muestral es adecuado,  
queda probado que no son equivalentes

# Poblaciones de análisis

## ¿Sobre qué población se realiza cada diseño?

- Población por Intención de Tratar (ITT): pacientes asignados a un tratamiento y de los cuales se tiene al menos una medida de eficacia
- Población por Protocolo (PP): pacientes de la ITT que no presentan violaciones del protocolo.

## ¿Qué población es más conservativa para cada contraste?

**Superioridad**



**ITT**

**No-inferioridad, Equivalencia**



**ITT-PP**

# 4. Intercambio de diseños

---

- ✓ Interpretar un estudio de no-inferioridad como uno de superioridad
- ✓ Interpretar un estudio de superioridad como uno de no-inferioridad
- ✓ Los diseños de equivalencia por su objetivo no permiten cambios a otros diseños, aunque es posible cambiar los márgenes de equivalencia

## 4.1 No inferioridad $\Rightarrow$ Superioridad

---

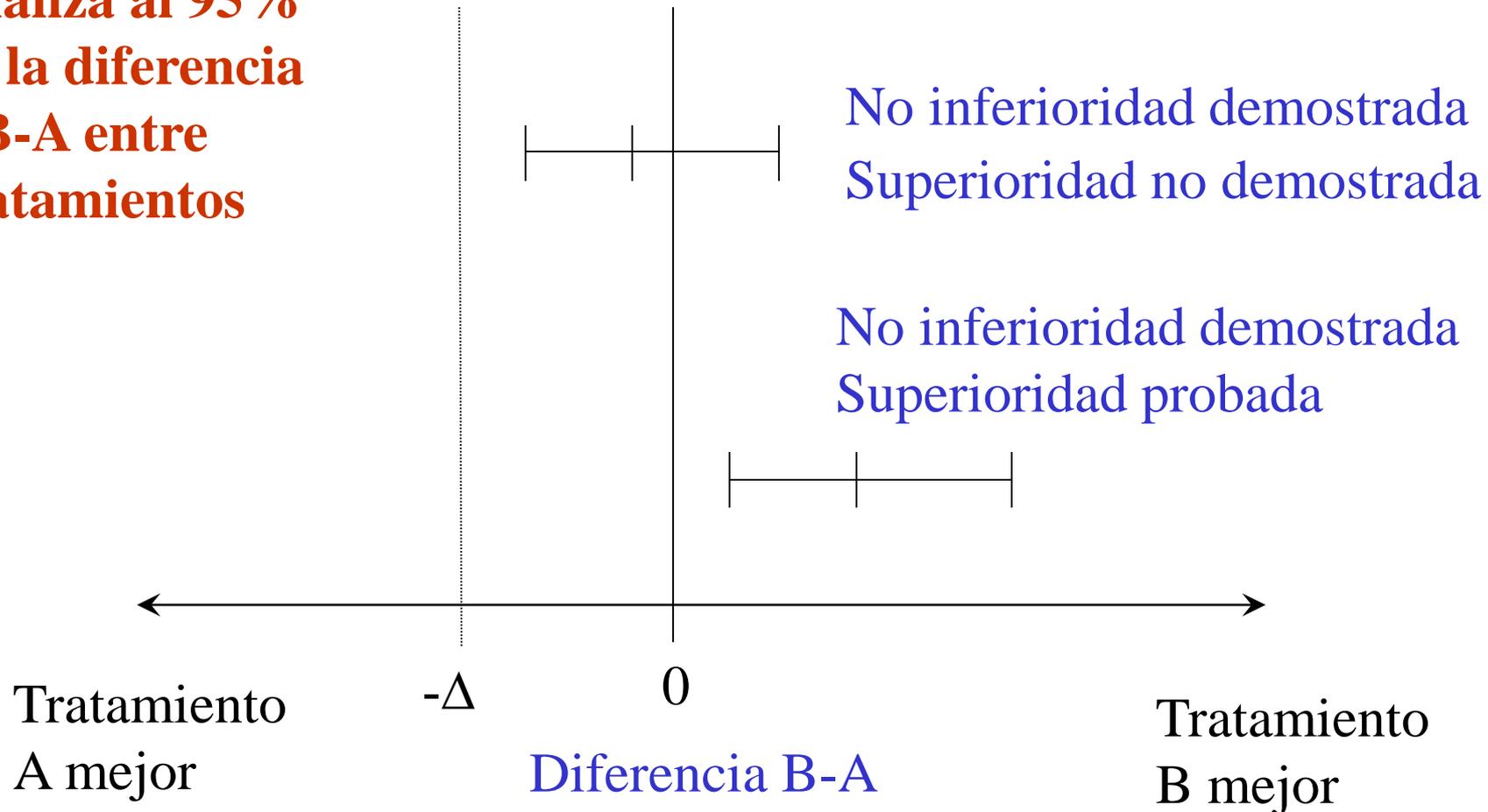
Si en un estudio de no inferioridad, se llega a que **B *no es inferior a* A...**

**¿Se puede concluir además que B es superior a A?**

**INTERVALOS DE CONFIANZA**

# 4.1 No inferioridad => Superioridad

Intervalos de confianza al 95% para la diferencia B-A entre tratamientos



# 4.1 No inferioridad => Superioridad

---

## OBSERVACIONES

- El estudio ha sido diseñado y realizado en las condiciones de un estudio de no-inferioridad
- p-valor para superioridad es calculado para obtener una valoración de la evidencia
- Análisis sobre ITT

## 4.2 Superioridad $\Rightarrow$ No inferioridad

---

Si en un estudio de superioridad, se llega a que **B *no es superior a A...***

**¿Se puede concluir por lo menos que B no es inferior a A?**

**INTERVALOS DE CONFIANZA**

## 4.2 Superioridad => No inferioridad

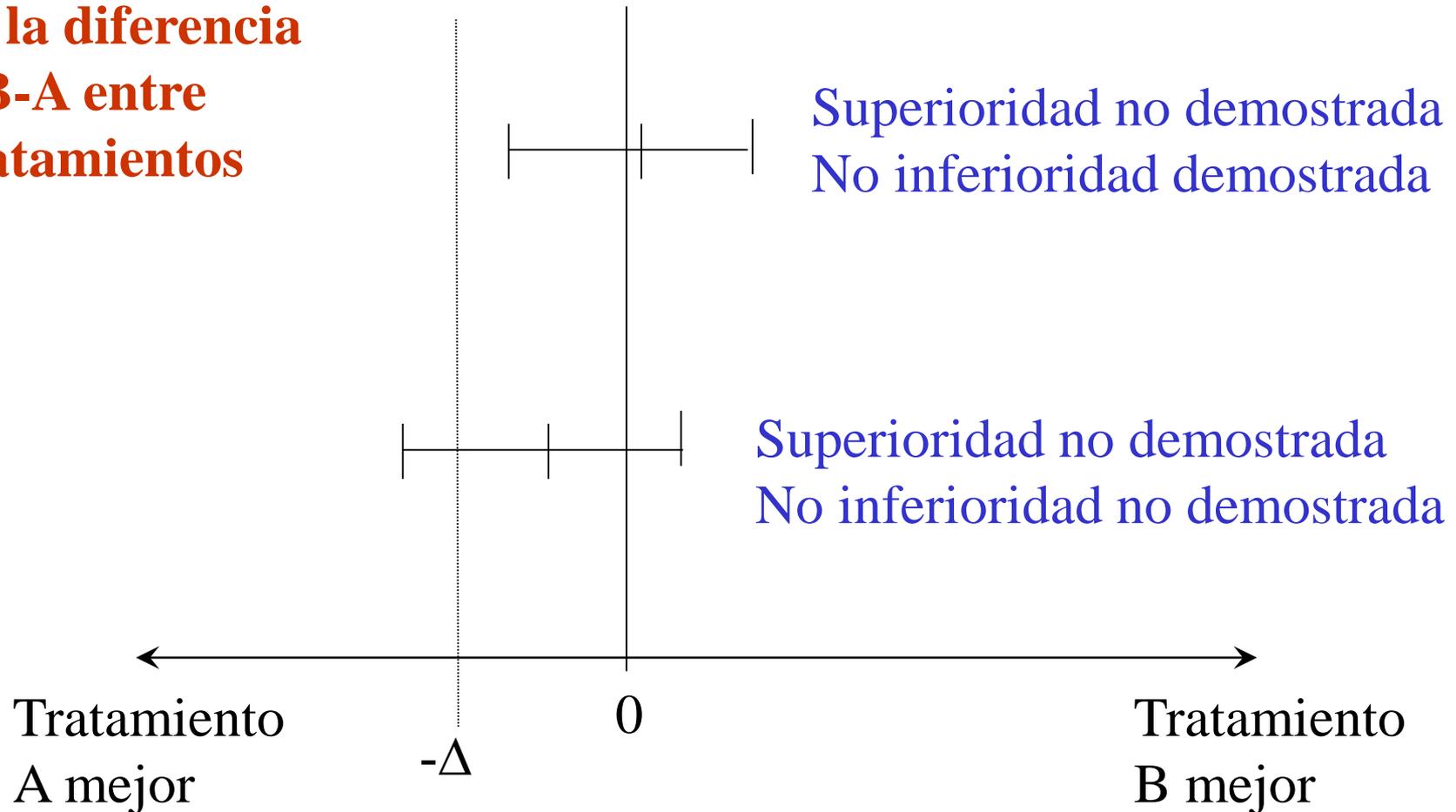
---

**Pero antes hay que tener en cuenta que...**

- El  $\Delta$  está predefinido en protocolo.
- Análisis sobre ITT y PP proporcionan IC y p-valores similares para la hipótesis de no inferioridad.
- Evidencia de que el control (A) tiene demostrada su eficacia.
- El estudio ha sido diseñado y realizado en las condiciones de un estudio de no-inferioridad.

# 4.2 Superioridad => No inferioridad

Intervalos de  
confianza al 95%  
para la diferencia  
B-A entre  
tratamientos



## 4.3 Márgenes de equivalencia

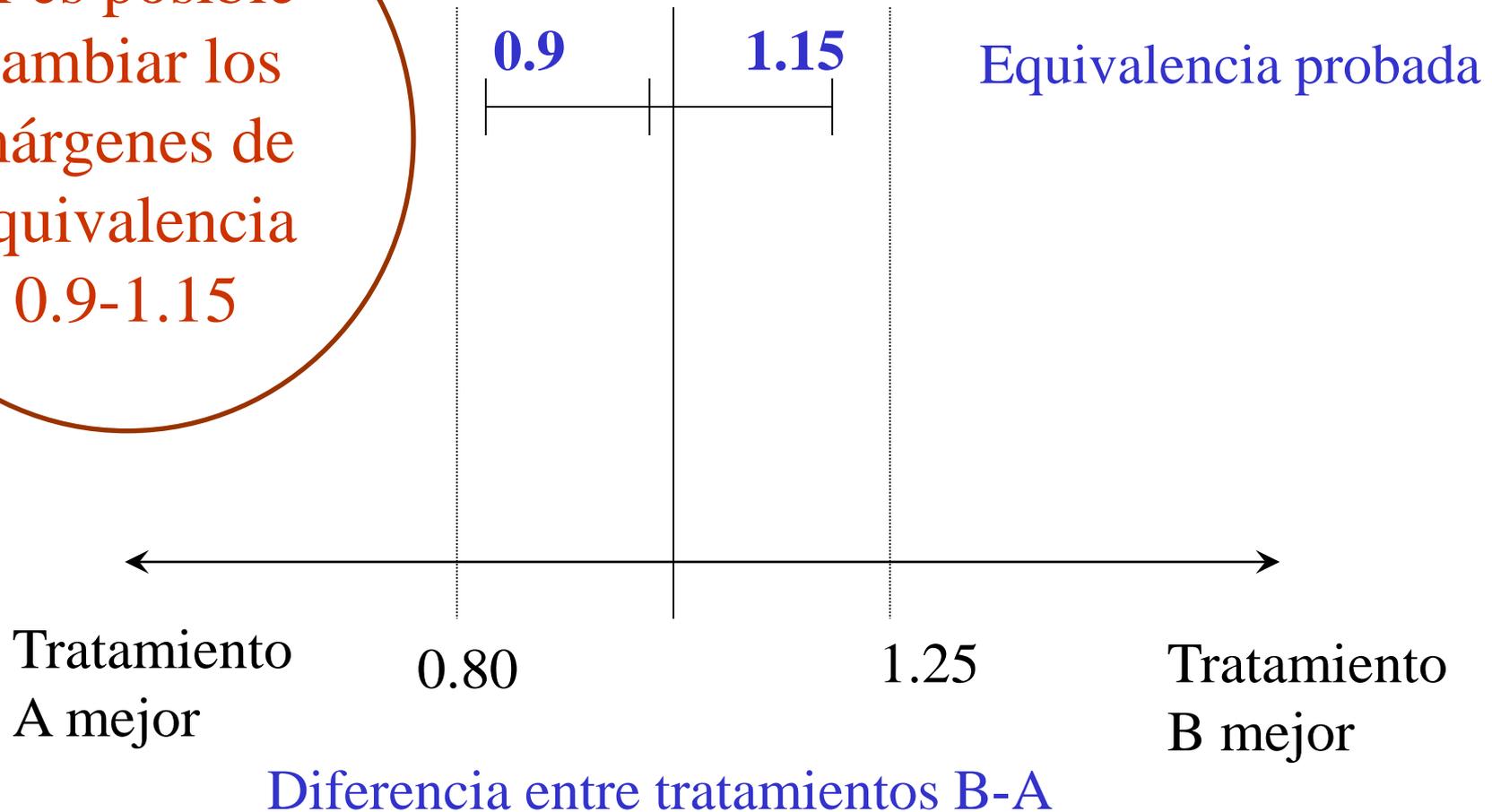
---

¿Es posible realizar cambios en los márgenes de los diseños de equivalencia (no-inferioridad) una vez que el estudio está completado?

*Ejemplo:* En un estudio de bioequivalencia se han definido los márgenes de equivalencia (0.80,1.25). Se obtiene IC(90%) para la biodisponibilidad relativa (RB) de una formulación nueva

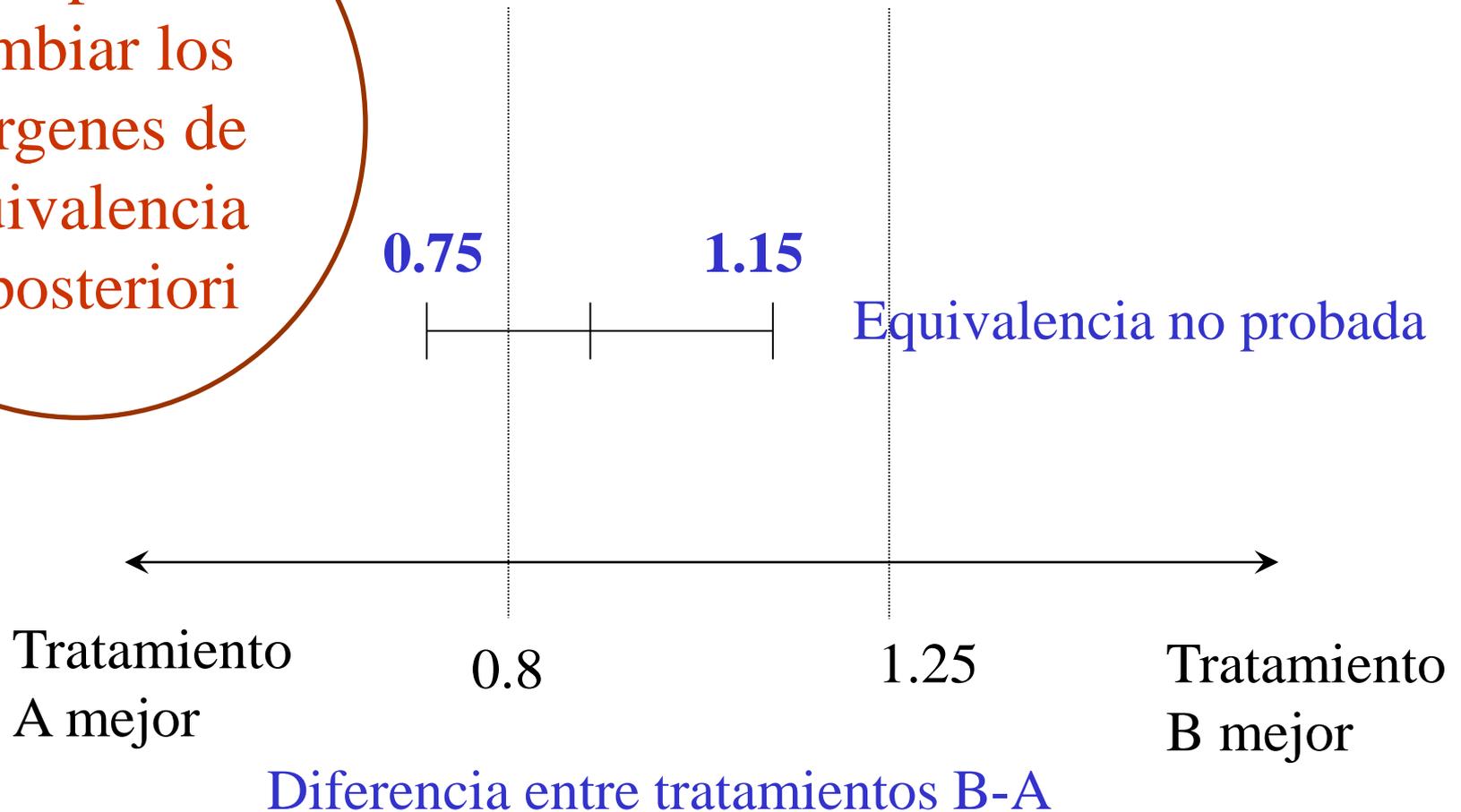
# Márgenes de Equivalencia

Sí es posible  
cambiar los  
márgenes de  
equivalencia  
0.9-1.15



# Márgenes de Equivalencia

No es posible  
cambiar los  
márgenes de  
equivalencia  
a posteriori



# 5 Estudios As Good as or Better trials

Demostrar que B no es clínicamente inferior o que es clínicamente superior a un control A: “B es tan bueno como o mejor que A

**Test no inferioridad**

$$HI_0: B \leq A - \Delta ;$$

$$HI_1: B > A - \Delta$$

**Test superioridad**

$$HS_0: B = A ;$$

$$HS_1: B \neq A$$

**SE PLANTEAN Y RESUELVEN  
SIMULTANEAMENTE PARA LA DIFERENCIA  
DE TRATAMIENTOS B-A**

## ¿Cómo se resuelve?

.Sí  $\mu_B - \mu_A > 0$  y  $HS_0$  rechazada entonces  $HI_0$  rechazada: **B es superior a A**

.Sí  $\mu_B - \mu_A < 0$  y  $HS_0$  rechazada y  $HI_0$  no rechazada : **B es inferior a A**

.Sí  $\mu_B - \mu_A < 0$  y  $HS_0$  rechazada y  $HI_0$  rechazada : **B no es inferior a A**

.Sí  $HS_0$  no rechazada y  $HI_0$  rechazada : **B no es inferior a A**

.Sí  $HS_0$  no rechazada y  $HI_0$  no rechazada : **B es inferior a A**

# 6. Conclusiones

---

1. En un **estudio de superioridad**, si se demuestra la significación estadística, hay que probar si es clínicamente relevante
2. En un **estudio de superioridad**, si no se demuestra la superioridad, esto **NO** implica haber demostrado equivalencia o no inferioridad.

Puede plantearse probar la no inferioridad, para lo que es necesario que se den las condiciones de este tipo de estudios

# 6. Conclusiones

---

3. En un estudio de **no inferioridad** para que ésta quede demostrada, es imprescindible probar la sensibilidad del diseño, probando la eficacia del efecto del control del estudio. En esta situación, puede plantearse probar superioridad.
4. En un estudio de **equivalencia**, en el que se demuestre la equivalencia, pueden obtenerse unos márgenes de equivalencia más estrechos que los fijados.

# 6. Conclusiones

---

5. La **sensibilidad de un estudio** es la capacidad que tiene para distinguir tratamientos efectivos de tratamientos no efectivos.  
En estudios de superioridad se cuestiona cuando no es posible demostrar la superioridad.  
En estudios de no inferioridad (equivalencia) se cuestiona cuando hay evidencia de no inferioridad (equivalencia).
6. En todos ellos es muy importante la elección adecuada del **tamaño muestral**