

Capítulo III

Tema C

¿CUÁNDO PODEMOS DECIR QUE UN MEDICAMENTO ES UN EQUIVALENTE TERAPÉUTICO?

1-LA EQUIVALENCIA TERAPÉUTICA. CONCEPTO

2-TIPOS DE ESTUDIOS PARA DETERMINAR EQUIVALENCIA
TERAPÉUTICA.

3. EVALUACIÓN DE EVIDENCIAS EN LA EQUIVALENCIA TERAPÉUTICA

1-LA EQUIVALENCIA TERAPÉUTICA. CONCEPTO

Desde el punto de vista de una compañía farmacéutica y de las expectativas de los investigadores y de los clínicos lo deseable al estudiar un fármaco nuevo es encontrar evidencias de superioridad del nuevo medicamento respecto al tratamiento estándar que se esté utilizando, de forma que el nuevo producto permita un avance en la terapia de los pacientes.

Por ello, tradicionalmente, los ensayos clínicos se han venido realizando con el objetivo de demostrar la superioridad del medicamento estudiado respecto a otro medicamento tomado de referencia. Esta superioridad puede ser de eficacia o de seguridad.

Pero cada vez es más frecuente que el objetivo del ensayo sea demostrar que el fármaco de estudio es igual de eficaz que el de referencia o, al menos, no inferior, bien porque el tratamiento control tenga una eficacia difícil de superar o porque el tratamiento experimental presente otras ventajas adicionales, como una administración más cómoda o una menor duración del tratamiento.

Desde el punto de vista de la selección de medicamentos, lo que nos interesa es determinar cuándo un medicamento puede ser clasificado como superior al de referencia y cuándo como equivalente terapéutico, bien porque se haya mostrado equivalencia terapéutica o porque las diferencias encontradas no sean importantes clínicamente.

El establecimiento de la Equivalencia Terapéutica tiene importancia tanto desde el punto de vista clínico, como de gestión. Clínicamente podremos establecer si es necesario disponer de varios equivalentes terapéuticos en la Guía Farmacoterapéutica (como segundas opciones de tratamiento) o bien disponer sólo de uno de ellos. Y desde el punto de vista de la gestión, si vale la pena disponer de un solo equivalente o bien definirlos como "medicamentos homólogos", lo que nos permite la inclusión de varios medicamentos (principios activos) definidos como equivalentes terapéuticos y que se utilizarán indistintamente en función de los costes o de la disponibilidad.

Por estas razones cada vez ha tomado más importancia el manejo correcto de la Equivalencia Terapéutica para la evaluación y selección de medicamentos.

2-TIPOS DE ESTUDIOS PARA DETERMINAR EQUIVALENCIA TERAPÉUTICA.

La Equivalencia Terapéutica se puede establecer a partir de diferentes estudios y grados de evidencia. Los estudios de equivalencia requieren, además, unas normas de lectura crítica y de análisis distintos de los estudios de superioridad a los que estamos acostumbrados, que es importante conocer bien. Como expone Greene WL (Ann Intern Med 2000), muchos de los estudios publicados que concluyen que dos opciones terapéuticas son equivalentes, tienen importantes defectos conceptuales o metodológicos.

Los siguientes tipos de estudio nos aportan información útil para clasificar dos medicamentos como equivalentes terapéuticos

Ensayos de equivalencia o no inferioridad

Son ensayos clínicos en los que el objetivo es demostrar que las opciones terapéuticas son equivalentes entre sí o que una de ellas es, al menos, no inferior a la otra.

A nivel conceptual un ensayo comparativo de superioridad suponemos que la hipótesis nula (H_0) es que no hay diferencias entre los dos tratamientos y al rechazarla se abraza la hipótesis alternativa, que es que son diferentes (H_1). En un ensayo de equivalencias es al revés, suponemos que la H_0 es que al menos hay una diferencia entre los dos tratamientos y al final demostraremos la hipótesis de que no hay diferencias.

Este tipo de estudios requiere de unas características diferentes de los de superioridad, tanto en el diseño como en el análisis de los resultados:

- *Debe constar en el diseño:* El que un ensayo tenga por objetivo demostrar la equivalencia o no inferioridad debe quedar definido en el diseño del mismo, y suele constar en el apartado de métodos, y especialmente en el análisis estadístico.

- *Definir el valor "delta":* La base metodológica de este tipo de estudios se basa en definir cual es la máxima diferencia entre los tratamientos que no se considera clínicamente importante, es el llamado valor "**delta**" (d), que viene definido por dos límites, uno inferior y otro superior ($-d$; $+d$). Para que los tratamientos se consideren equivalentes, la diferencia del resultado debe estar dentro de estos márgenes establecidos. Es importante señalar que lo que se mide no es el valor medio de la diferencia, sino que todos los valores comprendidos en el IC 95% deben estar contenidos en el margen establecido como equivalente. Para la definición del valor "delta" se debe considerar la opinión de expertos clínicos y el consenso con las agencias de evaluación. Este valor en general es mucho menor (al menos la mitad) que el empleado en los estudios de superioridad en esta patología.

- *Cálculo del tamaño muestral:* se debe hacer un cálculo del tamaño muestral adecuado para este tipo de estudios. De esta forma, si existe una diferencia real igual o menor al margen establecido, el estudio tiene las suficientes garantías (potencia estadística) para detectarla y confirmar la equivalencia terapéutica entre los tratamientos. Normalmente, el tamaño muestral para este tipo de estudios será mayor que en un estudio de superioridad.

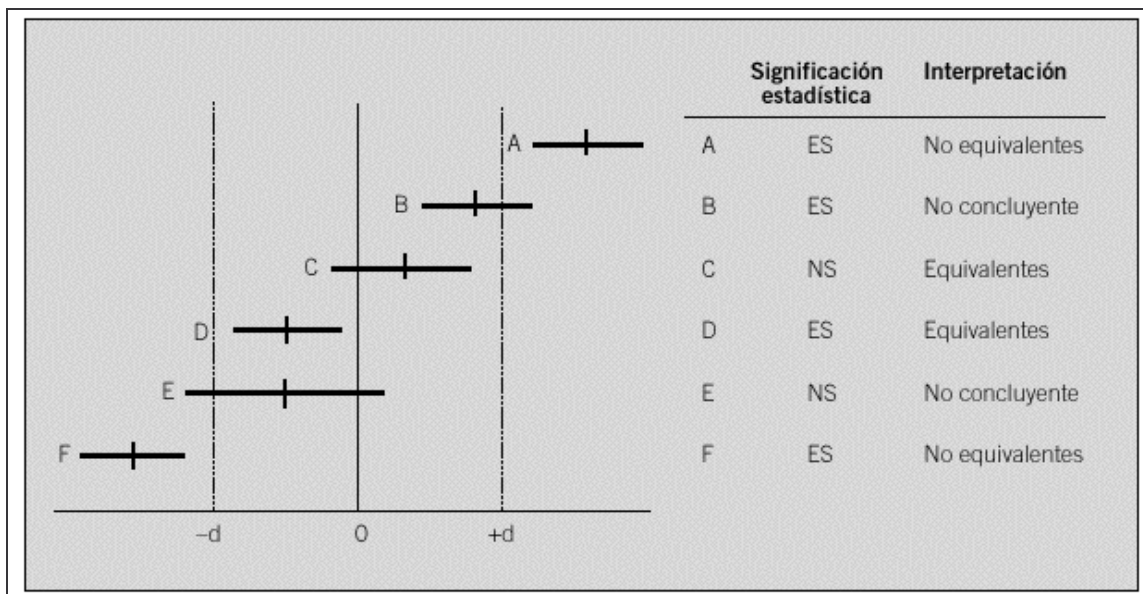
- *Estrategias de análisis.* El análisis de resultados de un ensayo clínico de superioridad se recomienda que se realice por "intención de tratar" porque se garantiza la posición más conservadora, que permitirá valorar los resultados en el caso más desfavorable. Sin embargo en los ensayos clínicos de equivalencia, es necesario también realizar un análisis "per protocol", o de datos observados, es decir, que han sido asignados a un grupo, que cumplen los criterios de inclusión y exclusión, que han recibido el tratamiento completo y que han sido seguidos hasta el final del estudio. Con este enfoque aumentarán las diferencias entre tratamientos, lo que dificulta poder concluir que dos tratamientos son equivalentes, manteniendo la posición más cauta en la interpretación de los resultados. Si los resultados de los dos tipos de análisis no coinciden, es necesario investigar y analizar los subgrupos de pacientes que se han desviado del protocolo.

Ejemplos prácticos de ensayos de equivalencia y no inferioridad:

Ver como ejemplo, la representación gráfica $-d$ y $+d$ es el rango de valores mínimo y máximo compatible con equivalencia terapéutica. Los valores de la Reducción del Riesgo Absoluto (RAR) por encima de $+d$ indican que uno de los fármacos es superior a otro, los valores de RAR por debajo de $-d$ indican que uno de los fármacos es inferior al otro. Hay que determinar tanto el valor de RAR como el de su IC95%. En la gráfica los casos C y D **se concluye que son equivalentes** (el IC95% no atraviesa el valor $+d$ ó el $-d$). Los casos A y F que no lo son. Los casos B y E no son concluyentes (ya que el IC95% atraviesa el valor $+d$ ó el $-d$). Como se observa en la gráfica la interpretación es independiente de si el resultado es estadísticamente significativo, es decir que el IC95% del RAR atravesase el cero.

Gráfica 1

Ejemplo: Posibles resultados de un estudio que compara dos tratamientos con la finalidad de comprobar su equivalencia. $-d$ a $+d$ es el rango de valores compatible con equivalencia terapéutica, preestablecida por los investigadores. ES: Estadísticamente significativo. NS: estadísticamente no significativo. Ref: Argimon JM Med Clin (Barc) 2002, 118:382-4.



Se describen a continuación algunos ejemplos. En todos los ejemplos se evalúa únicamente la eficacia. (Para poder definirlos como equivalentes terapéuticos tendríamos que evaluar además la seguridad).

<p>Tabla 1 Ejemplo: Voriconazol vs Anfoterina B deoxicolato en aspergilosis invasora <i>Hebrecht R. N Eng J Med 2002, 347,6: 408-15</i></p> <p>Objetivo primario: Demostrar la no inferioridad de voriconazol, comparado con Anfotericina B deoxicolato, en la semana 12, en la población por intención de tratar modificada. Se considera voriconazol no inferior si límite inferior del IC 95% de la RAR era superior a -20%.</p> <p>Resultado: RAR: 21,2% (IC95%: +10,4% a +32,9%)..." como el límite inferior del RAR estaba por encima de 0 (y también por tanto a -20%), se consideró que voriconazol no solo no era inferior a Anfo B deoxicolato, sino que era superior"</p>

En el ejemplo 1 se concluye que no hay equivalencia porque el medicamento en estudio es superior (similar a caso A del gráfico).

<p>Tabla 2 Ejemplo: Caspofungina vs Anfoterina B deoxicolato en candidiasis invasora <i>Mora-Duarte J N Eng J Med 2002; 347:2020-9</i></p> <p>Objetivo primario: Demostrar la no inferioridad de caspofungina, comparado con Anfotericina B deoxicolato en una variable compuesta de eficacia "favorable". Se considera Caspofungina no inferior si límite inferior del IC 95% de la RAR era superior a -20%. Se estudia en la población por intención de tratar modificada (MITT). Y per protocolo (PP)</p> <p>Resultado sobre MITT: RAR: 12,7% (IC95%: -0,7% a +26,0%). La respuesta es mejor para Caspofungina aunque no hay diferencia estadística significativa (EL IC incluye el valor 0, p=0,09). Como el límite inferior del RAR esta por encima de -20% el estudio demuestra que Caspofungina es al menos no inferior a Anfotericina B.</p> <p>Resultado sobre PP: RAR: 15,4% (IC95%: +1,1% a +29,7%). La respuesta es mejor para Caspofungina con diferencias estadísticas significativas (EL IC no incluye el valor 0, p=0,03). Como el límite inferior del RAR esta por encima de -20% el estudio demuestra que Caspofungina es al menos no inferior a Anfotericina B.</p>

En el ejemplo 2 el fármaco en estudio es equivalente en eficacia y puede ser superior (en el análisis PP, similar a caso B de la gráfica).

<p>Tabla 3 Ejemplo: Tenecteplasa vs Alteplasa en IAM <i>ASSENT-2 investigators: Lancet 1999; 354: 716-22</i></p> <p>El ensayo clínico ASSENT-II se diseñó para establecer la equivalencia de eficacia y evaluar la seguridad entre alteplasa y tenecteplasa. Se incluyeron 16.949 pacientes de 1.021 hospitales y 29 países. La variable fue la mortalidad a los 30 días, y las complicaciones hemorrágicas.</p> <p>Objetivo: Demostrar la no inferioridad de Tenecteplasa comparado con Alteplasa. Se considera Tenecteplasa equivalente si la diferencia en RAR de mortalidad era <1% o <14% de RR. Se estudia en la población por</p>

intención de tratar modificada (MITT).

Resultado: RAR: 0,028% (IC90%: -0,55% a +0,61%). RR: 1,004 (IC90%: 0,914-1,104) . La respuesta es mejor para Alteplasa aunque no hay diferencia estadística significativa (EL IC incluye e valor 0). Como el límite de los IC del RAR están por debajo del 1% y los del RR del 14% el estudio demuestra que Alteplasa y Tenecteplasa pueden considerarse equivalentes en eficacia.

En el ejemplo 3 son equivalentes en eficacia (similar a caso C del gráfico).

Tabla 4

Ejemplo: Voriconazol vs Anfotericina B Liposomal en neutropenia con fiebre

Walsh N Eng J Med 2002; 346:225-234

Objetivo primario : Demostrar la no inferioridad de voriconazol, comparado con Anfotericina B liposomal en una variable compuesta de eficacia. Se considera voriconazol no inferior si límites del IC 95% de la RAR estaba dentro del -10%.

Resultado: RAR: -4,5% (IC95%: -10,6% a +1,6%)%. La respuesta es peor para Voriconazol aunque no hay diferencia estadística significativa (EL IC incluye el valor 0). Como además el límite inferior del RAR está justo por debajo de -10% el estudio **no consigue demostrar que voriconazol es equivalente terapéutico** de Anfotericina B liposomal en la indicación de cobertura de pacientes neutropénicos.

En el ejemplo 4 no hay equivalencia (similar a caso E del gráfico)

Ensayos de superioridad significativos sin relevancia clínica

Son ensayos clínicos de superioridad cuyos resultados muestran diferencias estadísticamente significativas, pero no se consideran importantes clínicamente. Es el caso más habitual, ya que la mayoría de estudios publicados encuentran diferencias estadísticas significativas y se sabe que desde el punto de vista estadístico aumentando el tamaño muestral de forma suficiente se puede lograr $p < 0,05$, en algunos casos en variables de escasa repercusión clínica, se puede conseguir el valor significativo.

El punto clave es valorar si la mayor eficacia mostrada por el nuevo fármaco, además de ser estadísticamente significativa, es también clínicamente significativa. Para ver si la diferencia es relevante tendremos que:

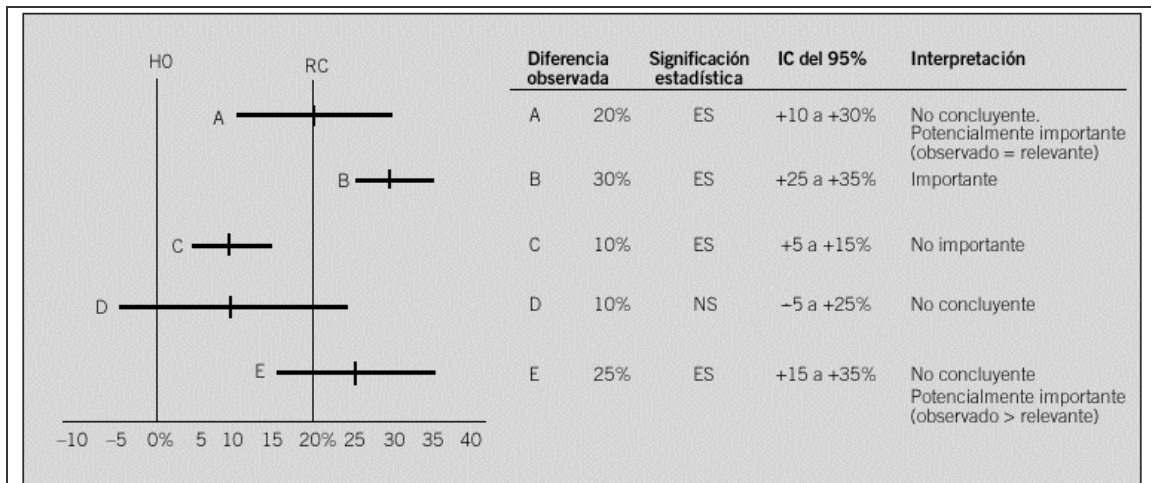
-Establecer a partir de qué valor se considera clínicamente relevante la diferencia. (ejemplo RAR 20 % en la gráfica)

-Valorar si la magnitud de la diferencia (RAR) entre los medicamentos estudiados es superior a la considerada de importancia clínica, y ver si los valores límites del IC 95 % de esta diferencia incluyen el valor establecido como clínicamente relevante. Ver ejemplo gráfico

Gráfica 2.

Resultados hipotéticos de un estudio que compara un tratamiento activo frente a un placebo y utiliza como medidas la diferencia de eficacia (RAR).

Se considera que la mínima RAR relevante clínicamente es del 20 %. ES: Diferencia significativa estadísticamente. NS: No significativa estadísticamente. Ref: Argimon JM Med Clin (Barc) 2002, 118:382-4.



En el caso B, concluimos que un tratamiento es superior al otro desde el punto de vista estadístico y clínico.

En el caso C que es superior desde el punto de vista estadístico pero no se considera importante desde el punto de vista clínico y por tanto podemos considerarlo que son **equivalentes terapéuticamente**.

El resto de opciones A, E no son concluyentes en cuanto a superioridad, pero dado que no podemos excluir que su IC incluya el valor de mínima relevancia clínica, tampoco podemos considerarlos como equivalentes terapéuticos. En el caso E posiblemente sea superior uno de los grupos.

Ensayos de superioridad con resultados no significativos

Son estudios directos entre dos medicamentos con el objetivo de demostrar superioridad pero no encuentran diferencias estadísticas significativas. Son los ensayos “negativos”, que no son habituales, pues la mayoría de estudios publicados encuentran diferencias estadísticas significativas, al menos en los resultados principales. Además existe el sesgo de publicación, de manera que muchos de los ensayos clínicos con resultados de diferencias estadísticas no significativas no llegan a publicarse.

El punto clave a tener en cuenta en este tipo de ensayos, es que la ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica (ver cuadro).

El rechazo de la hipótesis nula :

El método estadístico habitual se basa en la utilización de los denominados contrastes de hipótesis, que obligan al investigador a concretar su problema en dos hipótesis (H0=nula y H1=alternativa). El objetivo es demostrar la existencia de diferencias y por tanto, en relación a las hipótesis planteadas, el rechazo de la hipótesis de igualdad o hipótesis nula (H0).

Por convenio, generalmente se acepta que la probabilidad de equivocarse al concluir que hay diferencias (el llamado error de primera especie) es de un 5 % (alfa=0,05). Por otro lado, se establece que la probabilidad de equivocarse al concluir que no existen diferencias entre ambos tratamientos (error de segunda especie) es menos exigente, a menudo del 20 % (beta=0,20) (Ref Gich 1995).

*Cuando el valor de $p > 0,05$, se considera que la probabilidad de que el resultado obtenido sea debido al azar es demasiado elevada y se **concluye que no son estadísticamente significativos**. No se puede concluir que ambas intervenciones son iguales, sino que **no hay suficientes pruebas para afirmar que son diferentes** (Ref Argimon 2002).*

La falta de significación estadística puede deberse a que el tamaño de la muestra es pequeña, la dispersión de datos elevada y/o la magnitud de la diferencia más pequeña de la prevista.

Al diseñar un estudio de este tipo, los investigadores establecen la **magnitud mínima de la diferencia o asociación que consideran de relevancia clínica** y de acuerdo con ella calculan el **tamaño de la muestra** necesario.

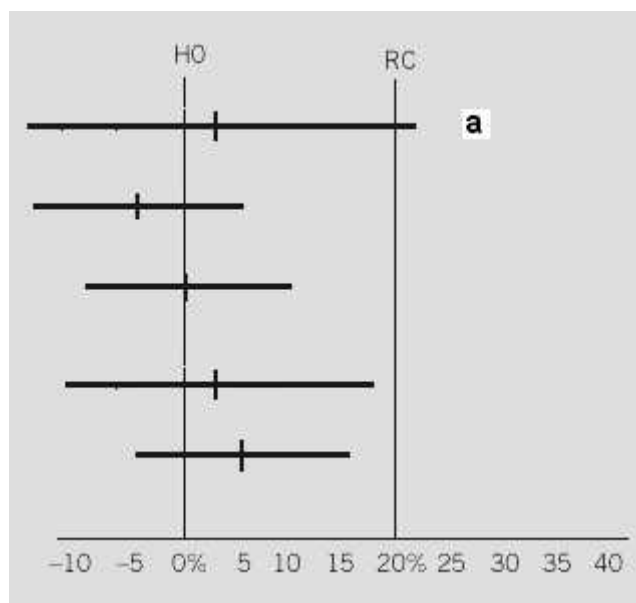
Al finalizar el estudio la interpretación del resultado debe tener en cuenta esta magnitud. Se ha de determinar, por ejemplo, si el RAR obtenido y su IC 95% es superior o no, a dicho valor de relevancia clínica establecido previamente.

Ver ejemplo en la gráfica :

-El valor de RAR que se considera clínicamente relevante en el ejemplo es 20 %.

-En el primer caso (parte superior (A)) no se presenta diferencias estadísticamente significativas pero no podemos descartar que haya una diferencia clínicamente relevante superior a 20%. El resultado es que posiblemente la eficacia sea comparable o superponible, pero no se puede ser concluyente (no se puede descartar que sea superior).

-El resto de resultados de la gráfica no muestran diferencias estadísticamente y que además su IC no atraviesa el valor 20 %. Si el tamaño de la muestra se ha calculado respecto a este objetivo, podremos considerar que de existir diferencias, estas serán **de poca entidad clínica** y en la práctica los podemos considerar de eficacia **comparable o equivalente**.



En los análisis de **subgrupos** o en caso de resultados de objetivos secundarios, en general de tamaño muestral no será suficiente, se dará una situación similar, en general en este caso los IC 95% serán más amplios y se darán mas situaciones del tipo "no concluyente".

Limitaciones:

Un punto importante es contrastar si la magnitud de la diferencia considerada de relevancia clínica para el cálculo del tamaño muestral, es realmente un valor que según nuestros conocimientos o el conocimiento de expertos en el tema, es adecuada. Muchas veces el tamaño muestral depende de factores como el número de pacientes que se estima es posible reclutar, de los recursos disponibles para realizar el ensayo o de otros factores.

En general, en los estudios de superioridad la definición de la magnitud de la diferencia clínicamente importante es bastante superior a la de los estudios de equivalencia (ejemplo x 2). Este hecho además de la naturaleza y objetivos distintos de cada ensayo hace más difícil el ser concluyente. Por ello, el nivel de evidencia aportado por los ensayos clínicos comparativos

4º Curso. Evaluación y selección de medicamentos
Palma de Mallorca, 28-31 Marzo 2006
Manual para la evaluación de nuevos fármacos en el hospital
Versión nº 6. Marzo 2006.

cuando se analizan para definir equivalencias, es inferior al de los ensayos diseñados para determinar equivalencia o no inferioridad.

La elección y aplicación de valores de la que van a considerarse de significación clínica de un RAR siempre es difícil. El problema reside en definir qué se entiende por una magnitud no relevante clínicamente. Una orientación sobre cual es la magnitud de un resultado clínicamente significativo, es que al menos debe superar la variabilidad producida en el medio asistencial por la propia característica de la atención sobre un tipo de paciente y enfermedad. Es importante decir que este margen de equivalencia depende de la categoría de la variable. Si es mortalidad el margen es muy estrecho o inadmisible, si hablamos de variables blandas el margen puede ser amplio (hemoglobina glicosilada, glucemias). Es decir, el margen de equivalencia depende de la calidad de la variable.

En un segundo nivel de nuestra estimación podemos sustituir el valor de RAR establecido como aceptable por el que creamos más apropiado, y analizar los resultados del ensayo desde este punto de vista.

En general para dar más robustez a las conclusiones, en este tipo de ensayos, debemos contrastar los resultados con información adicional. Ejemplo, existencia de varios estudios con resultados similares, conclusiones y discusión expresada por los autores del mismo y otras fuentes secundarias.

Ejemplo:

Tabla 5						
Ejemplo: Ensayos clínicos de Atosiban versus estimulantes beta-2. Efectividad superponible en el tratamiento de la amenaza parto prematuro.						
<i>Tres ensayos clínicos publicados (ver tabla siguiente) comparan Atosiban con fármacos adrenérgicos beta estimulantes (Ritodrina, Salbutamol, Terbutalina). Se dispone de un metanálisis que resume los tres estudios anteriores (Moutquin JM et al Br J Obstet Gynaecol 2001).</i>						
<i>El tamaño de la muestra se calculó para detectar un 18 % de aumento de eficacia tocolítica .</i>						
Ensayos clínicos versus estimulantes beta-2						
	Tto de referencia (pac de cada grupo)	Tiempo de eval.	Resultados		P	RAR (IC 95%)
			Atosiban	Tto de referencia		
Moutquin ,2000 247 mujeres	Atosiban n=126 Ritodrina n=121	48 h 7 días	84,9% 73,0%	86,8% 76,0%	0,99 0,85	-1,9 % (-10,5% a +6,8%) -3 % (-13,9% a +7,8%)
French/Austr Atosiban invest group 2001 241 mujeres	Atosiban n=119 Salbutamol n= 122	48 h 7 días	93,3% 89,9%	95,0% 90,1%	0,67 0,93	-1,8 % (-7,7% a +4,1%) -0,2 % (-7,8% a +7,3%)
European Atosiban invest group 2001 245 mujeres	Atosiban n=116 Terbutalina n=129	48 h 7 días	86,1% 76,5%	85,3% 67,4%	0,78 0,07	+0,9% (-7,8% a +9,7%) +9,3% (-1,9% a +20,4%)
Sumario de los tres estudios anteriores						
Moutquin. 2001 733 mujeres	Atosiban n=363 Beta adren n=379	48 h 7 días	88,1% 79,7%	88,9% 77,6%	0,99 0,28	-0,8% (-5,4% a +3,8%) +2% (-3,9% a +7,9%)
Resultado:						
-En dos de los estudios se muestra una tendencia a menor efectividad de Atosiban sin sobrepasar el límite definido.						
-En el tercero, la evaluación a los 7 días con un IC (-1,9% a +20,4%), podría considerarse como no concluyente.						
-Globalmente y los resultados del metanálisis muestran una equivalencia de efectividad de Atosiban y beta-adrenérgicos a los 48 h y a los 7 días.						
(Nota: Ver informe EMEA que confirma estos resultados de efectividad equivalente y al mismo tiempo indica que hay algunas diferencias al analizar por separado la toxicidad y la eficacia.)						
Ref:						
Moutquin , Am J Obstet Gynecol 2000, 182: 1191-9						
French/Austr Atosiban invest group , Europ J Obstet Gynecol 2001 98:188-95						
European Atosiban invest group Acta Obstetr Gynecol Scand 2001; 80: 413-22						
Moutquin. Br J Obstet Gynaecol 2001; 108:133-42						
Informe EMEA 2000 (CPMP 253/99)						

Ensayos clínicos diferentes pero con un tercer comparador común: placebo u otro fármaco

4º Curso. Evaluación y selección de medicamentos
Palma de Mallorca, 28-31 Marzo 2006
Manual para la evaluación de nuevos fármacos en el hospital
Versión nº 6. Marzo 2006.

Ocurre cuando el ensayo pivotal compara el medicamento nuevo con un medicamento de referencia C, y al mismo tiempo disponemos de otro ensayo clínico del medicamento similar comparado también con C. El medicamento C o tercer comparador suele ser placebo, aunque a veces puede ser un medicamento de referencia. Como se ha explicado en el capítulo I y dado el sistema de autorización y registros de las Agencias como la EMEA, se da con bastante frecuencia.

En este caso hay que tener en cuenta si los dos ensayos que comparamos se han realizado sobre una muestra de una misma población, o bien se trata de poblaciones con un riesgo base diferente, e incluso, en patología diferentes. Dos medicamentos pueden ser igualmente efectivos pero uno puede parecer superior porque se ha empleado en el ensayo sobre una población que responde mejor a la intervención. Ello puede ocurrir por ejemplo si comparamos una estatina empleada en un ensayo en prevención primaria, con otra estatina empleada en prevención secundaria. El RRR y sobre todo el RAR puede ser mucho mayor en los ensayos de prevención secundaria que en el de prevención primaria, independientemente de la estatina estudiada.

En este caso, convendría:

1-Ver si se realizan sobre la misma o similar población e indicación : Ver reclutamiento, criterios de inclusión y exclusión, la llamada tabla 1, las condiciones del tratamiento (dosis, duración) tipo de seguimiento y evaluación.

2-Ver si los resultados del grupo control de cada ensayo son superponibles.

3-Si se cumple el anterior punto, ver si hay diferencias significativas del RAR entre ambos estudios (ver si hay solapamiento de IC 95%), en este caso podemos decir que no hay evidencias de superioridad de uno u otro medicamento, y a nivel práctico hasta no disponer de estudios más concluyentes se pueden **consideran equivalentes**. También se puede mirar las medidas relativas (RR, OR). En general las medidas relativas (RR, OR) se mantienen más constantes que las absolutas y se confirmará lo obtenido con el RAR.

4- Si hay diferencias significativas entre ambos, ver si la diferencia es clínicamente relevante.

5- Si es sobre distinta población, con factores de riesgo base muy distintos se puede emplear el mismo método. La interpretación debe ser cautelosa, pero si ambos IC se solapan orientan hacia una equivalencia pero la evidencia es mucho más débil.

Tabla 6		
Ejemplo: Peg-interferon alfa 2-a vs Peg-interferon alfa 2-b.		
No existen estudios comparativos directos, nos basamos en los ensayos de ambos Peg-interferones vs Interferon alfa 2-b no pegilado		
Comparación de los dos preparados de Peg-interferon disponibles. Resultados de comparar con un medicamento común (vs Interferon alfa2-b)		
	PEG-Interferón alfa-2-a vs Interferon alfa2-b	PEG-Interferón alfa 2-b vs Interferon alfa2-b
Eficacia Respuesta Viral Sostenida	Fried. N Engl J Med 2002;347:975-982.(n=1121) Peginterferón alfa 2a 180mcg/semana+ 1000/1200mg/día ribavirina	Manns.Lancet 2001; 358: 958-ñ65. (n=1530) Peginterferon alfa 2-b 1,5mcg/Kg/semana+ ribavirina 800mg/día
Global vs IFN	56% vs 44% (p<0.001)	54% vs 47% (p=0.01)
Genotipo 1	46% vs 36% (p=0.01)	42% vs 33% (p=0.02)
Genotipo 2-3	76% vs 61% (p=0.005)	82% vs 79% (p=NS)
Global <2M copias/ml	62% vs 52 % (0,04)	78% vs 56 % (0,01)
Global >2Mcopias/ml	53% vs 41 % (0,003)	42% vs 42 % (ns)

Tabla 7:
Ejemplo Peg-interferon alfa 2-a vs Peg-interferon alfa 2-b.

Cálculo de RAR y de RR con IC.
 Estimaciones aproximadas realizadas con programa EBM calc y redondeos

Variable de eficacia: Respuesta Viral Sostenida	RAR (IC95)		RR (IC95)	
	Peg-a vs Int	Peg-b vs Int	Peg-a vs Int	Peg-b vs Int
Global	12%(5-18)	7% (1-13)	0,8(0,7-0,9)	0,9(0,8-1)
Genotipo 1	10 %(2-18)	14%(7-21)	0,8(0,7-1)	0,9(0,8-1)
Genotipo 2 o 3	15%(4-26)	3%(-6 a +12)	0,6(0,4-0,9)	0,8 (0,5-1,33)
<2Mill copias/ml	10 %(0-21)	22%(12-32)	0,8 (0,6-1)	0,5(0,3-0,7)
>2Mill copias/ml	12%(4-20)	0 %(-1 a +1)	0,8(0,7-0,9)	1(0,9-1,1)

Consideraciones

Los resultados de los IC se solapan en todos los casos, excepto en el subgrupo de carga viral > 2 millones de copias/ml

Las dosis de ribavirina empleadas en dichos estudios ha sido diferente, por lo que no son totalmente equiparables, habría que revisar también el resto de factores de ambos estudios se han realizado sobre una misma población y con un tipo similar de intervención

Ver si los resultados del grupo control son superponibles. Esto da una idea, no sólo de los resultados, sino del riesgo basal de la población estudiada. Por ejemplo los resultados con los 2 peginterferon alfa, en los que los resultados del grupo control son 44 % y 47% (tienen una diferencia del 3% global, relativamente pequeña) pero en genotipo no 1, es de 61 % y 79 % (un 18% de diferencia, que puede ser indicativo de poblaciones diferentes). La comparabilidad de los grupos control es una premisa para poder comparar 2 ensayos.

Tabla 8
Anti-TNF en artritis reumatoide

- Ejemplo Anti-TNF en artritis reumatoide: 4 ensayos independientes de Etanercept, Adalimumab e Infliximab vs Placebo+MTX

ACR50 A LAS 24 SEMANAS

¿Similares? **¿Se solapan?**

Tabla 9. Estudios comparación con Placebo+ MTX
Eficacia comparada entre los distintos fármacos. Estudios independientes frente a un tercer comparador común (en este caso MTX +placebo).
 Hay cuatro ensayos:
 -Etanercept + MTX versus Placebo + MTX (Weinblatt . N Eng J Med 1999). Dosis Etanercept 25mg 2vec/sem
 -Adalimumab + MTX versus Placebo + MTX (ARMADA Arthritis Rheum 2003). Dosis Adalimumab 40 mg/2 sem
 -Adalimumab + MTX versus Placebo + MTX (Estudio DE019 datos en informe EMEA e informe FDA). Dosis Adalimumab 40 mg/2 sem
 -Infliximab + MTX versus Placebo + MTX (Maini Lancet 1999). Dosis Infliximab 3 mg/Kg pauta gradual hasta cada 2 meses

Fármaco (estudio)	ACR 50 % a 24 semanas (30 sem Infliximab)		RAR (IC 95%)	NNT (IC 95%)
	Fármaco Si:No (% de si)	Control Si:No (% de si)		
Etanercept+MTX (Weinblatt 1999)	23:36 39%	1:29 3%	35,65 (21,6 a 49,7)	2,8 (2 a 4,6)
Adalimumab+MTX (ARMADA)	33:30 52,4%	4:56 6,7%	45,7 (31,8 a 59,6)	2,2 (1,7 a 3,1)
Adalimumab+MTX (DE019)	81:126 39,1%	19:181 9,5%	29,6 (21,9 a 37,4)	3,4 (2,7 a 4,6)
Infliximab+MTX (Maini 1999)	22: 61 27%	4:80 5%	21,7 (11,2 a 32,3)	4,6 (3,1 a 8,9)

* Cálculos sobre datos brutos de pacientes extraídos de la publicación y aplicación del programa "Risk Reduction Calculator - Illinois"

En este tipo de estudios es difícil llegar a resultados concluyentes de equivalencia terapéutica, pues comparamos estudios de poblaciones diferentes, seguimiento diferentes, etc. Además del estudio en si, y el análisis pormenorizado, debemos tener apoyo de otras fuentes y opiniones.

4º Curso. Evaluación y selección de medicamentos
Palma de Mallorca, 28-31 Marzo 2006
Manual para la evaluación de nuevos fármacos en el hospital
Versión nº 6. Marzo 2006.

En general se dispone de fuentes secundarias que exponen de forma razonada las ventajas e inconvenientes de un nuevo medicamento y sobre si aporta o no algo nuevo, y sobre si su eficacia y seguridad es similar. Las editoriales de las revistas, artículos de revisión y las GPC son fuentes que deben consultarse para confirmar las conclusiones del método propuesto.

Estudios observaciones

Se publican con frecuencia estudios, la mayoría de tipo observacional, en que se describen los resultados clínicos derivados de su aplicación práctica.

Este tipo de estudios aporta cierta información para estimar una equivalencia terapéutica de medicamentos que llevan tiempo en el mercado y tienen la ventaja de que se han realizado en un medio asistencial hospitalario similar al habitual (ej ver cuadro).

Ejemplos de estudios sobre intercambio terapéuticos en la práctica

Oh T, Franko TG. Implementing therapeutic interchange of intravenous famotidine for cimetidine and ranitidine. Am J Hosp Pharm 1990; 47(7):1547-1551.

Frighetto L, Nckoloff D, Jewesson P. Antibiotic therapeutic interchange program. Hosp Formul 1995; 30(93):105.

Gutensohn A, Bunz D, Frighetto L, Jewesson P. Outcome of a ceftriaxone/cefotaxime interchange programme in a major teaching hospital. Chemotherapy 1991; 37 Suppl 3:15-21.

Chase SL, Peterson AM, Wordell CJ. Therapeutic-interchange program for oral histamine H2-receptor antagonists. Am J Health Syst Pharm 1998; 55(13):1382-1386.

Howard RT, Wilson JP, Smeeding JE. Therapeutics interchange of H₂ antagonists, cost savings and impact on outcomes. Hosp Pharm 2000; 35(4):387-392

Bollinger KA, Vermeulen LC, Davis SN, Geurkink EA. Comparative effectiveness of low-molecular-weight heparins after therapeutic interchange. Am J Health Syst Pharm 2000; 57(4):368-372.

Walters JM, Noel H, Folstad J. Prospective evaluation of the therapeutic interchange of felodipine ER for amlodipine in patients with hypertension. Hosp Pharm 2000; 35:48-51.

Clay DR, Bourg MP, Lawrence DB. Outcomes of an amlodipine-to-felodipine therapeutic interchange program. Am J Health-Syst Pharm 2000; 57:1604-1607.

Mason B. Therapeutics Interchange of Lisinopril to Quinapril: Clinical Experience. Hosp Pharm 1996; 31(5):542-544.

.Stock AJ, Kofoed L. Therapeutic interchange of fluoxetine and sertraline: experience in the clinical setting [see comments]. Am J Hosp Pharm 1994; 51(18):2279-2281.

Documentación basada en juicios clínicos, opiniones de expertos, recomendaciones

En el caso de no existir evidencia de ensayos clínicos controlados, la documentación de establecer la equivalencia se basa en la experiencia clínica, reflejada en publicaciones de editoriales, consensos, opiniones de expertos o recomendaciones.

3. EVALUACION DE EVIDENCIAS EN LA EQUIVALENCIA TERAPEUTICA

Desde un punto de vista práctico podemos enfocar el tema de la siguiente forma:

Evidencia de equivalencia: ensayos que permiten **evidenciar** la equivalencia terapéutica:

Nivel 1a : Ensayos de equivalencia

Nivel 1 b: Ensayos de no inferioridad

Estimación de la equivalencia: ensayos que permiten **estimar** una equivalencia terapéutica:

4º Curso. Evaluación y selección de medicamentos
Palma de Mallorca, 28-31 Marzo 2006
Manual para la evaluación de nuevos fármacos en el hospital
Versión nº 6. Marzo 2006.

- Nivel 2: Ensayos de superioridad
 - 2a: ensayos de superioridad significativos sin relevancia clínica
 - 2b: ensayos de superioridad no significativos
- Nivel 3: Ensayos clínicos diferentes con un tercer comparador común
- Nivel 4: Ensayos pragmáticos
- Nivel 5: Documentación basada en juicios clínico, opiniones expertos, recomendaciones.

Referencias:

Altman DG, Bland JM. *Absence of evidence is not evidence of Absence. BMJ* 1995; 311:485-7

Jones B, Jarvis P, Lewis JA, Ebbutt AF. *Trials to assess equivalence: the importance of rigorous methods. BMJ* 1996; 313:36-39.

Greene WL, Concato J, Feinstein AR. *Claims of Equivalence in Medical Research: Are They Supported by the Evidence? Ann Intern Med* 2000; 132: 715-722.

Song F et al: *Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from publisher meta analyses. BMJ* 2003, 326:427-7

Mc Alister et al: *Drug class effects in Users Guide to the medical literature. 2B3 pp415-31. AMA ed 2002*

Argimon JM. *El intervalo de confianza, algo más que un vaor de significación estadística. Med Clin (Barc)* 2002; 118:382-4

Argimon JM: *La ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica. Med Clin (Barc)* 2002, 118: 701-3

Gich I: *Probando la hipótesis nula. Investigación clínica y bioética* 1995; 16: 15-16

Crespí Monjo M et al *La evidencia de la equivalencia: Evaluación de fármacos equivalentes .Comunicación al 49 congreso de la SEFH. 2004*