



## Artículo especial

### ¿Cómo comparar fármacos biológicos?



Xavier Calvet<sup>a,b,\*</sup> y Juan Vicente Esplugues<sup>b,c,d</sup>, en representación del Foro para la Adecuación de las Técnicas de Evaluación de Nuevos Anticuerpos monoclonales (Foro ATENAS)<sup>◇</sup>

<sup>a</sup> Servicio de Aparato Digestivo, Corporació Sanitària Universitària Parc Taulí, Universitat Autònoma de Barcelona, Sabadell, España

<sup>b</sup> CIBERhd, Madrid, España

<sup>c</sup> Departamento de Farmacología, Facultad de Medicina, Universidad de Valencia, Valencia, España

<sup>d</sup> Farmacología Clínica, Hospital Universitario Doctor Peset, Valencia, España

#### INFORMACIÓN DEL ARTÍCULO

##### Historia del artículo:

Recibido el 27 de marzo de 2014

Aceptado el 15 de junio de 2014

On-line el 16 de julio de 2014

##### Palabras clave:

Fármacos biológicos

Comparaciones indirectas

Estudios de no inferioridad

##### Keywords:

Biologic drugs

Indirect comparisons

Non inferiority studies

#### R E S U M E N

El presente documento de consenso revisa la evidencia sobre evaluación de fármacos biológicos. Las conclusiones principales del grupo son: *a*) la evidencia actual sobre comparación de biológicos se basa en comparaciones indirectas y es, en general, poco fiable y con importantes limitaciones metodológicas; por ello, *b*) se considera necesario modificar las directivas regulatorias en el sentido de favorecer decididamente los estudios aleatorizados de no inferioridad comparando cara a cara los nuevos biológicos con los actuales estándares de tratamiento, evitando los ensayos frente a placebo; *c*) un elemento clave en este proceso será la determinación por consenso entre las agencias reguladoras, las sociedades científicas, la industria farmacéutica y las autoridades sanitarias de las diferencias clínicas que deben considerarse relevantes en cada una de las patologías evaluadas.

© 2014 Elsevier España, S.L.U. Todos los derechos reservados.

#### How to compare biologic drugs

##### A B S T R A C T

This consensus document reviews the evidence on the evaluation of biological drugs. The main conclusions of the group are: *a*) the current evidence on biological comparisons is based on indirect comparisons and is generally unreliable and with important methodological limitations. Therefore, *b*) it is considered necessary to amend the regulatory directives in the sense of strongly favoring randomized non-inferiority studies comparing face to face the new biological treatment with current standards, avoiding trials versus placebo, *c*) A key element in this process will be determined by consensus among regulatory agencies, scientific societies, the pharmaceutical industry and health authorities regarding the clinical differences that should be considered relevant in each of the conditions tested.

© 2014 Elsevier España, S.L.U. All rights reserved.

#### Introducción

Los fármacos biológicos han representado una revolución terapéutica en las enfermedades reumatológicas —artritis reumatoide (AR), espondilitis anquilosante (EA), artritis psoriásica (Aps)—, las enfermedades inflamatorias intestinales (EII) —enfermedad de Crohn y colitis ulcerosa— y en determinadas enfermedades cutáneas —psoriasis moderada o grave<sup>1,2</sup>. No solamente han

demostrado su eficacia sobre los síntomas, sino que también pueden modificar la historia natural de dichas enfermedades, evitando sus complicaciones y la discapacidad asociada<sup>3-8</sup>.

A diferencia de los fármacos tradicionales obtenidos mediante síntesis química, los biológicos son moléculas generadas por células vivas, de base proteica. Su tamaño y peso molecular resultan variables (desde cadenas peptídicas hasta moléculas completas de anticuerpo), pudiendo ser muy elevados<sup>9</sup>. Aunque, por definición, no existen 2 moléculas de un mismo biológico idénticas al 100%, las diferencias entre miembros de una familia —p. ej., los anti-TNF que comparten una diana terapéutica pueden ser importantes. Las discrepancias radican en su cadena de aminoácidos o —en el caso de los fármacos biosimilares, que en general presentan una secuencia

\* Autor para correspondencia.

Correo electrónico: [xcalvet@tauli.cat](mailto:xcalvet@tauli.cat) (X. Calvet).

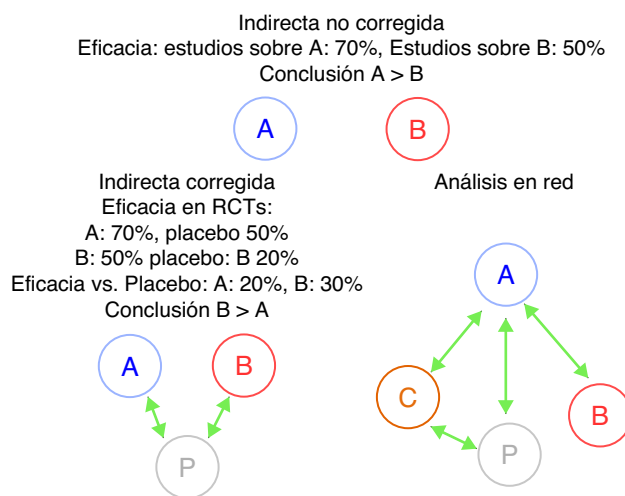
◇ Los componentes del foro ATENAS se relacionan en el [anexo](#).

de aminoácidos idéntica al fármaco original— en las modificaciones —p. ej., glucosilaciones o fucosilaciones— que sufre la cadena de aminoácidos después de su síntesis y que, al condicionar su plegamiento tridimensional, pueden originar variaciones en su afinidad por el sustrato o el grado de inmunogenicidad y ser causa de diferencias en eficacia o seguridad<sup>10</sup>. De hecho, como ejemplo, se ha descrito una incidencia significativamente más alta de anemia aplásica grave asociada a determinadas formulaciones de eritropoyetina recombinante pero no de otras<sup>9</sup>; recientemente, también se han demostrado diferencias en el patrón de afucosilación, en la afinidad por el receptor Fc $\gamma$ RIIIa, y, en estudios *in vitro*, la citotoxicidad dependiente de anticuerpos mediada por células entre infliximab y su biosimilar Inflectra<sup>®</sup>. La Agencia del Medicamento canadiense ha justificado, sobre la base de estos datos, que la aprobación concedida a Inflectra<sup>®</sup> para enfermedades reumáticas no se extienda a la EII<sup>11</sup>.

La oferta de fármacos biológicos ha aumentado rápidamente en los últimos años. Además, tras la caducidad de la patente y el final del periodo de protección de datos de los fármacos originales innovadores, han aparecido en el mercado los fármacos biosimilares, término entendido como «las copias de fármacos biológicos ya autorizados en los que se ha demostrado similares características físico-químicas, eficacia y seguridad tras efectuar las comparaciones necesarias»<sup>12</sup>. La definición hace énfasis en 2 aspectos: a) el hecho de que nunca van a ser iguales al fármaco original, de ahí el término «similar» frente al concepto de «idéntico», que sería aplicable a la comparación de un fármaco genérico respecto de su molécula original, y b) que tratándose de fármacos biológicos la noción farmacológica de bioequivalencia —la demostración de áreas bajo la curva séricas similares entre el fármaco original y la copia que se utiliza para demostrar la equivalencia terapéutica de los fármacos genéricos—, no es un criterio de equivalencia definitivo para considerar que una copia y su biológico original tienen la misma eficacia y seguridad. Se requiere, por tanto, una amplia evaluación de cada nuevo fármaco. No deben analizarse solamente las características físico-químicas, sino que también es necesario realizar una cuidadosa valoración clínica de su eficacia y seguridad para considerar una determinada copia como biosimilar<sup>12</sup>.

Los parámetros a determinar en dicha evaluación son motivo de debate<sup>9,13-19</sup>, aunque la orientación general de las agencias reguladoras y, en particular, de la European Medicines Agency ha sido la de requerir ensayos clínicos aleatorizados de equivalencia clínica o no inferioridad comparando la eficacia y la seguridad del biosimilar con la de sus originales<sup>10</sup>. Esto contrasta con el proceso de aprobación de fármacos biológicos innovadores, donde la mayoría de los estudios comparan el fármaco con un grupo control tratado con placebo. Como norma, los estudios de «equivalencia» o «no inferioridad» pretenden demostrar que estos términos son aplicables para un nuevo fármaco frente a un estándar terapéutico conocido —el nuevo fármaco es «equivalente» o «no inferior al conocido»— y en la mayoría de los casos no se utiliza placebo.

Una dificultad adicional a la hora de comparar fármacos biológicos radica en que, hasta hace poco, no se han publicado ensayos clínicos comparando directamente la eficacia y la seguridad de 2 biológicos. Esta carencia de comparaciones directas ha dado lugar a intentos de contrastarlos utilizando otros métodos de medicina basada en la evidencia, específicamente el análisis mediante comparaciones indirectas<sup>20-23</sup>. El método indirecto más sencillo son las comparaciones indirectas no ajustadas. Consiste en comparar la eficacia de 2 biológicos —que denominaremos A y B— utilizando la eficacia de A en los estudios que evalúan este fármaco comparándola directamente con la eficacia del fármaco B en sus respectivos estudios en la misma enfermedad, sin efectuar ningún tipo de corrección. Este método da muy a menudo resultados erróneos y, por tanto, se desaconseja totalmente su uso<sup>24</sup>. Una alternativa más



**Figura 1.** Comparaciones indirectas, corregidas, no corregidas o en red. Se muestra un ejemplo donde la comparación indirecta no corregida entre 2 fármacos, A y B, da un resultado probablemente incorrecto y totalmente distinto de la corregida debido a las diferentes características de la población estudiada. En los análisis en red pueden incluirse múltiples comparaciones entre diferentes fármacos (A, B, C) y/o con placebo. RCT: acrónimo inglés para estudios aleatorizados y controlados.

correcta desde el punto de vista metodológico son las comparaciones indirectas ajustadas. En este caso, debemos de disponer de ensayos clínicos aleatorizados comparando A y B frente al comparador común P (en el caso de los biológicos habitualmente se trata del placebo). La eficacia de A y B se compara a través de P con objeto de corregir —al menos parcialmente— las diferencias entre las poblaciones de los distintos estudios. Las comparaciones indirectas pueden complicarse mucho más, por ejemplo, si evaluamos múltiples fármacos (análisis en red o network analysis)<sup>24-26</sup>. Para estas evaluaciones se utilizan técnicas estadísticas más complejas, como la metarregresión (fig. 1). Sin embargo, si no se utilizan con un extremo rigor metodológico, estas herramientas pueden generar resultados imprecisos. Veremos que un buen ejemplo de ello es la evaluación de los fármacos biológicos.

Aunque la complejidad de los fármacos biológicos crea serias dificultades a la hora de compararlos, la demostración de que 2 fármacos son clínicamente equivalentes tiene importantes implicaciones asistenciales y económicas. El presente artículo pretende reflexionar sobre algunos aspectos relevantes para facilitar la evaluación de fármacos biológicos: a) la utilidad y las implicaciones éticas del diseño de determinados estudios, y en especial, del uso de placebo; b) la utilización de estudios de no inferioridad, las variables de evaluación a considerar y la importancia de establecer qué diferencia ( $\delta$ ) en eficacia o seguridad puede considerarse clínicamente no significativa, y c) la utilidad de los métodos de medicina basada en la evidencia y especialmente de las comparaciones indirectas. Todo ello se ha realizado a través de un abordaje multidisciplinar mediante una revisión no sistemática de la literatura y un posterior debate y consenso en el que han participado especialistas en reumatología, dermatología, gastroenterología, farmacólogos clínicos y expertos en estadística.

## Método

La elaboración del documento se realizó a partir de una revisión no sistemática realizada y consensuada por dos de los autores (XC y JVE). El resto de los participantes en el foro recibieron el documento por correo electrónico, revisaron el documento y realizaron aportaciones que se recogieron en un documento inicial. En una única reunión presencial, se discutieron los puntos de debate, se consensuaron la estructura y el contenido final del documento, y

se distribuyeron la responsabilidad a cada uno de los participantes. Así, los respectivos especialistas elaboraron las bases para proponer un valor de delta en cada una de las indicaciones de los fármacos biológicos. Una vez elaborada esta parte, los coordinadores (XC y JVE) integraron las distintas aportaciones elaborando un segundo documento, que se discutió a través del correo electrónico. Finalmente, todos los participantes en el foro dieron su aprobación al contenido final del documento.

### Conclusiones del consenso

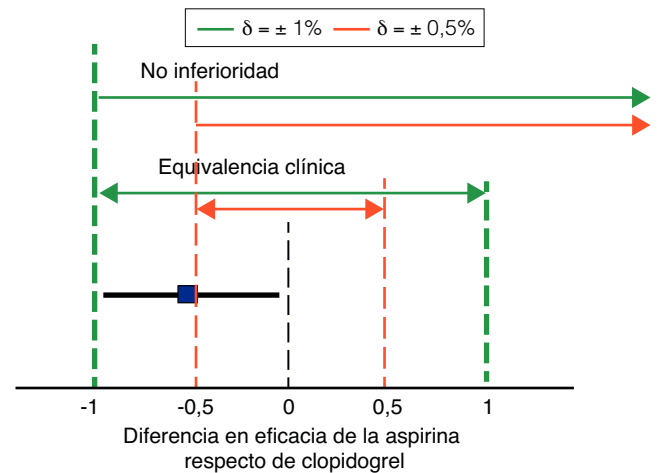
*Aspectos éticos: el uso de placebo es actualmente inaceptable en la comparación de fármacos biológicos*

La mayoría de los estudios en curso contienen un grupo tratado con placebo a pesar de que incluyen a pacientes con enfermedad moderada/grave en los que los biológicos son el estándar de tratamiento<sup>3-8</sup>. Los riesgos inherentes a dejar a pacientes con enfermedad grave sin tratamiento efectivo durante periodos que pueden alcanzar las 12 semanas en EII, las 16 semanas en dermatología o hasta 54 semanas en reumatología resultan, en opinión de los miembros del foro, inaceptables desde el punto de vista ético. Parece difícilmente justificable que las agencias reguladoras acepten, propongan e incluso exijan todavía estudios contra placebo para los nuevos fármacos biológicos. Más si se tiene en cuenta que para evaluar fármacos biosimilares las agencias solicitan estudios de no inferioridad comparando el biosimilar con su molécula original sin necesidad de incluir un grupo placebo. También parece poco razonable que los comités éticos acepten estudios frente a placebo cuando el estándar de tratamiento en pacientes con enfermedad moderada a grave es un fármaco biológico y el uso de un tratamiento subóptimo puede tener consecuencias negativas sobre la evolución de la enfermedad.

### Características de los estudios de equivalencia y no inferioridad

Clásicamente, los ensayos clínicos comparativos han utilizado la superioridad como criterio de comparación estadística directa entre distintos fármacos. Sin embargo, este planteamiento puede no ser el óptimo para la comparación entre biológicos o para nuevos biosimilares. En muchas ocasiones —y por definición, en el caso de los biosimilares—, los nuevos fármacos no pretenden ser más eficaces que el fármaco estándar con el que se comparan y a lo que se aspira es a demostrar que su eficacia o seguridad son comparables con los fármacos disponibles. Además, los estudios de no inferioridad no requieren un brazo de tratamiento con placebo, lo que reduce el riesgo que asume el paciente que participa en el estudio. La publicación de unas recomendaciones específicas dentro del *CONSORT statement*<sup>27,28</sup> muestra la importancia creciente<sup>29</sup> de los estudios de no inferioridad.

Una característica fundamental de los estudios de no inferioridad es la necesidad de establecer «a priori» qué diferencia ( $\delta$ ) en eficacia o seguridad puede considerarse clínicamente relevante o significativa, y solo después se puede realizar el análisis estadístico<sup>30</sup>. Así, en los estudios de no inferioridad no ocurre lo mismo que en los estudios comparativos convencionales, puesto que cuando estos incluyen un número muy elevado de pacientes, pequeñas diferencias en eficacia pueden resultar a la vez estadísticamente significativas e irrelevantes desde un punto de vista práctico. Un ejemplo es lo que sucedió con el estudio CAPRIE, que comparó los efectos de clopidogrel con ácido acetil-salicílico (AAS) en 19.185 pacientes seguidos durante 2 años. Se observó una tasa de eventos cardiovasculares del 5,83% por persona y año con AAS y del 5,32% con clopidogrel. La reducción del riesgo de eventos cardiovasculares fue del 0,51% por persona y año, estadísticamente significativa ( $p = 0,043$ ) pero de una relevancia clínica



**Figura 2.** Márgenes de no inferioridad y equivalencia clínica. La figura muestra los intervalos de no inferioridad y equivalencia para 2 valores de la diferencia en eficacia entre aspirina y clopidogrel que puede considerarse clínicamente relevante ( $\delta$ ) en líneas discontinuas. El cuadrado y la línea gruesa horizontal muestran la diferencia entre aspirina y clopidogrel en el estudio CAPRIE y su IC del 95%. Así con un  $\delta$  de  $\pm 1\%$  podríamos considerar que aspirina es tanto no inferior como equivalente a clopidogrel (líneas gruesas discontinuas). Por el contrario, con una estimación de  $\delta$  más restrictiva (0,5%, líneas finas discontinuas verticales), aspirina no cumpliría ni criterios de equivalencia ni de no inferioridad.

cuestionable<sup>31</sup>. Utilizaremos este estudio como ejemplo a lo largo de este apartado.

En los estudios de no inferioridad, debe calcularse la diferencia de eficacia entre el tratamiento evaluado y el tratamiento estándar, y se determina el intervalo de confianza del 95% (IC del 95%) de dicha diferencia. Si el IC de la diferencia está dentro del intervalo definido por  $\pm \delta$ , se considera que ambos tratamientos son equivalentes. Si el límite inferior del IC de la diferencia está por encima del valor de  $-\delta$ , consideramos que el fármaco evaluado cumple criterios de no inferioridad. Por ejemplo, imaginemos que en el caso del estudio CAPRIE consideramos clopidogrel como tratamiento estándar y AAS como tratamiento evaluado (fig. 2). Podemos utilizar una calculadora online para establecer el IC del 95% de la diferencia entre los 2 fármacos<sup>32</sup>, que es de  $-0,51\% \pm 0,48\%$ . Por tanto, el IC del 95% iría del  $-0,99\%$  al  $-0,02\%$ . Si asumimos que la máxima diferencia no significativa desde un punto de vista clínico es, por ejemplo, del  $\pm 1\%$  o superior, el AAS cumpliría criterios tanto de equivalencia clínica como de no inferioridad. Si, por el contrario, disminuimos el valor  $\delta$  a 0,5%, la aspirina no cumpliría criterios ni de equivalencia clínica ni de no inferioridad. Vemos, por tanto, una de las características de los estudios de no inferioridad: es fundamental seleccionar adecuadamente el intervalo numérico de la «relevancia» clínica, ya que el resultado de la comparación dependerá totalmente del intervalo  $\delta$  que elijamos<sup>33</sup>.

*Cómo establecer la máxima diferencia en eficacia que no se considera clínicamente relevante ( $\delta$ )*

Sin existir una fórmula para el cálculo de  $\delta$  aceptada universalmente<sup>33</sup>, sí hay una serie de factores definidos que deben tenerse en cuenta a la hora de determinar la mínima diferencia clínicamente relevante en un determinado parámetro de eficacia o seguridad. En el caso de los biológicos estos serían:

- El tipo de parámetro de eficacia evaluado: así, si se evalúa mortalidad, las diferencias que se han considerado clínicamente relevantes en los distintos estudios son habitualmente muy

pequeñas, entre el 0,4 y el 1%, en porcentajes absolutos. En otros parámetros menos «duros» —p. ej., la remisión o la respuesta clínica en el caso de la EII o la tasa de pacientes con respuesta moderada en AR se considerarían razonables márgenes más amplios.

- b. *La tasa de respuesta y la diferencia observada entre el tratamiento estándar y el placebo:* en ocasiones, se ha recomendado la mitad del efecto del tratamiento estándar vs. placebo. También se ha recomendado determinar el intervalo  $\delta$  en función de la dispersión de los resultados de los fármacos a comparar, por ejemplo, mediante el uso de 0,5 desviaciones estándar de la variable que se evalúa. Finalmente, también se han establecido porcentajes variables del intervalo  $\delta$  en función del grado de eficacia del fármaco<sup>29</sup>. En cualquier caso, el intervalo  $\delta$  debe ser suficientemente restrictivo para que el placebo no se incluya como no inferior o equivalente.
- c. *En caso de comparaciones múltiples, los valores  $\delta$  deben calcularse respecto del mismo tratamiento estándar,* que debe ser el más efectivo en términos absolutos. En caso contrario, podría producirse lo que se denomina como un efecto de arrastre, de modo que la comparación sucesiva 2 a 2 de fármacos progresivamente menos efectivos podría llevar a la aceptación de fármacos no superiores al placebo como equivalentes al estándar inicial.

#### ¿Cuál debe ser el valor $\delta$ en enfermedades reumáticas?

Actualmente, existen 9 agentes biológicos con indicaciones en la ficha técnica para el tratamiento de diversas artropatías inflamatorias crónicas. Para la AR, se ha aprobado un antagonista del receptor de interleucina (IL)-1 (anakinra), 5 agentes inhibidores del TNF- $\alpha$  (adalimumab, certolizumab, etanercept, golimumab e infliximab), un agente deplecionante de células B CD20+ (rituximab), un inhibidor del receptor soluble de IL-6 (tocilizumab) y un inhibidor de moléculas de coestimulación (abatacept).

En la AR, la eficacia terapéutica se evalúa habitualmente mediante el uso de índices combinados de respuesta que incluyen recuentos articulares, reactantes de fase aguda y valoraciones globales de la actividad de la enfermedad por parte del paciente y del médico. Los índices de respuesta y actividad más utilizados son las respuestas American College of Rheumatology (ACR) 20, 50 y 70, que miden el porcentaje de mejoría respecto a la situación basal, sin tener en cuenta la actividad final de la enfermedad, el índice Disease Activity Score (DAS) 28 y el Simplified Disease Activity Index (SDAI). Estos últimos son medidas absolutas de la actividad de la enfermedad definiendo la respuesta como el porcentaje de pacientes que alcanzan un determinado estado de la enfermedad (baja actividad o remisión). Los desenlaces en los ensayos clínicos se cuantifican como el porcentaje de pacientes que alcanzan una determinada respuesta; en el caso del ACR, el porcentaje de pacientes que alcanzan un determinado estado de la enfermedad (baja actividad o remisión) o como el descenso medio en el valor del DAS28 o el SDAI.

No existe un valor de  $\delta$  universalmente aceptado como clínicamente relevante ni para el porcentaje de pacientes que alcanza una determinada respuesta, por ejemplo, un ACR20 o un DAS28 inferior a 3,2 (que se considera un estado de baja actividad de la AR) ni para la reducción mínima clínicamente relevante del valor del DAS28 o del SDAI. No obstante, recientes ensayos clínicos de elevada calidad metodológica pueden orientar hacia qué  $\delta$  sería razonable elegir como dintel en las comparaciones entre agentes biológicos. En el estudio ADACTA, un ensayo clínico aleatorizado de comparación directa entre adalimumab y tocilizumab en monoterapia en pacientes con AR con respuesta inadecuada a metotrexato (MTX), los autores consideraron que la diferencia relevante entre grupos debería ser de al menos 0,6 unidades en el DAS28<sup>34</sup>. Por otra parte, en el estudio AMPLE, un ensayo clínico aleatorizado de

comparación directa entre adalimumab y abatacept combinados con MTX, también en pacientes con AR con respuesta inadecuada a MTX, los autores asumieron —sin una base científica clara— que el margen de no inferioridad en el porcentaje de pacientes que alcanzaran una respuesta ACR 20 se situaría en el 12% entre grupos<sup>35</sup>. Sin embargo, no se puede utilizar un delta válido en ACR20 para comparar ACR50 y ACR70. El porcentaje de respuesta ACR 50 suele ser del 40% de los pacientes en tratamiento combinado respecto al 20% de los pacientes con placebo o MTX, utilizar un 15% de diferencia en la tasa de respuesta, puede incluir la respuesta placebo y, por tanto, no sería un valor de  $\delta$  válido. En el caso de ACR70 el margen es todavía más estrecho, 20% vs. 5%-10%, por lo que los valores de  $\delta$  tendrían que ser aún inferiores.

Para la EA y la Aps, se ha aprobado el uso de 4 agentes inhibidores del TNF- $\alpha$  (adalimumab, etanercept, golimumab e infliximab). En el caso de la EA, como variable de desenlace en los ensayos clínicos se utiliza habitualmente el porcentaje de pacientes que alcanzan un índice combinado ASDAS de enfermedad inactiva ( $\leq 1,3$ ) o de enfermedad con actividad baja ( $\leq 2,1$ ) y se considera que un  $\delta \geq 1,1$  es un cambio clínicamente relevante<sup>36</sup>, por lo que este valor podría considerarse una referencia. Otro índice de actividad utilizado ampliamente en la EA es el BASDAI. Se considera que un BASDAI  $\leq 2$  refleja actividad mínima, mientras que un BASDAI  $\leq 4$  es considerado como de baja actividad<sup>37</sup>. Se ha propuesto que un  $\delta$  clínicamente relevante sería una variación superior a una unidad absoluta o al 22,5% del valor basal del BASDAI<sup>38</sup>.

En las formas axiales de la Aps, se asumen los índices de respuesta que se utilizan en la EA<sup>39</sup>, mientras que en las formas periféricas poliarticulares se asumen los índices de respuesta utilizados en la AR<sup>39</sup>. Actualmente, no existe un índice de respuesta comúnmente aceptado para valorar la respuesta al tratamiento de las formas oligoarticulares periféricas de la Aps, por lo que es difícil indicar recomendaciones sobre qué  $\delta$  sería el más adecuado en estos casos.

#### ¿Cuál debe ser el valor $\delta$ en psoriasis?

En dermatología, los fármacos actualmente autorizados para psoriasis son 3 fármacos anti-TNF: adalimumab, etanercept e infliximab, y un inhibidor de la p40, proteína que comparten la IL-12 y la IL-23, ustekinumab. La variable principal de evaluación más comúnmente utilizada en ensayos clínicos en psoriasis es la respuesta Psoriasis Area and Severity Index (PASI)-75, y como variables secundarias PASI-90 y PASI-100, valores estos últimos que se consideran indicativos de remisión<sup>40,41</sup>.

Respecto de cuál puede ser el valor del incremento de estos parámetros que puede considerarse clínicamente irrelevante, hay que tener en cuenta, en primer lugar, la eficacia sobre placebo de los diversos fármacos. Así, en distintos metaanálisis que incluyeron los estudios pivotales, las diferencias en el PASI-75 respecto a placebo del fármaco menos efectivo, etanercept a dosis bajas, fueron del 31 al 45%<sup>42</sup>. Para el resto de los fármacos, las diferencias con placebo oscilaron entre el 40 y el 78%.

No existen datos en la literatura sobre la diferencia mínima clínicamente relevante en la variable PASI-75 entre 2 medicamentos biológicos para el tratamiento de pacientes con psoriasis en placas. El único dato disponible hasta el momento viene del único ensayo clínico comparando 2 biológicos publicado. El estudio ACCEPT<sup>43</sup> comparó ustekinumab con etanercept en el tratamiento de la psoriasis moderada/severa; en este estudio, aunque no se estableció un valor  $\delta$ , el cálculo de la muestra se realizó sobre una diferencia esperada en términos de PASI-75 entre ustekinumab y etanercept del 14%. El estudio detectó una diferencia entre ustekinumab a dosis bajas y etanercept del 10,7%. Los autores interpretan esta diferencia como claramente relevante desde el punto de vista clínico. Finalmente, en otro ensayo clínico que comparó adalimumab con MTX<sup>44</sup>, la diferencia esperada en términos de PASI-75 entre ambos



tratamientos fue del 20% de pacientes que alcanzaban dicho grado de respuesta.

Por tanto, cuando se utiliza la respuesta PASI-75 como variable principal de evaluación, los escasos datos disponibles indican que un valor  $\delta$  apropiado debería moverse entre el 5 y el 15%, y es probable que debiera estar ligeramente por debajo del 10% observado en el estudio ACCEPT. Sin embargo, faltan argumentos sólidos para defender una cifra concreta.

#### *¿Cuál debe ser el valor $\delta$ en la enfermedad inflamatoria intestinal?*

Las escalas que se han empleado más a menudo en los ensayos clínicos más recientes son el Crohn's Disease Activity Index (CDAI) para la enfermedad de Crohn<sup>45</sup> y el índice de Mayo para la colitis ulcerosa<sup>46</sup>. En el caso del CDAI, se considera que el paciente está en remisión cuando los valores disminuyen por debajo de 150 y como respuesta un descenso de entre 70-100 puntos<sup>47</sup>. En el caso del índice de Mayo, se considera remisión valores de 2 o inferiores y respuesta un descenso de al menos 3 puntos y un 30% de los valores iniciales<sup>48</sup>.

En el caso de la EII, las características de la enfermedad y el diseño de los estudios condicionan que las diferencias respecto de placebo en los estudios pivotaes sea menor que en otras patologías. De hecho, en los estudios iniciales de los fármacos aprobados para enfermedad de Crohn y colitis ulcerosa, esta oscila entre el 33 y el 7,2%<sup>48</sup>, siendo menor en los pacientes que ya habían recibido tratamiento previo con un fármaco biológico<sup>49</sup>.

En el estudio SONIC<sup>50</sup>, se compararon azatioprina, infliximab y la combinación de ambos en pacientes con enfermedad de Crohn que no habían recibido previamente inmunosupresores. Las tasas de remisión clínica a la semana 26 fueron del 30% para azatioprina, el 44,4% para infliximab y el 56,8% para el tratamiento combinado. Se interpretaron como claramente significativas las diferencias entre azatioprina e infliximab, así como las diferencias con el tratamiento combinado. Finalmente, un consenso internacional que realizó una revisión sistemática de todos los índices para la evaluación de la colitis ulcerosa<sup>46</sup> indica un valor de  $\delta$  para no inferioridad del 10%, aunque, sorprendentemente, no especifica para qué variable de evaluación.

En conclusión, en el caso de la EII, teniendo en cuenta el menor beneficio sobre placebo que se ha observado generalmente en los ensayos pivotaes, se podría indicar como orientativos unos valores de  $\delta$  de aproximadamente un 10% para respuesta clínica y valores ligeramente inferiores, entre el 5 y el 10%, para el caso de la remisión clínica.

### **Uso de revisiones sistemáticas, meta-análisis y comparaciones directas e indirectas para establecer equivalencia**

#### *Comparaciones directas*

Evidentemente, el método más eficaz para comparar 2 fármacos biológicos es la comparación directa dentro de un ensayo clínico aleatorizado. Sin embargo, como estos ensayos comparativos son, hasta el momento, excepcionales, no es posible realizar revisiones sistemáticas y metaanálisis clásicos.

#### *Comparaciones indirectas*

Aunque son el único recurso disponible en ausencia de comparaciones directas, estos estudios tienen importantes limitaciones metodológicas.

En primer lugar, una asunción fundamental de los estudios indirectos es la consistencia de la evidencia o la comparabilidad de los estudios incluidos. Diferencias en el diseño o en la población evaluada pueden causar importantes sesgos y condicionar que los

resultados de los estudios no puedan ser comparados con estas técnicas<sup>51</sup>. Un problema importante y habitual es el cambio de la población de los estudios a lo largo del tiempo. Así, los pacientes que entran en estudios recientes tienen mayor probabilidad de haber fracasado a varios tratamientos biológicos previos, lo que los hace más refractarios a cualquier nuevo tratamiento y condiciona una peor respuesta. Por ello, se recomienda que en toda comparación indirecta se expliciten y discutan extensivamente todas aquellas diferencias en el diseño de estudios incluidos que puedan sesgar los resultados del análisis<sup>52</sup>. También existe la posibilidad de intentar controlar estas diferencias entre estudios mediante técnicas de metaregresión<sup>53</sup>. Sin embargo, en la práctica, esta posibilidad está limitada por el reducido número de estudios y por el riesgo a incurrir en sesgos ecológicos.

En segundo lugar, las comparaciones indirectas reducen marcadamente la potencia de las comparaciones y requieren muestras mucho mayores que las comparaciones directas<sup>24,54</sup>. Así, una reciente revisión comparando métodos directos e indirectos<sup>55</sup> evaluó 39 comparaciones entre intervenciones terapéuticas en las que las comparaciones directas encontraban una diferencia estadísticamente significativa. En 14 de las 39 comparaciones la significación estadística desaparecía cuando se realizaba un análisis combinando las estimaciones directas e indirectas. De manera similar, en una evaluación previa, los mismos autores evaluaron 19 comparaciones directas que encontraban diferencias significativas entre 2 intervenciones terapéuticas; únicamente en 9 de ellas las comparaciones indirectas detectaron diferencias significativas<sup>24,56</sup>.

Las consecuencia práctica de esta limitada potencia es que es muy difícil que, a través de estudios indirectos, 2 fármacos resulten significativamente distintos<sup>24,54-56</sup>. Si esto ocurre (esto es, si un fármaco es significativamente superior al otro), el resultado es razonablemente fiable. Sin embargo, el hecho de que no se detecten diferencias no demuestra la equivalencia entre fármacos. Debido a la naturaleza acentuadamente conservadora de los análisis indirectos, estos requieren diferencias muy marcadas entre tratamientos y un número muy elevado de pacientes para detectar una diferencia significativa. Este aspecto es muy importante, ya que algunos estudios comparando fármacos biológicos han interpretado erróneamente esta falta de potencia estadística de las comparaciones indirectas como prueba de que los fármacos son equivalentes<sup>20</sup>.

Al intenso debate sobre la fiabilidad de las comparaciones indirectas contribuye, y no poco, el hecho de que estudios que valoran el efecto de comparaciones indirectas sobre determinadas enfermedades lleguen a conclusiones divergentes. Como ejemplo, en el caso de la psoriasis, el estudio de Signorovitch et al.<sup>57</sup> concluye que adalimumab es superior a etanercept, mientras que un estudio recientemente publicado que plantea la misma comparación concluye que todos los biológicos son equivalentes<sup>21</sup>. De hecho, la significación estadística e incluso la dirección del efecto pueden variar en función del método que se emplee para la comparación indirecta. Así, O'Regan et al. estudiaron a 51 comparaciones indirectas entre fármacos mediante 2 métodos estadísticos de análisis diferentes. De las 51 comparaciones, encontraron 3 en que con un método la diferencia era significativa y con el otro no, 6 en los que la dirección del efecto era distinta dependiendo del método utilizado y 9 donde los IC variaban ampliamente en función del método<sup>58</sup>. Por este motivo, se considera recomendable en las comparaciones indirectas analizar no solamente la variable principal, sino también las variables secundarias de manera conjunta (p. ej., ACR20, ACR50 y ACR70 en estudios sobre AR). Los resultados de un análisis serán más consistentes si las diferencias a favor de un fármaco se mantienen en los distintos parámetros de medida.

Finalmente, se han detectado importantes deficiencias en la calidad de los estudios indirectos<sup>52,59</sup>; Donegan et al.<sup>52</sup> propusieron una serie de parámetros específicos para evaluar la calidad de estos estudios. Adicionalmente, sería recomendable que, como cualquier

otra revisión sistemática o metaanálisis, las comparaciones indirectas se realicen de acuerdo con las recomendaciones PRISMA<sup>60,61</sup>.

Vistos los riesgos de sesgo, resulta comprensible que muchos autores desconfíen de la fiabilidad de las comparaciones indirectas. Todo esto ha llevado a que la *International Society for Pharmacoeconomic and Outcomes Research*, en un intento de mejorar la fiabilidad de estos estudios, designara un grupo de trabajo específico que ha realizado recomendaciones para la evaluación, la interpretación y la realización de comparaciones indirectas<sup>62,63</sup>. En todo caso, debido a las importantes limitaciones del método, los resultados obtenidos mediante comparaciones indirectas deberán considerarse como datos exploratorios, útiles para generar hipótesis susceptibles de posterior confirmación pero nunca como una prueba definitiva ni de superioridad ni, mucho menos, de equivalencia<sup>52,55</sup>. De hecho, las agencias reguladoras en ningún momento han considerado las comparaciones indirectas como métodos adecuados para evaluar fármacos biosimilares.

## Conclusiones

La presente revisión aporta elementos de juicio para la comparación de fármacos biológicos. Las conclusiones principales del foro han sido:

- Tanto desde el punto de vista científico como desde el punto de vista ético, consideramos recomendable modificar las directivas regulatorias en el sentido de favorecer decididamente los estudios aleatorizados de no inferioridad, comparando cara a cara los nuevos biológicos con los actuales estándares de tratamiento, evitando los ensayos frente a placebo.
- Estos estudios aportarán datos fiables sobre la eficacia y la seguridad comparativas de los distintos fármacos de los que actualmente carecemos, dada la escasa fiabilidad y las importantes limitaciones metodológicas de las comparaciones indirectas.
- Un elemento clave en este proceso será la determinación por consenso con la participación de las agencias reguladoras, las sociedades científicas, la industria farmacéutica y las autoridades sanitarias de las diferencias clínicas que deben considerarse relevantes en cada una de las patologías evaluadas.

## Responsabilidades éticas

**Protección de personas y animales.** Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

**Confidencialidad de los datos.** Los autores declaran que en este artículo no aparecen datos de pacientes.

**Derecho a la privacidad y consentimiento informado.** Los autores declaran que en este artículo no aparecen datos de pacientes.

## Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses en el contenido de este artículo.

## Anexo.

**Dermatología:** Carlos Ferrandiz, Servicio de Dermatología, Hospital Universitario Germans Trias i Pujol, Universidad Autónoma de Barcelona; Hugo Vázquez Veiga, Servicio de Dermatología, Hospital de Conxo, Complejo Hospitalario Universitario de Santiago de Compostela, A Coruña. **Farmacología clínica:** Juan Vicente Esplugues.

**Reumatología:** José Luís Andreu, Servicio de Reumatología, Hospital Universitario Puerta de Hierro, Majadahonda, Madrid; Antoni Gómez, Servicio de Reumatología, Hospital de Sabadell, Institut Universitari Parc Taulí, Universitat Autònoma de Barcelona. **Gastroenterología:** Fernando Gomollón, Servicio de Aparato Digestivo, Hospital Clínico «Lozano Blesa», Zaragoza; Xavier Calvet. **Estadística:** David Suarez, Unidad de Epidemiología y Evaluación, Fundació Parc Taulí, Hospital Parc Taulí, Universitat Autònoma de Barcelona.

## Bibliografía

1. Emery P, Dorer T. Optimising treatment in rheumatoid arthritis: a review of potential biological markers of response. *Ann Rheum Dis*. 2011;70:2063–70.
2. Rutgeerts P, Van Deventer S, Schreiber S. Review article: The expanding role of biological agents in the treatment of inflammatory bowel disease –focus on selective adhesion molecule inhibition. *Aliment Pharmacol Ther*. 2003;17:1435–50.
3. Loftus EV, Feagan BG, Colombel JF, Rubin DT, Wu EQ, Yu AP, et al. Effects of adalimumab maintenance therapy on health-related quality of life of patients with Crohn's disease: Patient-reported outcomes of the CHARM trial. *Am J Gastro*. 2008;103:3132–41.
4. Revicki D, Willian MK, Saurat JH, Papp KA, Ortonne JP, Sexton C, et al. Impact of adalimumab treatment on health-related quality of life and other patient-reported outcomes: Results from a 16-week randomized controlled trial in patients with moderate to severe plaque psoriasis. *Br J Dermatol*. 2008;158:549–57.
5. Van der Heijde DM, Revicki DA, Gooch KL, Wong RL, Kupper H, Harnam N, et al. Physical function, disease activity, and health-related quality-of-life outcomes after 3 years of adalimumab treatment in patients with ankylosing spondylitis. *Art Res Ther*. 2009;11:R124. <http://dx.doi.org/10.1186/ar2790>.
6. Van Vollenhoven RF, Cifaldi MA, Ray S, Chen N, Weisman MH. Improvement in work place and household productivity for patients with early rheumatoid arthritis treated with adalimumab plus methotrexate: Work outcomes and their correlations with clinical and radiographic measures from a randomized controlled trial companion study. *Art Care Res*. 2010;62:226–34.
7. Colombel JF, Rutgeerts P, Reinisch W, Esser D, Wang Y, Lang Y, et al. Early mucosal healing with infliximab is associated with improved long-term clinical outcomes in ulcerative colitis. *Gastroenterology*. 2011;141:1194–201.
8. Furst DE, Keystone EC, So AK, Braun J, Breedveld FC, Burmester GR, et al. Updated consensus statement on biological agents for the treatment of rheumatic diseases, 2012. *Ann Rheum Dis*. 2013;72 Suppl 2:ii2–34.
9. Bennett CL, Luminari S, Nissenson AR, Tallman MS, Klinge SA, McWilliams N, et al. Pure red-cell aplasia and epoetin therapy. *New Eng J Med*. 2004;351:1403–8.
10. Fox A. Biosimilar medicines –new challenges for a new class of medicine. *J Biopharm Stat*. 2010;20:3–9.
11. Health Canada. Summary basis of decision for Inflectra [consultado 29 Abril 2014]. Disponible en: <http://www.hc-sc.gc.ca/dhp-mps/prodpharma/sbd-smd/drug-med/sbd.smd.2014.inflectra.159493-eng.php>
12. Weise M, Bielsky MC, de Smet K, Ehmann F, Ekman N, Narayanan G, et al. Biosimilars-why terminology matters. *Nat Biotechnol*. 2011;29:690–3.
13. Scheinberg MA, Kay J. The advent of biosimilar therapies in rheumatology –«O brave new world». *Nat Rev Rheumatol*. 2012;8:430–6.
14. Danese S, Gomollon F. ECCO position statement: The use of biosimilar medicines in the treatment of inflammatory bowel disease (IBD). *JCC*. 2013;7:586–9.
15. Chow SC, Wang J, Endrenyi L, Lachenbruch PA. Scientific considerations for assessing biosimilar products. *Stat Med*. 2013;32:370–81.
16. Chow SC, Endrenyi L, Lachenbruch PA. Comments on the FDA draft guidance on biosimilar products. *Stat Med*. 2013;32:364–9.
17. Chow SC. Assessing biosimilarity and interchangeability of biosimilar products. *Stat Med*. 2013;32:361–3.
18. Chen SL, Wu SL, Huang LJ, Huang JB, Chen SH. A global comparability approach for biosimilar monoclonal antibodies using LC-tandem MS based proteomics. *J Pharm Biomed Anal*. 2013;80:126–35.
19. Bourgoin AF, Nuskey B. White paper: An outlook on U.S. biosimilar competition. *Drugs Today (Barc)*. 2013;49:399–410.
20. Gallego-Galisteo M, Villa-Rubio A, Alegre-del Rey E, Marquez-Fernandez E, Ramos-Baez JJ. Indirect comparison of biological treatments in refractory rheumatoid arthritis. *J Clin Pharm Ther*. 2012;37:301–7.
21. Galvan-Banqueri M, Marin Gil R, Santos Ramos B, Bautista Paloma FJ. Biological treatments for moderate-to-severe psoriasis: Indirect comparison. *J Clin Pharm Ther*. 2013;38:121–30.
22. Migliore A, Broccoli S, Bizzi E, Lagana B. Indirect comparison of the effects of anti-TNF biological agents in patients with ankylosing spondylitis by means of a mixed treatment comparison performed on efficacy data from published randomised, controlled trials. *J Med Econ*. 2012;15:473–80.
23. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Stat Med*. 2007;26:1237–54.
24. Glenny AM, Altman DG, Song F, Sakarovich C, Deeks JJ, D'Amico R, et al. Indirect comparisons of competing interventions. *Health Technol Assess*. 2005;9:1–134, iii–iv.

25. Catala-Lopez F. Evaluation of comparative treatment effects using indirect comparisons. *Rev Esp Cardiol*. 2013;66:156–7.
26. Catala-Lopez F, Tobias A. Síntesis de la evidencia clínica y metaanálisis en red con comparaciones indirectas. *Med Clin (Barc)*. 2013;140:182–7.
27. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. *JAMA*. 2006;295:1152–60.
28. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. Reporting of noninferiority and equivalence randomized trials: Extension of the CONSORT 2010 statement. *JAMA*. 2012;308:2594–604.
29. Delgado-Sánchez O, Puigventós-Latorre F, Pinteño-Blanco M, Ventayol-Bosch P. Equivalencia terapéutica: concepto y niveles de evidencia. *Med Clin (Barc)*. 2007;129:9.
30. Huson L. Statistical assessment of superiority, equivalence and non-inferiority in clinical Trials. *Clinical Research Focus* [Internet]. 2001 [consultado mayo 2014]. Disponible en: <http://www.ais.up.ac.za/med/tnm800/EssentialTNM800/DayTwo/InterventionResearch/equivalence.pdf>
31. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). CAPRIE Steering Committee. *Lancet*. 1996;348:1329–1339.
32. Lowry R. The confidence interval for the difference between two independent proportions. [consultado mayo 2014]. Disponible en: [http://www.vassarstats.net/prop2\\_ind.html](http://www.vassarstats.net/prop2_ind.html)
33. Kaul S, Diamond GA. Good enough: A primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med*. 2006;145:62–9.
34. Gabay C, Emery P, van Vollenhoven R, Dikranian A, Alten R, Pavelka K, et al. Tocilizumab monotherapy versus adalimumab monotherapy for treatment of rheumatoid arthritis (ADACTA): A randomised, double-blind, controlled phase 4 trial. *Lancet*. 2013;381:1541–50.
35. Weinblatt ME, Schiff M, Valente R, van der Heijde D, Citera G, Zhao C, et al. Head-to-head comparison of subcutaneous abatacept versus adalimumab for rheumatoid arthritis: Findings of a phase IIIb, multinational, prospective, randomized study. *Art Rheum*. 2013;65:28–38.
36. Machado P, Landewe R, Lie E, Kvien TK, Braun J, Baker D, et al. Ankylosing Spondylitis Disease Activity Score (ASDAS): Defining cut-off values for disease activity states and improvement scores. *Ann Rheum Dis*. 2011;70:47–53.
37. Juanola X, Zarco P, Sanz J, Muñoz S, Mulero J, Linares L, et al. Documento SER de consenso sobre el uso de terapias biológicas en la espondilitis anquilosante y otras espondiloartritis, excepto la artritis psoriásica. *Reumatol Clin*. 2011;7:10.
38. Pavy S, Brophy S, Calin A. Establishment of the minimum clinically important difference for the bath ankylosing spondylitis indices: A prospective study. *J Rheumatol*. 2005;32:80–5.
39. Fernández-Sueiro J, Juanola XJC, Torre-Alonso J, Garcia-de-Vicuña R, Queiró R, Ariza Ariza R, et al. Documento SER de consenso sobre el uso de terapias biológicas en la artritis psoriásica. *Reumatol Clin*. 2011;7:9.
40. Puzenat E, Bronsard V, Prey S, Gourraud PA, Aractingi S, Bagot M, et al. What are the best outcome measures for assessing plaque psoriasis severity? A systematic review of the literature. *JEADV*. 2010;24 Suppl 2:10–6.
41. Robinson A, Kardos M, Kimball AB. Physician Global Assessment (PGA) and Psoriasis Area and Severity Index (PASI): Why do both? A systematic analysis of randomized controlled trials of biologic agents for moderate to severe plaque psoriasis. *J Am Acad Dermatol*. 2012;66:369–75.
42. Lucka TC, Pathirana D, Sammain A, Bachmann F, Rosumeck S, Erdmann R, et al. Efficacy of systemic therapies for moderate-to-severe psoriasis: a systematic review and meta-analysis of long-term treatment. *JEADV*. 2012;26:1331–44.
43. Young MS, Horn EJ, Cather JC. The ACCEPT study: Ustekinumab versus etanercept in moderate-to-severe psoriasis patients. *Exp Rev Clin Immunol*. 2011;7:9–13.
44. Saurat JH, Stingl G, Dubertret L, Papp K, Langley RG, Ortonne JP, et al. Efficacy and safety results from the randomized controlled comparative study of adalimumab vs. methotrexate vs. placebo in patients with psoriasis (CHAMPION). *B J Dermatol*. 2008;158:558–66.
45. Best WR, Beckett JM, Singleton JW, Kern Jr F. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. *Gastroenterology*. 1976;70:439–44.
46. D'Haens G, Sandborn WJ, Feagan BG, Geboes K, Hanauer SB, Irvine EJ, et al. A review of activity indices and efficacy end points for clinical trials of medical therapy in adults with ulcerative colitis. *Gastroenterology*. 2007;132:763–86.
47. Colombel JF, Sandborn WJ, Rutgeerts P, Enns R, Hanauer SB, Panaccione R, et al. Adalimumab for maintenance of clinical response and remission in patients with Crohn's disease: The CHARM trial. *Gastroenterology*. 2007;132:52–65.
48. Sandborn WJ, van Assche G, Reinisch W, Colombel JF, D'Haens G, Wolf DC, et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology*. 2012;142:257–65, e1–3.
49. Reinisch W, Sandborn WJ, Hommes DW, D'Haens G, Hanauer S, Schreiber S, et al. Adalimumab for induction of clinical remission in moderately to severely active ulcerative colitis: Results of a randomised controlled trial. *Gut*. 2011;60:780–7.
50. Colombel JF, Sandborn WJ, Reinisch W, Mantzaris GJ, Kornbluth A, Rachmilewitz D, et al. Infliximab, azathioprine, or combination therapy for Crohn's disease. *N Engl J Med*. 2010;362:1383–95.
51. Coory M, Jordan S. Frequency of treatment-effect modification affecting indirect comparisons: A systematic review. *PharmacoEconomics*. 2010;28:723–32.
52. Donegan S, Williamson P, Gamble C, Tudur-Smith C. Indirect comparisons: A review of reporting and methodological quality. *PLoS One*. 2010;5:e11054.
53. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med*. 2009;28:1861–81.
54. Mills EJ, Ghement I, O'Regan C, Thorlund K. Estimating the power of indirect comparisons: A simulation study. *PLoS One*. 2011;6:e16237.
55. Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, et al. Inconsistency between direct and indirect comparisons of competing interventions: Meta-epidemiological study. *BMJ*. 2011;343:d4909.
56. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: Empirical evidence from published meta-analyses. *BMJ*. 2003;326:472.
57. Signorovitch JE, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao Y, et al. Comparative effectiveness without head-to-head trials: A method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. *PharmacoEconomics*. 2010;28:935–45.
58. O'Regan C, Ghement I, Eyawo O, Guyatt GH, Mills EJ. Incorporating multiple interventions in meta-analysis: An evaluation of the mixed treatment comparison with the adjusted indirect comparison. *Trials*. 2009;10:86.
59. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: Survey of published systematic reviews. *BMJ*. 2009;338:b1147.
60. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*. 2009;339:b2700.
61. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*. 2009;339:b2535.
62. Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 2. *Value Health*. 2011;14:429–37.
63. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1. *Value Health*. 2011;14:417–28.