



Special article

How to Compare Biologic Drugs[☆]



Xavier Calvet,^{a,b,*} Juan Vicente Esplugues^{b,c,d}, on behalf of the Forum for the Adecuación de las Técnicas de Evaluación de Nuevos Anticuerpos monoclonales (Foro ATENAS)[◇]

^a Servicio de Aparato Digestivo, Corporació Sanitària Universitària Parc Taulí, Universitat Autònoma de Barcelona, Sabadell, Spain

^b CIBERehd, Madrid, Spain

^c Departamento de Farmacología, Facultad de Medicina, Universidad de Valencia, Valencia, Spain

^d Farmacología Clínica, Hospital Universitario Doctor Peset, Valencia, Spain

ARTICLE INFO

Article history:

Received 27 March 2014

Accepted 15 June 2014

Available online 3 August 2014

Keywords:

Biologic drugs

Indirect comparisons

Non inferiority studies

Palabras clave:

Fármacos biológicos

Comparaciones indirectas

Estudios de no inferioridad

ABSTRACT

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.

¿Cómo comparar fármacos biológicos?

RESUMEN

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.

Introduction

Biological drugs have constituted a therapeutic revolution in rheumatic diseases as rheumatoid arthritis, (RA), ankylosing

spondylitis (AS), psoriatic arthritis (PsA)–inflammatory bowel disease (IBD), Crohn's disease and ulcerative colitis, and in certain skin diseases^{1,2}–moderate or severe psoriasis. Not only has this group of drugs demonstrated their effectiveness on symptoms, but can also in modifying the natural history of these diseases, preventing complications and the associated disability.^{3–8}

Unlike traditional drugs obtained by chemical synthesis, biological molecules are protein based and generated by living cells. Their size and molecular weight are variable (from peptide chains to whole antibody molecules), and may be very high.⁹ Although, by definition, there are no two biological molecules 100% identical, differences between family members–p, e.g., anti-TNF sharing a therapeutic target, may be important. The differences lie in their

[☆] Please cite this article as: Calvet X, Esplugues JV, en representación del Foro para la Adecuación de las Técnicas de Evaluación de Nuevos Anticuerpos monoclonales (Foro ATENAS). ¿Cómo comparar fármacos biológicos? Reumatol Clin. 2014;10:353–359.

* Corresponding author.

E-mail address: xcalvet@tauli.cat (X. Calvet).

¹ ATENAS Forum components are listed in [Annex](#).

amino acid chain or—in the case of biosimilar drugs, which generally have a sequence identical to the original in drug-p-amino acid modifications, e.g., glycosylations or fucosylations of amino acids side chains after synthesis and thereby conditioning three-dimensional folding, which can cause variations in substrate affinity or the degree of immunogenicity and cause differences in efficacy or safety.¹⁰ Indeed, as an example, there is a significantly higher incidence of severe aplastic anemia associated with certain formulations of recombinant erythropoietin but not others⁹; recently, differences have also been seen in the fucosylation pattern of FcR1a affinity γ , and, in in vitro studies, the antibody-dependent cytotoxicity mediated by cells between infliximab and the biosimilar Inflectra®. The Canadian Agency for Drugs justified, on the basis of this data, that the approval granted to Inflectra® for rheumatic disease does not spread to IBD.¹¹

The market for biologics has increased rapidly in recent years. Moreover, after the expiration of the patent and the end of the period of data protection for innovative original drugs, biosimilar drugs have appeared on the market, a term understood as copies of biological drugs already approved where similar physicochemical, efficiency and safety features have been demonstrated after undertaking the necessary 12 comparisons. The definition emphasizes two aspects: *a*) the fact that they will never be equal to the original drug, hence the term “similar” as opposed to the concept of ‘identical’, which would apply to the comparison of a generic drug with respect to the original molecule and, *b*) in the case of biologics drugs, bioequivalence a concept that involves similar areas under the curve between the parent drug serum levels and the copy, used to demonstrate the therapeutic equivalence of generic drugs—which is not definite equivalence or criteria for classifying a biological copy and original as having the same efficacy and safety. A comprehensive assessment of each new drug is therefore required. Not only must we analyze the physico-chemical characteristics, but we also require careful clinical assessment of efficacy and safety to consider a given copy as biosimilar.¹²

The parameters to be determined in this evaluation are debated,^{9,13–19} although the overall orientation of regulatory agencies, and in particular, the European Drug Agency, has been to require randomized clinical trials comparing equivalence or non-inferiority of the efficacy and safety of biosimilar in relation to its original.¹⁰ This contrasts with the approval process for innovative biologics, where most require studies comparing the drug with a placebo control group. As a rule, studies of ‘equivalence’ or ‘non-inferiority’ seek to show that these terms apply for a new therapeutic drug against a known standard—the new drug is “equivalent” or “not less than the known drug”—and in most cases no placebo is used.

Furthermore, when comparing biologics, the difficulty that appears is that, until recently, there are no published clinical trials directly comparing the efficacy and safety of two biological drugs. This lack of direct comparisons has led to attempts to compare them using other methods of evidence-based medicine, specifically through indirect comparisons analysis.^{20–23} The simplest indirect methods are non-adjusted indirect comparisons. They consist in comparing the efficacy of two biologicals – which we will call A and B, using the efficiency of A in studies evaluating this drug, directly comparing the efficacy of drug B in their respecting studies of the same condition, without making any corrections. This method often gives incorrect results, therefore, its use is strongly discouraged.²⁴ A more correct alternative from the methodological point of view are adjusted indirect comparisons. In this case, we must have randomized trials comparing A and B vs the common comparator P (in the case of biological drugs, this is commonly a placebo). The effectiveness of A and B is compared through P in order to correct, at least partially, the differences between the populations of different studies. Indirect comparisons can be much more complicated, for

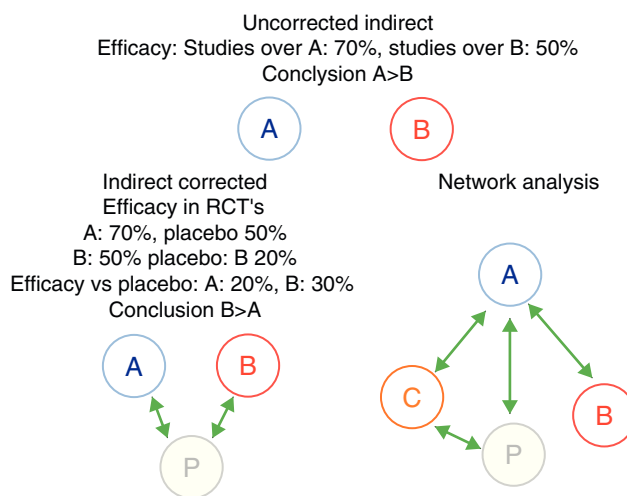


Fig. 1. Indirect comparisons, corrected, uncorrected or networked. An example is shown where the unadjusted indirect comparison between two drugs, A and B, is displayed, giving a totally different result, probably incorrect and the corrected result due to the different characteristics of the study population. Network analysis may include multiple comparisons between different agents (A–C) and/or with placebo. Uncorrected indirect efficacy: studies over A: 70%, studies over B: 50% conclusion A>B indirect corrected efficacy in RCT's A: 70%, placebo 50% B: 50% placebo: B 20% efficacy vs placebo conclusion B>A network analysis. RCT: randomized controlled trials.

example, if we evaluate multiple drugs (network analysis).^{24–26} For these evaluations more sophisticated statistical techniques, such as Metaregression are employed (Fig. 1). However, if not used with extreme methodological rigor, these tools can generate inaccurate results. We will see that a good example is the evaluation of biological drugs.

Although the complexity of biological drugs creates serious difficulties when compared, showing that 2 drugs are clinically equivalent has important assistance and economic implications. This article aims to reflect on some important aspects needed to facilitate the evaluation of biologic agents: *a*) the utility and the ethical implications of certain design studies, and in particular the use of placebo, *b*) the use of non-inferiority studies, assessment variables to consider and the importance of differences (δ) in efficacy or safety that can be considered clinically significant, and *c*) the usefulness of the methods of evidence-based medicine and, especially, indirect comparisons. This has been accomplished through a multidisciplinary approach using a non-systematic review of the literature and further discussion and consensus which involved specialists in rheumatology, dermatology, gastroenterology, clinical pharmacologists and statisticians.

Method

The preparation of the document was performed from a systematic review and a consensus reached by two of the authors (XC and JVE). The rest of the forum participants received the document by e-mail, reviewed the document and made contributions that were collected in an initial document. In a single-face-to-face meeting, the discussion points were agreed upon, the structure and content of the final document were set, and responsibility was distributed to each of the participants. Thus, the respective specialists developed the basis for proposing a value of delta in each of the indications for biological drugs. Once established, coordinators (XC and JVE) integrated the various contributions to develop a second document, which was discussed through email. Finally, all forum participants gave their approval to the final content of the document.

Consensus Conclusions

Ethical Issues: The Use of Placebo Is Currently Unacceptable in the Comparison of Biological Drugs

Most current studies contain a placebo group, although involving patients in spite of moderate/severe disease in which biological treatment is the standard treatment. Not treating patients with severe disease with effective treatment for periods that can reach 12 weeks in IBD, 16 weeks in dermatology or 54 weeks in rheumatology are, in the opinion of the forum members, unacceptable ethically. It seems difficult to justify how regulatory agencies accept, and even require placebo studies for new biologic drugs. This even when considering that to evaluate biosimilar drugs, agencies are asking for non-inferiority studies comparing the biosimilar with original molecule without the need to include a placebo group. It also seems unreasonable to accept that ethical committees accept studies versus placebo when standard treatment in patients with moderate to severe disease is a biological drug and that the use of suboptimal treatment can have negative consequences on the evolution of the disease.

Characteristics of Equivalence and Non-inferiority Studies

Classically, controlled clinical trials have used the criterion of superiority as a direct statistical comparison between different drugs. However, this approach may not be optimal for comparing biological or new biosimilar drugs. In many cases, and by definition, in the case of biosimilars, new drugs are not intended to be more effective than the standard drug with which they are being compared and strive only to demonstrate effectiveness and safety comparable to available drugs. In addition, non-inferiority studies require no placebo arm, which reduces the risk for the patient participating in the study. Publication of specific recommendations in the *CONSORT statement*^{27,28,29} shows the growing importance of non-inferiority studies.

A key feature of non-inferiority studies is the need for establishing “a priori” what difference (δ) in efficacy or safety can be considered clinically relevant or significant, and only then can a statistical analysis be performed.³⁰ Thus, non-inferiority studies are not the case of conventional comparative studies, because when these include a very large number of patients, small differences in efficacy may be statistically significant and irrelevant from a practical standpoint. An example is what happened with the CAPRIE study, which compared the effects of clopidogrel with acetylsalicylic acid (ASA) in 19,185 patients followed for 2 years. The rate of cardiovascular events was 5.83% per person per year with aspirin and 5.32% for clopidogrel. Reducing the risk of cardiovascular events was 0.51% per person per year, statistically significant ($P=0.043$) but of questionable clinical relevance.³¹ We will use this study as an example throughout this section.

In studies of non-inferiority, the difference in efficacy must be calculated for the evaluated treatment and standard treatment effectiveness, and the confidence interval of 95% (95% CI) of the difference must be determined. If the CI of the difference is within the range defined by $\pm\delta$ both treatments are considered equivalent. If the lower limit of the CI for the difference is above the $-\delta$ value, the drug meets assessed noninferiority criteria. For example, imagine that in the case of CAPRIE clopidogrel is considered standard therapy and aspirin as the assessed drug (Fig. 2). We can use an online calculator to set the 95% of the difference between the two drugs as³² $\pm 0.48\%$ -0.51% . Therefore, the 95% would range from -0.99% to -0.02% . If we assume that the maximum non-significant difference from a clinical point of view is, for example, $\pm 1\%$ or more, ASA meets criteria for both clinical equivalence and non-inferiority. If, conversely, we decrease the value δ 0.5%, aspirin does not meet

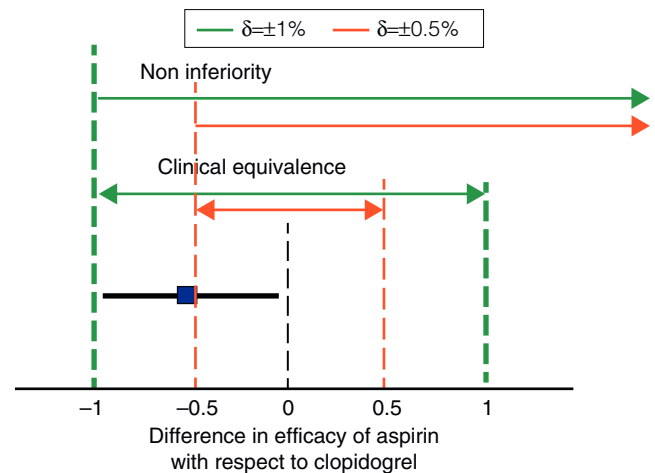


Fig. 2. Noninferiority margins and clinical equivalence. The figure shows the noninferiority intervals equivalent to two values of the difference in efficacy between aspirin and clopidogrel which can be considered clinically relevant (δ) in broken lines. The square and the horizontal bold line show the difference between aspirin and clopidogrel in the CAPRIE study and its 95% CI. So with $\delta \pm 1\%$, it could be considered that aspirin is not inferior or equivalent to clopidogrel (thick dashed lines). In contrast, with a more restrictive delta estimate δ (0.5%, thin vertical dashed lines), aspirin would not meet criteria for equivalence or non-inferiority. Noninferiority clinical equivalence difference in efficacy of aspirin with respect to clopidogrel.

clinical criteria of equivalence or non-inferiority. We see, therefore, one of the characteristics of non-inferiority studies: it is essential to properly select the numerical range of the ‘clinical relevance’, as the result of the comparison depends entirely on the range δ we choose.³³

Setting the Maximum Difference in Efficiency Is not Considered Clinically Relevant (δ)

Without a δ ³³ universally accepted formula, there are a number of defined factors to be taken into account when determining the minimum clinically relevant difference in a certain parameter of efficacy or safety. In the case of biological drugs these are:

- The type of parameter of effectiveness evaluated:* thus, if mortality is assessed, differences considered clinically relevant in the different studies are usually very small, between 0.4 and 1% in absolute percentages. In other, less “hard” parameters, e.g., remission or clinical response in the case of IBD or the rate of response in patients with moderate RA, wider margins can be reasonably considered.
- The response rate and the observed difference between the standard treatment and placebo* it has occasionally been recommended to be half the standard treatment effect vs placebo effect. It has also been recommended to determine the interval based δ on the dispersal of results of the drug to be compared, for example, by using 0.5 standard deviations of the variable to be evaluated. Finally, δ interval variable percentages have also been established depending on the degree of efficacy of the drug.²⁹ In any case, the interval must be restrictive δ enough for the placebo not to be included as inferior or non-equivalent.
- In case of multiple comparisons, δ values should be calculated from the same standard of care, which should be the most effective in absolute terms.* Otherwise, it may cause what is known as a domino effect, so that the successive comparison 2–2 of progressively less effective drugs could lead to the acceptance of drugs not superior to placebo as equivalent to the initial standard.

What Should Be the Value δ in Rheumatic Diseases?

Currently, there are 9 biological agents with indications in the data sheet for the treatment of various chronic inflammatory arthropathy. For RA, an antagonist of interleukin (IL)-1 (anakinra), 5 inhibitors of TNF- α (adalimumab, certolizumab, etanercept, golimumab and infliximab), an agent depleting CD20+B cells (rituximab), an inhibitor of soluble IL-6 receptor (tocilizumab) and an inhibitor of costimulation molecules (abatacept) have been approved.

In RA, the therapeutic efficacy is usually assessed by using combined response rates that include joint counts, acute phase reactants and global assessments of disease activity by the patient and physician. Response rates and activity are measured using American College of Rheumatology (ACR) 20, 50 and 70 criteria, which measure the percentage of improvement from baseline, regardless of the final activity of the disease, the Disease Index Activity Score (DAS) 28 and the Simplified Disease Activity Index (SDAI). The latter are absolute measures of disease activity by defining the response as the percentage of patients achieving a particular disease state (low activity or remission). Outcomes in clinical trials are quantified as the percentage of patients achieving a particular response; in the case of ACR, the percentage of patients achieving a certain state of the disease (low activity or remission) or as the mean decrease in the value of DAS28 or SDAI.

There is neither universally accepted clinical relevant value of δ nor for the percentage of patients who achieved a certain response, e.g., an ACR20 or DAS28 less than 3.2 (considered a state of low activity of RA) or clinically for the minimum clinically relevant reduction in the value of DAS28 or SDAI. However, recent clinical trials of high methodological quality point toward what would be reasonable to choose as δ a marker for comparisons between biological agents. In the ADACTA study, a randomized direct comparison between adalimumab and tocilizumab monotherapy in RA patients with inadequate response to methotrexate (MTX), the authors considered that the relevant difference between groups should be at least 0.6 units of DAS28.³⁴ Moreover, in the AMPLE study, a randomized direct comparison between adalimumab and abatacept in combination with MTX, also in RA patients with inadequate response to MTX, the authors assumed, with no scientific basis, that the margin of non inferiority in the proportion of patients who reached an ACR20 response would be at 12% among³⁵ groups. However, there is no valid delta in ACR20 ACR50 and ACR70 to compare. The ACR 50 response rate is usually 40% of patients receiving combination therapy compared to 20% of patients receiving placebo or MTX, using a 15% difference in the response rate, which can include the placebo response and thus, would not be valid δ . In the case of ACR70, the margin is still narrower, 20% vs 5%–10%, so δ values have to be even lower.

For SA and PsA, the use of 4 TNF inhibitors (adalimumab, etanercept, golimumab and infliximab) has been approved. In the case of AS, the outcome variable in clinical trials is the percentage of patients achieving a combined ASDAS index of inactive disease (≤ 1.3) or with low disease activity (≤ 2.1) and it is considered that a $\delta \geq 1.1$ is a clinically relevant change,³⁶ which may be considered a reference value. Another widely used index of activity in AS is the BASDAI. It is considered that a BASDAI ≤ 2 reflects minimum activity, while a BASDAI ≤ 4 is considered low activity.³⁷ It is proposed that a clinically relevant δ must be superior to an absolute unit or 22.5% of baseline BASDAI.³⁸

In axial forms of PsA, response rates used in AS³⁹ are assumed, while in the peripheral polyarticular forms, response rates used in the RA³⁹ response rates are assumed. Currently, there is no commonly accepted response rate to assess the response to treatment of peripheral oligoarticular forms of PsA, so it is difficult to indicate δ recommendations on what would be best suited in these cases.

What Should Be the Value δ in Psoriasis?

In dermatology, the drugs currently approved for psoriasis are three anti-TNF drugs: adalimumab, etanercept and infliximab, and an inhibitor of p40, a protein shared by IL-12 and IL-23, ustekinumab. The primary endpoint most commonly used in psoriasis clinical trials Psoriasis Area response and Severity Index (PASI)-75, and as secondary variables, the PASI-90 and PASI-100, the latter values are considered indicative of remission.^{40,41}

As to what may be the value of the increase in these parameters that can be considered clinically irrelevant, we must take into account, on the effectiveness of various drugs over placebo. Thus, different meta-analysis included the pivotal studies, differences in PASI-75 compared to the less effective drug's placebo drug, low-dose etanercept were 31%–45%.⁴² For other drugs, the differences with placebo ranged between 40% and 78%.

There is no data in the literature on the minimum clinically relevant difference in PASI-75 between 2 biologic drugs for the treatment of patients with plaque psoriasis. The only data available so far comes from the only clinical trial comparing 2 biological drugs.⁴³ The ACCEPT study compared ustekinumab with etanercept in the treatment of moderate/severe psoriasis; in this study, although no δ value was established, the calculation of the sample was performed on an expected difference in terms of PASI-75 between ustekinumab and etanercept of 14%. The study found a difference between low-dose ustekinumab and etanercept of 10.7%. They interpret this difference as clearly relevant from the clinical point of view. Finally, in another clinical trial comparing adalimumab with MTX,⁴⁴ the expected difference in terms of PASI-75 between the two treatments was 20% of patients achieving that degree of response.

Therefore, when the PASI-75 response is used as a primary endpoint, the limited data available indicate that an appropriate value should move δ between 5 and 15%, and probably should be slightly below 10% as observed in the ACCEPT study. However, strong arguments are lacking to defend a specific figure.

What Should Be the Value of δ in Inflammatory Bowel Disease?

The indexes that have been used most often in recent clinical trials are the Crohn's Disease Activity Index (CDAI) for Crohn's disease⁴⁵ and Mayo index for ulcerative colitis.⁴⁶ In CDAI's case, it is considered that the patient is in remission when the values fall below 150 and in response decreases of 70–100 points.⁴⁷ For the Mayo index, remission values of 2 or less and a decrease in response at least 3 points and 30% of the initial values are considered.⁴⁸

In the case of IBD, the disease characteristics and study design that determine differences from placebo in the pivotal studies is lower than in other diseases. In fact, in the initial studies of the drugs approved for Crohn's disease and ulcerative colitis, this ranges between 33 and 7.2%,⁴⁸ lower in patients who had received prior treatment with a biological agent.⁴⁹

In the SONIC study,⁵⁰ azathioprine, infliximab and the combination of both in patients with Crohn's disease who had not been previously received immunosuppressants were compared. The rates of clinical remission at week 26 were 30% for azathioprine, infliximab and 44.4%–56.8% for the combined treatment. These were interpreted as clearly significant differences between azathioprine and infliximab, as well as differences with combination treatment. Finally, an international consensus that conducted a systematic review of all the indexes for the evaluation of ulcerative colitis⁴⁶ indicates a value of δ for non-inferiority of 10%, although, surprisingly, it does not specify for what endpoint.

In conclusion, in the case of IBD, considering that the minor benefit over placebo that has been generally observed in the pivotal trials could indicate as approximate values of δ 10% for clinical response and slightly lower, between 5 and 10%, in case of clinical remission.

Using Systematic Reviews, Meta-Analysis and Direct and Indirect Comparisons to Establish Equivalence

Direct Comparisons

Obviously, the most effective method used to compare two biologic drugs is the direct comparison in a randomized clinical trial. However, as these comparative tests are, so far, exceptional, it is not possible to perform systematic reviews and classic metaanalysis.

Indirect Comparisons

Although they are the only available resource in the absence of direct comparisons, these studies have important methodological limitations.

First, a fundamental assumption of indirect studies is the consistency of the evidence or the comparability of the included studies. Differences in the design or in the evaluated population can cause significant biases and condition that the study results can not be compared with these techniques.⁵¹ An important and common problem is the change in population in studies over time, thus patients entering recent studies are more likely to have failed several previous biological treatments, making them more refractory to any new treatment and conditioning a worse response. Therefore, it is recommended that all indirect comparisons explicitly and extensively discuss all differences in the design of included studies that may skew the results of the analysis.⁵² There is also the possibility of trying to control these differences between studies using meta-regression techniques.⁵³ However, in practice, this possibility is limited by the small number of studies and the risk of incurring ecological bias.

Secondly, indirect comparisons markedly reduce potency comparisons and require much larger samples than direct comparisons.^{24,54} Thus, a recent review comparing direct and indirect methods⁵⁵ evaluated 39 therapeutic interventions in which direct comparisons found a statistically significant difference. In 14 of the 39 comparisons, statistical significance disappeared when an analysis combining the direct and indirect estimates was performed. Similarly, in a preliminary assessment, the same authors evaluated 19 direct comparisons that found significant differences between 2 therapeutic interventions; only 9 indirect comparisons detected significant differences.^{24,56}

The practical consequence of this limited power is that it is very difficult, through indirect studies, to show that 2 drugs have significantly different result.^{24,54–56} If this occurs (that is, if a drug is significantly superior to the other), the result is reasonably reliable. However, the fact that no differences are observed does not demonstrate the equivalence between drugs. Due to the markedly conservative nature of indirect analysis, these require very marked differences between treatments and a very large number of patients to detect a significant difference. This is very important because some studies comparing biologic agents have misinterpreted this lack of statistical power of the indirect comparisons as evidence that the drugs are equivalent.²⁰

The intense debate on the reliability of the indirect comparisons contributes, in no small measure, to the fact that studies assessing the effect of indirect comparisons on certain diseases reach divergent conclusions. As an example, in the case of psoriasis, the Signorovitch et al. study concluded that Adalimumab⁵⁷ exceeds etanercept, while a newly published study with the same comparison concludes that both are equivalent.²¹ In fact, statistical significance and even the direction of the effect may vary depending on the method used for indirect comparison. Thus, O'Regan et al. studied 51 indirect comparisons between drugs using 2 different statistical methods of analysis. Of the 51 comparisons, 3 found that with one method the difference was significant and with the

other, it was not, 6 in which the direction of the effect was different depending on the method used and 9 where the CI varied widely depending on the method.⁵⁸ For this reason, it is considered advisable in indirect comparisons to analyze not only the primary endpoint, but also secondary variables jointly (e.g., ACR20, ACR50 and ACR70 in RA studies). The analysis results will be more consistent if the differences in favor of a drug are maintained in the various measurement parameters.

Finally, we have identified significant shortcomings in the quality of indirect studies^{52,59}; Donegan et al. proposed a series of⁵² specific parameters to assess the quality of these studies. Additionally, it would be advisable, as any systematic review or meta-analysis, to perform indirect comparisons in accordance with the PRISMA recommendations.^{60,61}

Given the risks of bias, many authors are understandably wary of the reliability of the indirect comparisons. All this has led the *International Society for Pharmacoeconomic and Outcomes Research*, to attempt to improve the reliability of these studies, designating a specific working group to make recommendations for the evaluation, interpretation and implementation of indirect comparisons.^{62,63} In any case, because of the significant limitations of the method, the results obtained by indirect comparisons should be considered as exploratory data, useful for generating hypotheses subject to further confirmation but never as definitive proof of superiority, let alone of equivalence.^{52,55} In fact, regulatory agencies at no time has considered indirect comparisons as appropriate methods for evaluating biosimilar drugs.

Conclusions

This review provides evidence for the comparison of biological drugs. The main conclusions of the forum were:

- Both from a scientific point of view and from an ethical point of view, we consider advisable to modify regulatory policies in the sense of strongly favoring randomized noninferiority trials, comparing drugs face to face with the new biological treatment using current standards, avoiding trials against placebo.
- These studies provide reliable data on the comparative efficacy and safety of different drugs that they currently lack, given the unreliability and important methodological limitations of indirect comparisons.
- A key element in this process will be the consensus with the involvement of regulatory agencies, scientific societies, the pharmaceutical industry and health authorities of the clinical differences that should be considered relevant in each of the conditions tested.

Ethical Responsibilities

Protection of persons and animals. The authors declare that this research has not performed experiments on humans or animals.

Data confidentiality. The authors state that no patient data appears in this article.

Right to privacy and informed consent. The authors state that no patient data appears in this article.

Conflict of Interest

The authors declare no conflict of interest in the content of this article.

Annex.

Dermatology: Carlos Ferrandiz, Department of Dermatology, University Hospital Germans Trias i Pujol, Autonomous University of Barcelona; Hugo Vázquez Veiga, Department of Dermatology, Hospital of Conxo, University Hospital Complex of Santiago de Compostela, A Coruña. **Clinical Pharmacology:** Juan Vicente Esplugues. **Rheumatology:** Jose Luis Andreu, Department of Rheumatology, Hospital Universitario Puerta de Hierro, Majadahonda, Madrid; Antoni Gómez, Department of Rheumatology, Hospital de Sabadell, Institut Universitari Parc Taulí, Universitat Autònoma de Barcelona. **Gastroenterology:** Fernando Gomollón, Department of Gastroenterology, Clinical Hospital “Lozano Blesa” Zaragoza; Xavier Calvet. **Statistics:** David Suarez, Epidemiology and Evaluation Unit, Fundació Parc Taulí, Taulí Parc Hospital, Universitat Autònoma de Barcelona.

References

- Emery P, Dorner T. Optimising treatment in rheumatoid arthritis: a review of potential biological markers of response. *Ann Rheum Dis.* 2011;70:2063–70.
- 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.
- Loftholm 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 Gastroenterol.* 2008;103:3132–41.
- 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.
- 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>.
- 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. *Arthritis Care Res (Hoboken).* 2010;62:226–34.
- 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.
- 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.
- Bennett CL, Luminari S, Nissenon AR, Tallman MS, Klinge SA, McWilliams N, et al. Pure red-cell aplasia and epoetin therapy. *N Engl J Med.* 2004;351:1403–8.
- Fox A. Biosimilar medicines—new challenges for a new class of medicine. *J Biopharm Stat.* 2010;20:3–9.
- Health Canada. Summary basis of decision for inflectra; 2014. Available from: <http://www.hc-sc.gc.ca/dhp-mpps/prodpharma/sbd-smd/drug-med/sbd-smd.2014.inflectra-159493-eng.php> [accessed 29.04.14].
- 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.
- Scheinberg MA, Kay J. The advent of biosimilar therapies in rheumatology—“O brave new world”. *Nat Rev Rheumatol.* 2012;8:430–6.
- Danese S, Gomollon F, Governing Board and Operational Board of ECCO. ECCO position statement: the use of biosimilar medicines in the treatment of inflammatory bowel disease (IBD). *J Crohns Colitis.* 2013;7:586–9.
- Chow SC, Wang J, Endrenyi L, Lachenbruch PA. Scientific considerations for assessing biosimilar products. *Stat Med.* 2013;32:370–81.
- Chow SC, Endrenyi L, Lachenbruch PA. Comments on the FDA draft guidance on biosimilar products. *Stat Med.* 2013;32:364–9.
- Chow SC. Assessing biosimilarity and interchangeability of biosimilar products. *Stat Med.* 2013;32:361–3.
- 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.
- Bourgoin AF, Nuskey B. White paper: an outlook on U.S. biosimilar competition. *Drugs Today (Barc).* 2013;49:399–410.
- 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.
- 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.
- 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.
- 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.
- 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.
- Catala-Lopez F. Evaluation of comparative treatment effects using indirect comparisons. *Rev Esp Cardiol.* 2013;66:156–7.
- Catala-Lopez F, Tobias A. Síntesis de la evidencia clínica y metaanálisis en red con comparaciones indirectas. *Med Clin (Barc).* 2013;140:182–7.
- Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *J Am Med Assoc.* 2006;295:1152–60.
- Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *J Am Med Assoc.* 2012;308:2594–604.
- 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.
- Huson L. Statistical assessment of superiority, equivalence and non-inferiority in clinical trials. *Clinical Research Focus [Internet].* 2001. Available from: <http://www.ais.up.ac.za/med/tnm800/EssentialTNM800/DayTwo/InterventionResearch/equivalence.pdf> [accessed May 2014].
- CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). *CAPRIE Steering Committee. Lancet.* 1996;348:1329–39.
- Lowry R. The confidence interval for the difference between two independent proportions; 2014. Available from: http://www.vassarstats.net/prop2_ind.html [accessed May 2014].
- Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med.* 2006;145:62–9.
- 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.
- 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. *Arthritis Rheum.* 2013;65:28–38.
- 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.
- 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.
- 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.
- Fernández-Sueiro J, Juanola XJC, Torre-Alonso J, García-de-Vicuña R, Queiro 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.
- 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. *J Eur Acad Dermatol Venereol.* 2010;24 Suppl. 2:10–6.
- 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.
- Lucka TC, Pathirana D, Sammain A, Bachmann F, Rosumek S, Erdmann R, et al. Efficacy of systemic therapies for moderate-to-severe psoriasis: a systematic review and meta-analysis of long-term treatment. *J Eur Acad Dermatol Venereol.* 2012;26:1331–44.
- 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.
- 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). *Br J Dermatol.* 2008;158:558–66.
- Best WR, Becktel 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.
- 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.
- 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.
- 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, Ghemment 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, <http://dx.doi.org/10.1136/bmj.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, Ghemment 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, <http://dx.doi.org/10.1136/bmj.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, <http://dx.doi.org/10.1136/bmj.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, <http://dx.doi.org/10.1136/bmj.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.