Biological drugs for the treatment of rheumatoid arthritis by the subcutaneous route: interpreting efficacy data to assess statistical equivalence

Andrea Messori, Valeria Fadda, Dario Maratea, Sabrina Trippoli, Roberta Gatto, Mauro De Rosa and Claudio Marinai

Abstract

Background: No equivalence analysis has yet been conducted on the effectiveness of biologics in rheumatoid arthritis. Equivalence testing has a specific scientific interest, but can also be useful for deciding whether acquisition tenders are feasible for the pharmacological agents being compared.

Methods: Our search covered the literature up to August 2014. Our methodology was a combination of standard pairwise meta-analysis, Bayesian network meta-analysis and equivalence testing. The agents examined for their potential equivalence were etanercept, adalimumab, golimumab, certolizumab, and tocilizumab, each in combination with methotrexate (MTX). The reference treatment was MTX monotherapy. The endpoint was ACR50 achievement at 12 months. Odds ratio was the outcome measure. The equivalence margins were established by analyzing the statistical power data of the trials.

Results: Our search identified seven randomized controlled trials (2846 patients). No study was retrieved for tocilizumab, and so only four biologics were evaluable. The equivalence range was set at odds ratio from 0.56 to 1.78. There were 10 head-to-head comparisons (4 direct, 6 indirect). Bayesian network meta-analysis estimated the odds ratio (with 90% credible intervals) for each of these comparisons. Between-trial heterogeneity was marked. According to our results, all credible intervals of the 10 comparisons were wide and none of them satisfied the equivalence criterion. A superiority finding was confirmed for the treatment with MTX plus adalimumab or certolizumab in comparison with MTX monotherapy, but not for the other two biologics.

Conclusion: Our results indicate that these four biologics improved the rates of ACR50 achievement, but there was an evident between-study heterogeneity. The head-to-head indirect comparisons between individual biologics showed no significant difference, but failed to demonstrate the proof of no difference (i.e. equivalence). This body of evidence presently precludes any option of undertaking competitive tenderings for the procurement of these agents.

Keywords: adalimumab, biologics, certolizumab, equivalence, etanercept, golimumab, meta-analysis, rheumatoid arthritis, tocilizumab

Correspondence to:
Andrea Messori, PharmD
HTA Unit, Area Vasta
Centro Toscana, Regional
Health System, Via San
Salvi 12, 50100 Firenze,
Italy
**andrea.messori.it@gmail.
com**

**Valeria Fadda, PharmD
Dario Maratea, PharmD
Sabrina Trippoli, PharmD
Roberta Gatto, PharmD**
HTA Unit, ESTAV Toscana
Centro, Regional Health
Service, Firenze, Italy

Mauro De Rosa, PharmD
President, SIFACT,
Italian Society for
Clinical Pharmacy and
Therapeutics, Milano, Italy

Claudio Marinai, PharmD
Department of
Pharmaceutical Logistics,
ESTAV Toscana Centro,
Regional Health Service,
Firenze, Italy

Introduction

The literature on the effectiveness of biological agents for the treatment of rheumatoid arthritis (RA) unresponsive to methotrexate (MTX) monotherapy covers more than a decade and is therefore considered to be sufficiently settled. In particular, several meta-analyses [Singh *et al.* 2009; Gallego-Galisteo *et al.* 2012; Devine *et al.*

2011; Singh *et al.* 2011a, 2011b; Barra *et al.* 2014] have been carried out on this topic even though the follow-up of the included studies did not generally go beyond 12 months.

On the other hand, the concepts of therapeutic equivalence and noninferiority have been widely debated in the scientific literature of the last years, particularly in the context of randomized controlled trials (RCTs) and meta-analyses. As a result, the statistical methodology for equivalence and noninferiority testing is now mature [Ahn *et al.* 2013; Christensen, 2007; Tunes da Silva *et al.* 2009] and in fact these techniques are increasingly being used in evidence-based analyses (especially those focused on comparative effectiveness) [Messori *et al.* 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2014g, 2014h; Maratea *et al.* 2014].

In the present study, we examined the data of comparative effectiveness for biological agents indicated for the treatment of RA in adults by the subcutaneous route. In particular, we used the information from the RCTs published thus far to determine to what extent the different biological agents available for this indication can be considered therapeutically equivalent with one another. Hence, the original part of our study was essentially represented by the statistical assessment of equivalence which in turn requires the definition of margins of incremental benefit.

Methods

Literature search
To identify the clinical material suitable for our purposes, firstly we carried out a literature search (phase 1) focused exclusively on systematic reviews and meta-analyses; then, we extracted from these articles all RCTs that were potentially pertinent to our analysis. This first search was conducted in PubMed (last query run on 14 August 2014) and covered the last 5 years according to PubMed's definition of this interval. The PubMed filters were 'systematic reviews', 'reviews', and 'meta-analysis' combined through the 'or' Boolean operator. A single search term ('rheumatoid arthritis') was employed since the number of citations extracted though the above procedure was relatively small (less than 4000). In this phase of our search, we analyzed the eligible papers selected this way and we identified the RCTs that met our inclusion criteria (see below).

Duplicate studies were included once only according to the degree of completeness and update of the respective reports.

The subsequent phase (phase 2) of our literature analysis relied on another search that was based on the same keywords and the same inclusion criteria as the first one, but had no date limits and used the filter of 'Randomized trial' instead of the filter of 'reviews and meta-analyses'. This phase was aimed at including RCTs that were not identified from phase 1.

Inclusion criteria
Our study included the trials that met the following criteria: (a) randomized design; (b) patients with RA unresponsive to MTX alone; (c) treatment group treated with MTX in combination with etanercept (50 mg/week or 25 mg twice weekly), adalimumab (40 mg/2 weeks), golimumab (50 mg/4 weeks), certolizumab (400 mg/2 weeks three times then 200 mg/2 weeks), or tocilizumab (8 mg iv/kg/4 weeks or 162 mg sc/week); (d) control group treated with either MTX alone or MTX in combination of one of the above biologics; (e) follow-up length of at least 12 months; (f) measurement of ACR50 achievement at 12 months. Subcutaneous tocilizumab was included because the drug has received US Food and Drug Administration (FDA) approval and is thought it will be shortly approved also by the European Medicines Agency (EMA). Studies focused on juvenile disease were excluded. The dosing regimens reported for the various biologics were employed in our analysis because they reflect the respective dosages approved by EMA or FDA. Our efficacy endpoint was ACR50 mainly because this is the endpoint most commonly employed in previous meta-analyses [Devine *et al.* 2011].

Data extraction
For each trial, we extracted the basic information needed for our analysis as well as information on the primary endpoint. Extraction was made in duplicate by AM and VF, and discrepancies were resolved by consensus.

The information about the endpoint achievement was aimed to reflect the intention-to-treat population. However, there were some occasional post-randomization exclusions in some trials, and so the clinical material actually adopted the

so-called modified intention-to-treat population [Montedori *et al.* 2011].

As regards the trials' primary endpoint, we also retrieved from the articles the details concerning how statistical power calculations were conducted.

Bayesian network meta-analysis

The Bayesian framework to make indirect comparisons [Greco *et al.* 2013; Mills *et al.* 2013] is increasingly being used and can be considered the current standard in the field [Jansen *et al.* 2011; Hoaglin *et al.* 2011]. As compared with the traditional frequentist approach, the Bayesian approach entails one main advantage in that all treatments included in the comparison are incorporated into a single model. Another advantage of the Bayesian approach is that this technique enables rank ordering of each treatment. As opposed to traditional confidence intervals adopted in frequentist analysis [Bucher *et al.* 1997], the Bayesian output reports credible intervals (CrIs), which can be directly interpreted as the probability of an event residing in the reported range.

The Bayesian analysis involves a formal combination of a prior probability distribution that reflects a prior belief of the possible values of the effect of interest, and the likelihood distribution of the effect based on the observed data, to obtain a posterior distribution. In the absence of real data, prior probabilities are assigned by using vague, flat, or noninformative priors (that are generally small numbers between 0 and 3).

The model adopted for our analysis was a random-effects logistic regression model run within a Bayesian framework. This model employs a random sequence of chains, called Markov chain Monte Carlo simulation. Each chain must be run for a length of time sufficient to allow model convergence (burn-in) before estimating posterior probabilities [Hoaglin *et al.* 2011]. We created the random-effects logistic regression model by using the binary outcome of whether each subject in each arm of each study met the ACR50 criteria at the specified time interval. Randomization within each study was preserved by specifying each arm in each study separately, thus accounting for the effect of the comparator. Results were presented as log odds ratio (OR). We accounted for heterogeneity among studies by applying meta-regression techniques and by consequently generating an index of heterogeneity.

In running our analysis, we first determined whether the ACR50 improvement for each biologic in combination with MTX was significantly different from that of the controls based on the pooled trial data. Then, the rank order was calculated, in terms of ACR50 improvement, for each combination treatment in comparison with MTX alone. Next, we estimated pairwise comparisons for each treatment with one another by calculating the difference in log OR and, finally, by determining the OR for each comparison (both direct and indirect). All values of OR were associated with their respective 5–95% CrI, that reflect a 90% CrI. Direct comparisons are those for which at least a single clinical trial was available while indirect comparisons are those for which a 'real' trial is lacking. Finally, as a sensitivity analyses, we changed the initial values from which each Markov chain Monte Carlo simulation began, as is customary in the Bayesian framework [Jansen *et al.* 2001].

Recent advances in computing power and the development of sophisticated software have greatly facilitated the use of Bayesian statistics. All of our analyses were conducted by using the software package WinBUGS 1.4.3 (Cambridge, United Kingdom) in combination with the meta-analysis code developed by the National Institute for Health and Care Excellence [NICE, 2010].

Equivalence testings

RCTs always incorporate power calculations and power calculations, in turn, require the declaration of a prespecified expected benefit [Norman *et al.* 2012; Jones *et al.* 2014; Sedgwick, 2014; Sobrero and Bruzzi, 2009]. For this reason, the magnitude of the benefit adopted for power calculations influences clinical research as well as the clinical decision process, because this magnitude tends to be inversely proportional to sample size. Only a few studies have been carried out in this area. The article by Norman and coworkers has explored the technical side of the problem by highlighting that, on the one hand, power calculations invariably imply a certain degree of arbitrariness. On the other hand, declaring a prespecified (incremental) benefit for power calculations (the so-called 'margin' or 'delta') reflects the concept of the minimal important difference and therefore aims at differentiating between a clinically relevant incremental benefit and an irrelevant one [Norman *et al.* 2012; Jones *et al.* 2014; Sedgwick, 2014; Sobrero and Bruzzi, 2009].
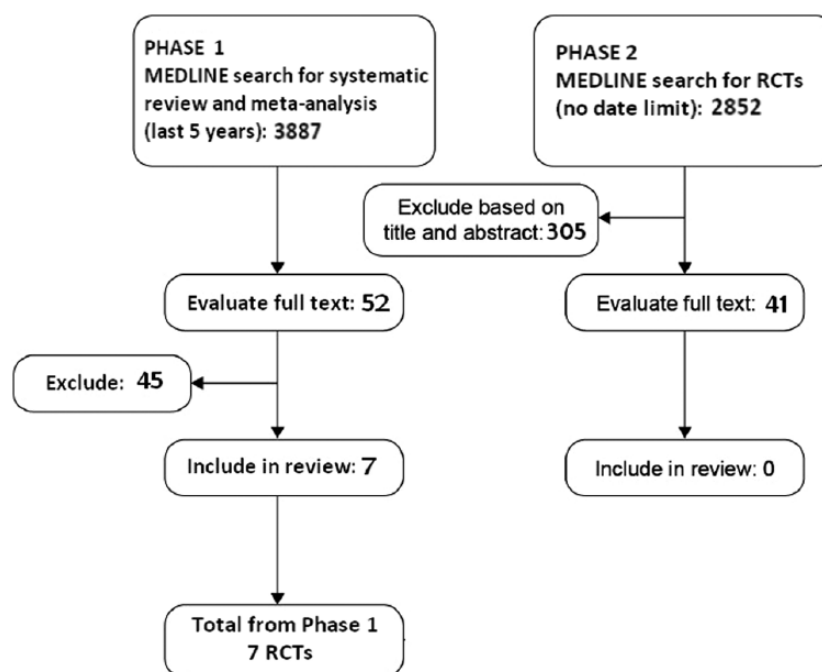
Figure 1. PRISMA schematic.
Summary of the phases of our literature search and their respective results.

In our study, first we tabulated the information that we extracted from each trial on both the primary endpoint and the design of sample size calculations (i.e. the margins adopted in each individual trial). Then, on the basis of this overall body of data, we empirically determined, by consensus among the authors (AM, VF, DM, ST), the maximum variation in ACR50 (or margin) representing the minimal clinically important difference.

Finally, this variation (or margin) was incorporated in the series of statistical tests in which we determined to what extent the available evidence demonstrated the presence of therapeutic equivalence across the different biologics. Since the outcome measure adopted in our network meta-analysis was the OR, the equivalence margin for ACR50 was expressed according this parameter. The results of the equivalence tests were interpreted according to standard criteria [Ahn *et al.* 2013; Christensen, 2007; Tunes da Silva *et al.* 2009]. Briefly, the demonstration of equivalence is accepted when, in a standard Forest plot, the 95% CrI of the outcome measure includes the identity line and does not cross (and does not touch) the margin(s). In these graphs, it is common practice to vertically draw both the identity line (at $x = 1$ for relative risk or OR or

hazard ratio, at $x = 0$ for risk difference) and the margin(s), while the 95% CrIs are plotted horizontally.

In summary, the available evidence from RCTs was interpreted by comparing the pooled incremental benefits estimated for each pairwise head comparison according to our network meta-analysis (both direct and indirect comparisons) with a threshold benefit representing the margin of therapeutic equivalence.

Results

Literature search, identification of included studies, and data extraction

Our literature search, which is summarized in Figure 1, extracted a total of 3887 citations in phase 1 and 2852 citations in phase 2. For a further scrutiny of the material eligible for our analysis, we examined the full text of 52 articles in phase 1 and 41 in phase 2. After examining these papers, we selected a total of 7 RCTs (total number of patients: 2846) that met our inclusion criteria (Table 1). The biological agents used in these trials were adalimumab, certolizumab, etanercept, and golimumab. No study utilizing tocilizumab satisfied our inclusion criteria.

Table 1. Characteristics of the seven RCTs identified by our literature search and included in our Bayesian network meta-analysis.

Study	Treatment	Patients achieving ACR20 at 12 months	Primary endpoint	Margin for sample size estimation
PREMIER-2006 [Breedveld *et al.* 2006]	MTX adalimumab+MTX	118/257 166/268	ACR50 at 12 months	Absolute improvement of at least 13%
KEYSTONE-2004 [Keystone *et al.* 2004]	MTX adalimumab+MTX	20/200 87/207	ACR20 at 24 weeks	Absolute improvement of at least 20%
BEJARANO-2008 [Bejarano *et al.* 2008]	MTX adalimumab+MTX	33/73 42/75	Job loss at 2 months	Absolute improvement from 22% to 5%
KEYSTONE-2008 [Keystone *et al.* 2008]	MTX certolizumab*+MTX	15/199 149/393	ACR20 at 24 weeks	Absolute improvement of at least 50%
KLARESKOG-2004 [Klareskog *et al.* 2004]	MTX etanercept+MTX	98/228 159/231	ACRN-AUC at 24 weeks	Absolute improvement of at least 4.5 units (with SD=14)
COMET-2008 [Emery *et al.* 2008]	MTX etanercept+MTX	129/263 188/265	DAS28 at 12 months	Absolute improvement from 23% to 37%
KEYSTONE-2013 [Keystone *et al.* 2011, 2009, 2013]	MTX golimumab+MTX	38/104 50/83	ACR20 at 14 weeks	Absolute improvement from 35% to 55%

*In this study, the patients did not receive the three 400 mg doses scheduled at the beginning of treatment according to summary of product characteristics.
Abbreviations: RCT, randomized controlled trial; MTX, methotrexate; CrI, credible interval; ACR, American College of Rheumatologists; AUC, area under the curve; DAS, Disease Activity Score; ACRN, American College of Rheumatologists index of improvement in rheumatoid arthritis; SD, standard deviation.

As regards the margins for power calculations, only the PREMIER trial adopted the endpoint of ACR50 at 12 months, while the other five (as shown in Table 1) employed other endpoints which in most cases were coprimary. Another drawback in the perspective of our equivalence analysis was that the PREMIER trial indicated an absolute improvement for the 'margin' (at least 13%), but did not specify the two values (for the treatment group and the controls, respectively) that generated this improvement.

In this context, the choice of a specific margin for our equivalence analysis was necessarily a discretional one. In previous research testing noninferiority/equivalence [Granger *et al.* 2011], it has been proposed that the pivotal randomized studies comparing the new treatment (in this case, a biologic combined with MTX) *versus* the old standard (in this case, MTX monotherapy) can suggest, on the basis of their results, the margin for future research in the same area because this 'appropriate' margin can be set midway (on a logarithmic scale) between the experimental OR found in the pivotal RCTs and 1. In other words, this method of margin determination is designed to preserve at least 50% of the relative risk reduction previously observed in RCTs evaluating the new standard *versus* the old one. This method has found a wide application in the research on stroke prevention in atrial fibrillation [Messori *et al.* 2014a, 2014h; Granger *et al.* 2011].

This method of preserving 50% of previous effectiveness was applied to equivalence margins of our analysis for the endpoint of ACR50 at 12 months. For this purpose, a traditional meta-analysis (random effect model) of the seven RCTs was conducted (Figure 2). This meta-analysis gave a pooled OR of 3.08 for the direct comparison of any biologic+MTX *versus* MTX. Finally, on the basis of this value of pooled OR our equivalence margins were calculated at 0.56–1.78. Interestingly enough, this traditional meta-analysis of direct comparisons showed a high degree of heterogeneity (I^2, 80% for the overall analysis and 88% for the three RCTs studying adalimumab).

Bayesian network meta-analysis
We ran 10,000 iterations within a single Markovian chain and then another 10,000 iterations for the so-called 'burn-in' of the parameters. This overall series of iterations demonstrated satisfactory convergence of the assigned prior distributions. The results of this analysis are shown in Table 2. Superiority was
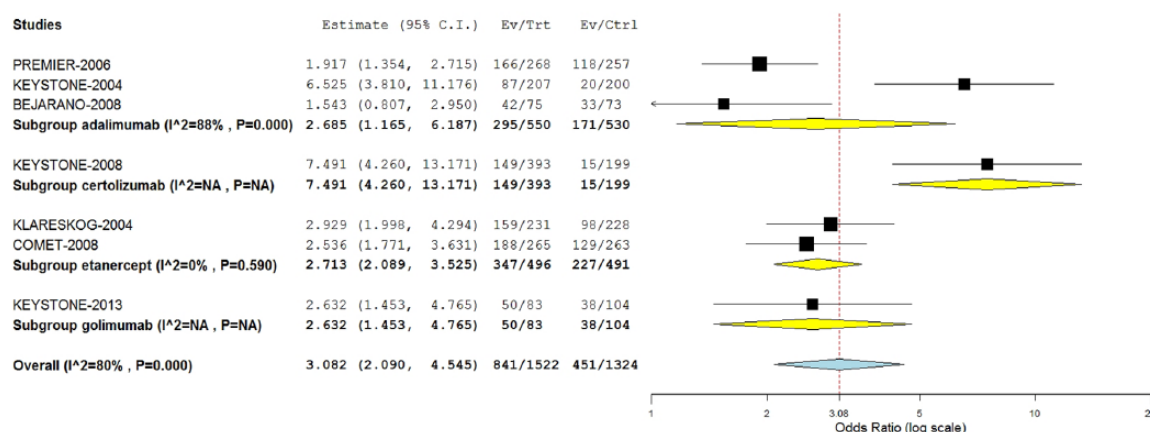
Figure 2. Traditional pairwise meta-analysis of RCTs comparing a biologic+MTX *versus* MTX monotherapy: Forest plot and values of OR.
Here we show the pairwise meta-analysis that we carried out to estimate the magnitude of the incremental benefit between any biologic+MTX *versus* MTX alone. The clinical material was represented by the same seven RCTs included in our Bayesian meta-analysis. The graph shows the Forest plot, the study-specific crude rates of incidence, and the meta-analytic results expressed as pooled OR (with 95% confidence interval) for the seven trials. Abbreviations: CI, confidence interval; Ev, number of patients with the event; Trt, total number of patients in the treatment group; Ctrl, total number of patients in the control group; I^2, index of heterogeneity; OR, odds ratio; RCT, randomized controlled trial; MTX, methotrexate. This analysis was carried out using the OMA software (Open Meta-Analyst, version 4.16.12, Tufts University, http://tuftscaes.org/open_meta/).

Table 2. Results of Bayesian network meta-analysis.

Comparison	OR (5–95% CrI) for ACR50 achievement at 12 months
Adalimumab + MTX *versus* MTX monotherapy*	2.72 (1.09 to 6.78)
Certolizumab + MTX *versus* MTX monotherapy*	7.81 (1.54–39.63)
Etanercept + MTX *versus* MTX monotherapy*	2.76 (0.92–8.28)
Golimumab + MTX *versus* MTX monotherapy*	2.65 (0.53–13.40)
Certolizumab *versus* adalimumab	0.35 (0.05–2.24)
Etanercept *versus* adalimumab	0.98 (0.24–4.10)
Golimumab *versus* adalimumab	1.02 (0.16–6.57)
Etanercept *versus* certolizumab	2.83 (0.40–20.08)
Golimumab *versus* certolizumab	2.94 (0.30–29.15)
Golimumab *versus* etanercept	1.04 (0.15–7.36)

*Direct comparisons.
Abbreviations: OR, odds ratio; MTX, methotrexate; CrI, credible interval; ACR (American College of Rheumatologists).

demonstrated only for certolizumab+MTX *versus* MTX and adalimumab+MTX *versus* MTX, and this confirms that the Bayesian approach differs to a certain extent from the traditional approach reported in Figure 2. The former, in fact, is more conservative (i.e. wider intervals) than the latter. In this case, this difference probably reflected the high degree of heterogeneity of the clinical material that affected not only treatment groups but also control groups (see Figure 2).

Running two further Markovian chains as a form of sensitivity analysis (data not shown) gave full confirmation of these results. Given that most of

pairwise comparisons included just a single RCT, we could not formally assess statistical heterogeneity and publication bias.

In the histogram of rankings (figure not shown), certolizumab ranked first in 80% of simulations; MTX monotherapy ranked last in 80% of simulations, while the remaining three treatments had intermediate rankings that were very close with one another.

Equivalence testing
Using the results shown in Table 2 and the margins determined by the empirical approach, our
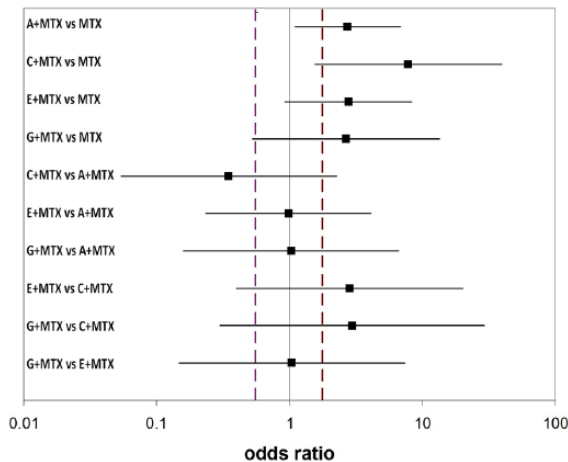
Figure 3. Forest plot and equivalence testing for meta-analytical values of OR in four direct comparisons of biologic+MTX *versus* MTX monotherapy and six head-to-head indirect comparisons between individual biologics.
The equivalence test is based on the area comprised between the two vertical dashed lines, that reflect the predetermined equivalence margins (set at OR of 0.56 and 1.78). The values of OR were those estimated by our Bayesian network meta-analysis. Each horizontal bar indicates the two-sided 90% CrI for the OR (solid square). The criterion for demonstrating equivalence is when both extremes of the CrI remain within the two vertical lines. Abbreviations: A, adalimumab; C, certolizumab; E, etanercept; G, golimumab; OR, odds ratio; CrI, credible interval; MTX, methotrexate.

equivalence tests were carried out on the basis of 90% CrI (that, in terms of percentiles, identify a CrI from 5% to 95%). These results are summarized in Figure 3. Among the 10 datasets shown in this figure, no comparison demonstrated equivalence according to the prespecified margin. Rather, all comparisons remained far from demonstrating equivalence.

Discussion

The main strength of our analysis lies in the attempt to study the degree of equivalence of the different biologics currently used in RA. Demonstrating equivalence (i.e. proof of no difference) requires much more stringent criteria than those needed to conclude no proof of difference [Messori *et al.* 2014a]. The latter is an inconclusive result in that the therapeutic question remains uncertain; in contrast, the former is much more informative (i.e. a conclusive result) because a proof is provided that no (clinically relevant) difference exists between the various comparators.

The studies published thus far on comparative effectiveness of biologics in RA have not incorporated any

formal assessment of equivalence or, at best, have presented a series of remarks that stressed 'no proof of difference' across the different biologics (i.e. a nonsignificant difference). In other words, no consideration was given in the previous literature as to whether the 'proof of no difference' was achieved or not. In other areas of pharmacotherapy, we have conducted several studies aimed at assessing equivalence [Messori *et al.* 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2014g, 2014h; Maratea *et al.* 2014], and many of them [Messori *et al.* 2014b, 2014f, 2014g, 2014h; Maratea *et al.* 2014] have been successful in demonstrating equivalence despite the use, as always, of stringent criteria. In particular, a recent network meta-analysis[10] comparing five biologics in moderate-to-severe psoriasis (adalimumab, low-dose ustekinumab, standard-dose ustekinumab, standard-dose etanercept, and high-dose etanercept) examined a total of 10 head-to-head comparisons (based on 16 RCTs involving more than 8000 patients) and found that two of these comparisons demonstrated equivalence and another four exceeded the prespecified margins to a small extent.

The main limitation of our study is related to the uncertainty that still surrounds the decision on which margins are appropriate for RCT power calculations and whether these margins are also satisfactory for the subsequent inclusion in equivalence analysis. However, the literature in this topic is growing rapidly and most controversies in this area are likely to be solved in the near future.

Apart from their scientific interest, these demonstrations have also practical implications because, in some countries including Italy, procurement of medicines through acquisition tenderings is increasingly associated to evidential reports to support the clinical feasibility of the procurement approach. In Italy, a national regulation [Italian Agency of Medicines, 2012] has been issued to improve the homogeneity of drugs' access across different regions; for this purpose, all tenderings run by the National Health System are mandatorily subjected to a preventive technical authorization issued by our national Agency for Medicines (Agenzia Italiana del Farmaco, AIFA) that declares which agents are 'equivalent' and can therefore be managed through these tenderings. It can therefore be seen that the world of drug regulation is increasingly involved in the assessment of equivalence. Finally, it should be recalled that, in most countries, acquisition tenderings do not adjudicate to the winner the whole amount of the expected drug consumption, but only a

majority; hence, residual amounts (up to 30% overall) can be adjudicated to the other products in order to meet specific clinical needs of individual patients. This of course facilitates the practical implementation of tenderings.

The present analysis failed to demonstrate equivalence among any of the four biologics examined. Understanding the reasons for this failed demonstration can be important. First, the high degree of clinical heterogeneity observed in the included trials is likely to be the best explanation for our results. Second, the population of patients enrolled in these trials was relatively small and, while our choice to include RCTs with at least 12 months of follow up has increased the external validity of the effectiveness data, the studies with 12 months of follow up were much less numerous than those based on a follow up of 6 months. Studying also the data sets with 6-month follow up will therefore be worthwhile.

In comparison with previous articles, our study was characterized by some specific features, namely: (a) a more updated literature search; (b) a specific focus on subcutaneous agents, the class of which is particularly suitable for being subjected to open tenderings by public health systems; (c) the presentation of equivalence testings according to standard Forest plots. Biologics for intravenous use were excluded from our study; these agents pose other questions to clinicians and decision makers (mainly in relation to the need of in-hospital administration) and, in our view, our choice of leaving them out was opportune.

In conclusion, on the one hand our study has a specific scientific interest in that another application of equivalence testing is described in the area of pharmacotherapy and, more specifically, on biological agents for RA. On the other hand, our experience can be useful in practical terms because our findings do not presently support the choice of running competitive tenderings for the procurement of this class of pharmacological agents.

Funding

Conflict of interest statement

The authors declare no conflict of interest in preparing this article.

References

Ahn, S., Park, S. and Lee, K. (2013) How to demonstrate similarity by using noninferiority and equivalence statistical testing in radiology research. *Radiology* 267: 328–338.

Barra, L., Ha, A., Sun, L., Fonseca, C. and Pope, J. (2014) Efficacy of biologic agents in improving the Health Assessment Questionnaire (HAQ) score in established and early rheumatoid arthritis: a meta-analysis with indirect comparisons. *Clin Exp Rheumatol* 32: 333–341.

Bejarano, V., Quinn, M., Conaghan, P., Reece, R., Keenan, A., Walker, D. *et al.* for the Yorkshire Early Arthritis Register Consortium (2008) Effect of the early use of the anti-tumor necrosis factor adalimumab on the prevention of job loss in patients with early rheumatoid arthritis. *Arthritis Rheum* 59: 1467–1474.

Breedveld, F., Weisman, M., Kavanaugh, A., Cohen, S., Pavelka, K., van Vollenhoven, R. *et al.* (2006) The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis Rheum* 54: 26–37.

Bucher, H., Guyatt, G., Griffith, L. and Walter, S. (1997) The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 50: 683–691.

Christensen, E. (2007) Methodology of superiority vs. equivalence trials and non-inferiority trials. *J Hepatol* 46: 947–954.

Devine, E., Alfonso-Cristancho, R. and Sullivan, S. (2011) Effectiveness of biologic therapies for rheumatoid arthritis: an indirect comparisons approach. *Pharmacotherapy* 31: 39–51. DOI: 10.1592/phco.31.1.39. PubMed PMID: 21182357.

Emery, P., Breedveld, F., Hall, S., Durez, P., Chang, D., Robertson, D. *et al.* (2008) Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial. *Lancet* 372: 375–382. DOI: 10.1016/S0140-6736(08)61000-4.

Gallego-Galisteo, M., Villa-Rubio, A., Alegre-del Rey, E., Márquez-Fernández, E. and Ramos-Báez, J. (2012) Indirect comparison of biological treatments in refractory rheumatoid arthritis. *J Clin Pharm Ther* 37: 301–307.

Granger, C., Alexander, J., McMurray, J., Lopes, R., Hylek, E., Hanna, M. *et al.* for the ARISTOTLE Committees and Investigators (2011) Apixaban versus

warfarin in patients with atrial fibrillation. *N Engl J Med* 365: 981–992.

Greco, T., Landoni, G., Biondi-Zoccai, G., D'Ascenzo, F. and Zangrillo, A. (2013) A Bayesian network meta-analysis for binary outcome: how to do it. *Stat Methods Med Res* 28 October 2013. [Epub ahead of print 28 October]

Hoaglin, D., Hawkins, N., Jansen, J., Scott, D., Itzler, R., Cappelleri, C. *et al.* (2011) 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* 14: 429–437. DOI: 10.1016/j.jval.2011.01.011.

Italian Agency of Medicines (2012) "Decreto Balduzzi", Art. 15, comma 11-bis, Legge 7 agosto 2012 n. 135. Available at: http://www.agenziafarmaco.gov.it/it/content/aggiornamento-al-15052013-delle-tabelle-con-lista-farmaci-di-classe-consentire-tutti-gli-ope (last accessed 16 January 2014).

Jansen, J., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N. *et al.* (2011) 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* 14: 417–428. DOI: 10.1016/j.jval.2011.04.002.

Jones, P., Beeh, K., Chapman, K., Decramer, M., Mahler, D. and Wedzicha, J. (2014) Minimal clinically important differences in pharmacological trials. *Am J Respir Crit Care Med* 189: 250–255.

Keystone, E., Genovese, M., Klareskog, L., Hsia, E., Hall, S., Miranda, P. *et al.* (2010) Golimumab in patients with active rheumatoid arthritis despite methotrexate therapy: 52-week results of the GO-FORWARD study. *Ann Rheum Dis* 69: 1129–1135. DOI: 10.1136/ard.2009.116319. Erratum in: (2011) *Ann Rheum Dis* 70: 238-239.

Keystone, E., Genovese, M., Hall, S., Miranda, P., Bae, S., Palmer, W. *et al.* (2013) Golimumab in patients with active rheumatoid arthritis despite methotrexate therapy: results through 2 years of the GO-FORWARD study extension. *J Rheumatol* 40: 1097–1103.

Keystone, E., Genovese, M., Klareskog, L., Hsia, E., Hall, S., Miranda, P. *et al.* (2009) GO-FORWARD Study. Golimumab, a human antibody to tumour necrosis factor alpha given by monthly subcutaneous injections, in active rheumatoid arthritis despite methotrexate therapy: the GO-FORWARD Study. *Ann Rheum Dis* 68: 789–796.

Keystone, E., Heijde, D., Mason, D. Jr, Landewé, R., Vollenhoven, R., Combe, B. *et al.* (2008) Certolizumab pegol plus methotrexate is significantly more effective than placebo plus methotrexate in active rheumatoid arthritis: findings of a fifty-two-week, phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. *Arthritis Rheum* 58: 3319–3329. DOI: 10.1002/art.23964.

Keystone, E., Kavanaugh, A., Sharp, J., Tannenbaum, H., Hua, Y., Teoh, L. *et al.* (2004) Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum* 50: 1400–1411.

Klareskog, L., van der Heijde, D., de Jager, J., Gough, A., Kalden, J., Malaise, M. *et al.* for the TEMPO (Trial of Etanercept and Methotrexate with Radiographic Patient Outcomes) study investigators (2004) Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *Lancet* 363: 675–681.

Maratea, D., Fadda, V., Trippoli, S., Gatto, R. and De Rosa, M. (2014) Letter: Biological drugs for inducing remission in patients with ulcerative colitis: determining statistical equivalence according to evidence based methods. *Aliment Pharmacol Ther* 5: 341–344.

Messori, A., Fadda, V., Gatto, R., Maratea, D. and Trippoli, S. (2014a) Differentiating between "no proof of difference" and "proof of no difference" for new oral anticoagulants. *BMJ* 348: g1955. DOI: 10.1136/bmj.g1955.

Messori, A., Fadda, V., Maratea, D., Gatto, R., Trippoli, S., Rosa, M. *et al.* (2014b) Intravenous proton pump inhibitors for stress ulcer prophylaxis in critically ill patients: determining statistical equivalence according to evidence-based methods. *Int J Clin Pharmacol Ther* 52: 825–829.

Messori, A., Fadda, V., Maratea, D. and Trippoli, S. (2014c) Intramyocardial bone marrow-derived cells for ischemic heart disease: is the benefit clinically relevant? *Atherosclerosis* 234: 152–153. DOI: 10.1016/j.atherosclerosis.2014.02.027.

Messori, A., Fadda, V., Maratea, D. and Trippoli, S. (2014d) Outcomes with short-term versus long-term antiplatelet dual therapy after drug-eluting stenting: quantifying the equivalence margins. *Int J Cardiol* 172: 469–470. DOI: 10.1016/j.ijcard.2013.12.228.

Messori, A., Fadda, V., Maratea, D., Trippoli, S., Gatto, R., De Rosa, M. *et al.* (2014e) Biological drugs for the treatment of moderate-to-severe psoriasis by subcutaneous route: determining statistical equivalence according to evidence-based methods.

Clin Drug Investig 34: 593–598. DOI: 10.1007/ s40261-014-0214-1.

Messori, A., Fadda, V., Maratea, D., Trippoli, S. and Marinai, C. (2014f) Anti-reabsorptive agents in women with osteoporosis: determining statistical equivalence according to evidence-based methods. *J Endocrinol Invest* Jul 10 [Epub ahead of print].

Messori, A., Fadda, V., Maratea, D., Trippoli, S. and Marinai, C. (2014g) Testing the therapeutic equivalence of alogliptin, linagliptin, saxagliptin, sitagliptin or vildagliptin as monotherapy or in combination with metformin in patients with type 2 diabetes. *Diabetes Ther* 5: 341–344. DOI: 10.1007/ s13300-014-0066-y.

Messori, A., Fadda, V., Maratea, D., Trippoli, S. and Marinai, C. (2014h) Testing the therapeutic equivalence of novel oral anticoagulants for thromboprophylaxis in orthopedic surgery and for prevention of stroke in atrial fibrillation. *Int J Clin Pharmacol Ther* (in press).

Mills, E., Thorlund, K. and Ioannidis, J. (2013) Demystifying trial networks and network meta-analysis. *BMJ* 346: f2914.

Montedori, A., Bonacini, M., Casazza, G., Luchetta, M., Duca, P., Cozzolino, F. *et al.* (2011) Modified versus standard intention-to-treat reporting: are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. *Trials* 12: 58.

NICE (2010) Clinical Guidelines, No. 92. National Clinical Guideline Centre – Acute and Chronic Conditions (UK). London: Royal College of Physicians (UK). Available at .http://www.ncbi.nlm. nih.gov/books/NBK116530/ (accessed 14 August 2014).

Norman, G., Monteiro, S. and Salama, S. (2012) Sample size calculations: should the emperor's clothes be off the peg or made to measure? *BMJ* 345: e5278.

Sedgwick, P. (2014) Understanding why "absence of evidence is not evidence of absence". *BMJ* 349: g4751.

Singh, J., Beg, S. and Lopez-Olivo, M. (2011a) Tocilizumab for rheumatoid arthritis: a Cochrane systematic review. *J Rheumatol* 38: 10–20. DOI: 10.3899/jrheum.100717.

Singh, J., Christensen, R., Wells, G., Suarez-Almazor, M., Buchbinder, R., Lopez-Olivo, M. *et al.* (2009) A network meta-analysis of randomized controlled trials of biologics for rheumatoid arthritis: a Cochrane overview. *CMAJ* 181: 787–796. DOI: 10.1503/cmaj.091391. Erratum in: (2010) *CMAJ* 182: 806.

Singh, J., Wells, G., Christensen, R., Tanjong Ghogomu, E., Maxwell, L., Macdonald, J. *et al.* (2011b) Adverse effects of biologics: a network meta-analysis and Cochrane overview. *Cochrane Database Syst Rev* 2: CD008794. DOI: 10.1002/14651858. CD008794.pub2.

Sobrero, A. and Bruzzi, P. (2009) Incremental advance or seismic shift? The need to raise the bar of efficacy for drug approval. *J Clin Oncol* 27: 5868–5873.

Tunes da Silva, G., Logan, B. and Klein, J. (2009) Methods for equivalence and non-inferiority testing. *Biol Blood Marrow Transplant* 15(1 Suppl.): 120–127.