

# Equivalence and *switching* between biosimilars and reference molecules in rheumatoid arthritis: protocol for two systematic reviews and meta-analyses

**Bruna de Oliveira Ascef** (✉ [brunaascef16@gmail.com](mailto:brunaascef16@gmail.com))

Universidade de Sao Paulo <https://orcid.org/0000-0002-2744-3974>

**Matheus Oliveira Almeida**

Universidade Ibirapuera

**Ana Cristina de Medeiros Ribeiro**

Universidade de Sao Paulo

**Danieli Castro Oliveira de Andrade**

Universidade de Sao Paulo

**Haliton Alves de Oliveira Júnior**

Universidade Federal de Minas Gerais

**Tiago da Veiga Pereira**

University of Leicester

**Patrícia Coelho de Soárez**

Universidade de Sao Paulo

---

## Protocol

**Keywords:** Biosimilar Pharmaceuticals, Etanercept, Infliximab, Adalimumab, Arthritis, Rheumatoid

**Posted Date:** September 21st, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-79492/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Systematic Reviews on July 17th, 2021.

See the published version at <https://doi.org/10.1186/s13643-021-01754-x>.

# Abstract

**Background** Biologic drugs such as adalimumab, etanercept, and infliximab represent major first-line and second-line treatments for rheumatoid arthritis (RA) patients. However, their high cost poses a massive burden on healthcare systems worldwide. The expiration of patents for these biologics has driven to the production of biosimilar drugs, which are potentially less costly and remarkably similar, albeit not identical to the reference molecules. These two systematic reviews aim to investigate the efficacy and safety profile of biosimilars compared to biologics (systematic review 1) and the impact of switching between biosimilar drugs and reference biologics on the management of RA patients (systematic review 2).

**Methods** Electronic searches will be performed through MEDLINE (via Pubmed), EMBASE, LILACS, and CENTRAL (from inception to September 2020). Risk of bias assessments will be carried out with the Cochrane risk of bias tool, supplemented with specific domains from equivalence trials. Random-effects models will be fitted to obtain summary estimates using either relative risk or standardized mean difference as a metric. Between-trial heterogeneity will be tested and quantified with the Q-test and  $I^2$  metric, respectively, whereas assessment of small-study bias will be examined through contour-enhanced funnel plots and Egger and Harbord's tests. Meta-regression models will be fitted when appropriate. The primary outcome will be the rate of treatment success according to the American College of Rheumatology 20 (ACR20) and the co-primary outcome will be the Patient's self-assessment of physical function (Health Assessment Questionnaire - Disability Index). Conclusions will be based on equivalence hypothesis testing using predefined margins of equivalence elicited from a group of experienced rheumatologists and prior studies. The overall certainty of evidence will be assessed based on the GRADE system.

**Discussion** The two systematic reviews described here, to the best of our knowledge, are the first ones proposing a comprehensive, clinician-oriented approach to assess the equivalence and the impact of switching between biosimilars and biologics on the management of patients with RA. The results will elucidate the efficacy, safety, immunogenicity of biosimilars, and the clinical consequences of substituting biologics with biosimilars in the management of RA.

**Systematic Review Registration** Protocol synopses were previously registered on PROSPERO (CRD42019137152 and CRD42019137155).

## Background

Rheumatoid Arthritis (RA) is a chronic inflammatory joint disease that affects up to 20 million people worldwide; thereby representing a major public health burden with important socioeconomic consequences (1–3).

Biological drugs, commonly known as “biologics”, are invaluable resources in the treatment of patients RA. Synthetic disease-modifying antirheumatic drugs (DMARDs) are the first line of therapy and

associated with biologics DMARDs have changed clinical outcome reducing the inflammatory burden of disease and therefore chronic articular deterioration (4). The effectiveness and safety of biologic DMARDs have been robustly established (5–7), and several studies have identified factors that affect the patient's response to these DMARDs (8–11). However, biologics pose an important challenge for the sustainability of healthcare systems worldwide, given the high direct costs associated with this drug category (1). For instance, expenses related to biologic treatments can represent almost 40% of the net drug spending in the United States (12).

Given the rapid evolution of pharmaceutical technologies over the past decade and patent expiration of previously approved biologic molecules, biosimilar drugs have been developed as less costly alternatives to their reference biologics (13). According to the U.S. Food and Drug Administration (FDA), biosimilars possess clinically similar benefits and safety profile compared to the existing FDA-approved biologics (14). In this regard, it is believed that biosimilars can accelerate the rheumatic disorders drug market competition, positively impacting the global health-care system through improved health-care affordability and increased patients' access to effective and safe drugs (13, 15)

However, despite the cost-savings potential of biosimilar drugs, there is still uncertainty as to whether currently marketed biosimilar drugs are equivalent to reference molecules in terms of efficacy, safety, and immunogenicity. Besides, the switching and interchangeability between biologics and biosimilars drugs are still topics of great debate in the treatment of RA (16–20).

Herein, we describe the protocols of two systematic reviews that will address the efficacy, safety and immunogenicity of biosimilars compared to biologics, and the impact of switching between biosimilar drugs and reference biologics on the management of RA patients. Unlike previous reviews, we will establish acceptable equivalence margins elicited from clinical specialists to conclude on the equivalence of biosimilars compared to biologics.

## Methods/ Design

### Reporting guidelines used in this protocol

The present protocol has been developed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols guidelines (PRISMA-P) 2015 statement (**Additional file 1**) (21). We will refer separately to the systematic review of efficacy and safety (systematic review 1) and switching (systematic review 2) because the two systematic reviews address two separate objectives and, therefore, need varying methodologies and approaches.

### PROSPERO synopses

Synopses for the two systematic reviews were prospectively registered in the International Prospective Register of Systematic Reviews (<https://www.crd.york.ac.uk/PROSPERO/>). The first systematic evaluation will focus on the efficacy and safety of biosimilars compared to biologics (PROSPERO number: CRD42019137152), whereas the second systematic review will examine the clinical impact of

switching from reference biologics to biosimilars on the management of RA patients whose treatment has already been started (PROSPERO: CRD42019137155).

### **Adopted reporting standards**

Both systematic reviews will be conducted and presented in accordance with Preferred reporting items for systematic review and meta-analysis (PRISMA Statement) (22). Besides, since there are specific aspects related to the conduct, interpretation and reporting of equivalence and noninferiority trials, we will also adopt the US Agency for Healthcare Research and Quality recommendations (23).

### **Electronic searches**

Search strategies were built using controlled vocabulary according to each database and free-text terms based on the research question. We will use the following electronic databases (from inception to September 2020): MEDLINE via PubMed, EMBASE, Cochrane Central Register of Controlled Trials (CENTRAL), and Latin American and Caribbean Health Science (LILACS). A detailed description of the search strategy is available in **the additional file 2**.

### **Other sources**

We will also search for non-published or ongoing trials in the EU Clinical Trial Register (<https://www.clinicaltrialsregister.eu>), International Clinical Trials Registry Platform-World Health Organization (<http://apps.who.int/trialsearch/>) and Clinicaltrials (<https://clinicaltrials.gov/>). The search strategies to be used in these platforms are described in **the additional file 2**. When necessary, we will contact corresponding authors for supplementary information. Additionally, we will manually screen the references of all included trials as well as previous systematic reviews. Finally, we will employ Google Scholar and Epistemonikos (<https://www.epistemonikos.org/>) to retrieve relevant reports citing all relevant included articles. No language limitation will be imposed.

### **Eligibility Criteria**

#### **Types of biosimilars**

We will assess any biosimilars of adalimumab, etanercept and infliximab. We chose these three main biologics because they are the most prescribed first-line biologics DMARDs in RA (24). Also, these three DMARDs have the highest numbers of approved biosimilars for RA in the market (13, 25).

#### **Types of control interventions**

We will consider as control interventions the reference biologic drugs (i.e, adalimumab, etanercept and infliximab originals). No restrictions on dosages, treatment schedules, co-treatment or combined therapies will be imposed.

#### **Types of trials**

#### **Types of trials: Systematic review on efficacy and safety (systematic review 1)**

To assess the efficacy and safety of biosimilars ("biosimilarity") (26), we will include randomized controlled trials (RCTs) or quasi-randomized controlled trials. We will include all trials comparing

biosimilars to biologic drugs irrespective of the type of statistical design (superiority, equivalence, or non-inferiority). A quasi-randomised trial was defined as a prospective interventional study whose allocation sequence was not truly random (e.g., consecutive order, day of the week, date of birth etc.). For trials with a 2-part study design, we will consider results from the first period (biosimilarity) only to avoid carry over effects.

### **Types of trials: Systematic review on switching (systematic review 2)**

To assess the impact of switching on clinical outcomes of RA patients, we will include RCTs with two or multiple-part designs. The following four main designs of switching trials will be considered:

- Single-switch design (27, 28): Trials in which there is a single switch from each treatment to the other. All patients receive the study interventions in successive periods. Firstly, patients are randomly allocated to either a biosimilar or a biologic drug (first period). Then, in the second period, treatments are randomly switched in both directions (Group 1: biologic → biosimilar, Group 2: biosimilar → biologic; Group 3: biologic → biologic; Group 4: biosimilar → biosimilar).
- Transition design 1 (two non-switching groups as a control): Trials in which there is a single switch from one treatment (biologic drug) to another (biosimilar drug), but not the contrary. Firstly, patients are randomly allocated to either a biosimilar or a biologic drug (first period). Then, in the second period, the trial becomes a three-arm trial in which patients in the biologic drug group are re-randomized to either continue in the biologic group or to switch to the biosimilar drug treatment. Patients initially allocated to the biosimilar group continue to receive a biosimilar throughout the study period (Experimental group: biologic → biosimilar; Control arm 1: biologic → biologic; Control arm 2: biosimilar → biosimilar).
- Transition design 2 (randomized trials with an open label extension; single non-switching group as a control): Trials in which there is a single switch from a biologic drug to a biosimilar drug, but not the contrary. Firstly, patients are randomly allocated to either a biosimilar or a biologic drug (first period). Then, in open label extended phase (second period), all patients (intervention and control groups) receive the biosimilar drug (Experimental group: biologic → biosimilar; Control arm 1: biosimilar → biosimilar).
- Multiple switches design: Also known as interchangeability design (27, 28), in which multiple switches between treatments are allowed throughout the trial follow-up.

### **Type of participants**

Trials will be included if patients with RA had been diagnosed with validated and established international criteria. No limitation will be imposed on age, baseline RA severity, sex, lines of treatment (e.g., treatment-naïve patients or second line of treatment), or any other major demographic characteristics.

### **Types of outcomes measures**

All outcomes were prespecified in the registered PROSPERO synopses and were categorized into three types: efficacy (encompassing outcomes related to disease activity, functional capacity, quality of life

and structural damage progression), safety, and immunogenicity. For efficacy outcomes, we will extract data at the following timepoints: 1 month ( $\pm$  2 weeks), 3 months ( $\pm$  4 weeks), 6 months ( $\pm$  4 weeks), 8 months ( $\pm$  4 weeks), 12 months ( $\pm$  4 weeks), 36 months ( $\pm$  4 weeks) and 48 months ( $\pm$  4 weeks). For safety and immunogenicity outcomes, we will collect data from *longest follow-up* available.

### **Primary outcomes (efficacy)**

For both systematic reviews, we prespecified a primary outcome, a co-primary outcome, and all secondary outcomes. A co-primary outcome was adopted because the demonstration of superiority or equivalence in a single outcome is insufficient to support clinical decisions. The choice of primary and co-primary outcomes was decided on a panel composed of two RA specialists supervised by two researchers with experience in evidence synthesis. The rationale was to evaluate equivalence between biosimilars, and reference biologic drugs based both on physician-reported and patient-reported outcomes. Similar approaches have been used previously in RA trials (29, 30).

### **Systematic review on efficacy and safety (systematic review 1)**

The primary outcome will be treatment success at 6 months according to the American College of Rheumatology 20 (ACR20) (31). If trials report results at different time points, we will use the time point closest to 6 months. ACR20 is a composite and binary outcome, requiring patients to have  $\geq$  20% improvement in the number of swollen and tender joint plus  $\geq$  20% improvement in at least three out of five domains:

- Patient's assessment of pain (measured on a 100 mm visual analog scale [VAS]).
- Patient's global assessment of disease activity (measured on a 0-to-10 Likert scale).
- Physician's global assessment of disease activity (measured on a 0-to-10 Likert scale).
- Patient's self-assessment of physical function (Health Assessment Questionnaire - Disability Index [HAQ-DI]).
- C-Reactive Protein levels.

The co-primary outcome will be HAQ-DI, which assesses the functional status of patients through the evaluation of eight domains of daily-life activities (dressing and grooming, arising, eating, walking, hygiene, reach, grip, and activities) with 20 questions in total. For each question, there are four possible responses: 0 = without difficulty, 1 = with some difficulty, 2 = with much difficulty, and 3 = unable to do. The highest score reported for any component question in each domain determines the final score for that domain. By convention, the overall disability index is expressed on a 0 to 3 scale, representing an average score across the domains. A HAQ-DI of 0 indicates no functional disability, whereas a HAQ-DI of 3 denotes severe functional disability (32).

### **Systematic review on switching (systematic review 2)**

The primary outcome will be rate of treatment success at 6 months after the first switch (i.e. 6 months after re-randomization or 6 months after the first switch on the open-label extension phase) defined by

the ACR20 (dichotomous outcome). The co-primary outcome will be the HAQ-DI index also measured at 6 months after the first switch (continuous outcome). If outcome data are reported at different time points, we will use the time point closest to 6 months.

### **Secondary outcomes (efficacy, safety, and immunogenicity)**

#### **Secondary outcomes: Efficacy**

Secondary outcomes of efficacy will include disease activity, prevention of structural damage progression and quality of life measures:

- Measures of disease activity: the American College of Rheumatology criteria with 50% (ACR50) and 70% (ACR70) responses, simplified disease activity score (SDAI), clinical disease activity score (CDAI), disease activity score in 28 joints based on the erythrocyte sedimentation rate (DAS28-ESR), disease activity score in 28 joints with four components based on C-reactive protein (DAS28-CRP) and the numeric index of the ACR response (ACR-N).
- Functional capacity/quality of life: scores of HAQ-DI and the Medical Outcomes Study 36-item Short-Form Health Survey (SF-36) (physical and mental components summaries).
- Prevention of structural damage progression: scores of Sharp/ Van der Heijde or Sharp-Van Der Heijde Modified Score Method (mTRSS). A full description of secondary outcomes can be found in **the additional file 3**.

#### **Secondary outcomes: Safety**

We will evaluate the safety of biosimilars compared to biologics by the proportion of patients with treatment-emergent adverse events (TEAEs), serious TEAEs, infusion-related reactions (IRRs), hypersensitivity, malignancies, active tuberculosis, serious infections, all-cause mortality, and treatment-related mortality. Also, we will evaluate discontinuation rates in both treatments. A full list of safety outcomes can be found in **the additional file 3**.

#### **Secondary outcomes: Immunogenicity**

Immunogenicity will be evaluated by the proportion of patients with positive anti-drug antibodies (ADAs) and the proportion of patients with positive neutralizing antibodies (Nabs).

#### **Study Screening And Selection**

We developed a customized web platform for data extraction and curation using *Ragic* (<https://www.ragic.com/>). This database was carefully designed to simultaneously allow for study screening and selection, and data extraction for both systematic reviews 1 and 2.

During the screening phase, two review authors will independently evaluate titles and abstracts. Disagreements will be solved by a consensus. Next, for each study selected, full-length articles will be downloaded, and two independent reviewers will re-assess the eligibility of each pre-selected trial. In cases of disagreements, a third reviewer will be consulted. Reasons for exclusions will be described in detail in subsequent publications.

## **Data Extraction And Management**

### *Analysis population*

Trials may report two populations for the analysis: an intention-to-treat (ITT) population and a per-protocol (PP) population (23). In both systematic reviews 1 and 2, preference will be given to results based on the PP population, because of the conservative effect of the per-protocol approach on equivalence testing (23). Since there may be a substantial variety in the definition of what constitutes a PP population or an ITT analysis, we will collect and tabulate in detail the definition of PP and ITT used in each trial.

### *Numerical and graphical results*

All data will be extracted independently by two investigators. Discrepancies will be solved via a consensus. We will extract all pertinent quantitative information, including the number of participants at baseline, the number of participants analyzed, and measures of central tendency, variability, and precision. Specifically, whenever available, we will collect means, mean changes, the difference between means at follow-up, medians, standard deviations, interquartile ranges, standard errors, confidence intervals (and their coverage, e.g., 90% or 95%), P-values (one- or two-sides), and t statistics. These data will be used to approximate means and standard deviations when necessary (33). For continuous outcomes, we will use preferentially follow-up data but will use the mean change from baseline when follow-up values are not available (34).

Quantitative data from figures and graphs will be extracted independently using digitizing software (Digitizelt 2.2.2, Germany, <https://www.digitizeit.de/>). Estimates from the digitizing software will be averaged out to generate the final value. When necessary, data for the same trial will be extracted from multiple sources (e.g., multiple related publications and trial registries).

### *Ongoing trials*

We will summarize all identified ongoing trials, detailing the primary author, research question(s), methods, outcome measures, study start date along with an estimate of study completion date.

## **Assessment Of Risk Of Bias**

Two review authors will independently assess the risk of bias in the included studies. Each domain will be classified as being at a low, unclear, or high risk of bias. Disagreements will be resolved by consensus or discussion with a third reviewer. The studies will be assessed by outcome level. If the trial has one or more domains with a high risk of bias, it will be considered as a high risk of bias study. If the trial has more than two domains at uncertain risk of bias, we will judge the risk of bias to be uncertain. If the trial has a low risk of bias in all domains or one domain as uncertain bias, it will be considered as a low risk of bias study.

## Assessment of risk of bias in efficacy and safety trials

We will use criteria recommended by the Cochrane collaboration (Cochrane Risk of bias tool 1.0) (35). The following domains will be evaluated: random sequence generation, allocation concealment, blinding of participants and investigators, blinding of outcome assessors, and incomplete outcome data (PP and ITT population analysis). To specifically address equivalence or non-inferiority trials, we will refer to the recommendations by the US Agency for Healthcare Research and Quality (23) (**Table 1**). Specifically, we will assess inconsistent application of inclusion/exclusion criteria, patients selected for anticipated nonresponse or good response in one arm, patient behavior changes (poor adherence, use of concomitant treatments, and protocol violations), inadequate outcome measurement techniques and incomplete outcome data (PP and ITT population analysis: ITT population analysis may underestimate the treatment effect in equivalence/non-inferiority trials). More information on the criteria used in each domain can be found in **the additional file 4**.

**Table 1. Risk of bias domains to be evaluated on equivalence or non-inferiority trials.**

Type of Bias	Domain	Source
Selection bias	Random sequence generation	Cochrane RoB tool 1.0
	Allocation concealment	Cochrane RoB tool 1.0
	Inconsistent application of inclusion/exclusion criteria	US Agency for Healthcare Research and Quality
Performance bias	Blinding of participants and investigators	Cochrane RoB tool 1.0
	Participants behavior changes	US Agency for Healthcare Research and Quality
Detection bias	Blinding of outcome assessment	Cochrane RoB tool 1.0
	<i>Outcome measurement techniques</i>	US Agency for Healthcare Research and Quality
Attrition bias	Incomplete outcome data	Cochrane RoB tool 1.0 and US Agency for Healthcare Research and Quality

Sources: Cochrane Risk of Bias in randomized trials (RoB 1.0.) described in the Cochrane Handbook for Systematic Reviews of Interventions (35) and the US Agency for Healthcare Research and Quality recommendations (23).

## *Assessment of risk of bias in switching trials*

For switching trials, we will use the recommendations of Moots et al. (27) and the FDA guidance for Considerations in demonstrating interchangeability with a reference product (14). The six specific domains to be evaluated are:

- Randomized and blinded design with appropriate control arms;
- At least 1-way switch from originator to biosimilar;
- The assessment of immunogenicity;
- The washout period between treatment;
- Enough power to assess efficacy and safety (equivalence phase);
- Enough follow-up periods.

More information about the criteria of judgments can be seen on **the additional file 5**.

## **Data synthesis**

### **Effect size measures**

For binary outcomes, we will combine study estimates using the relative risk (RR) as a measure of effect. For continuous outcomes, we will use the standardized mean difference (SMD) defined as the bias-adjusted method of Hedges. SMD will be used because has similar statistical power and is more generalizable than the mean difference (36).

### **Meta-analysis models**

Main analyses will be based on the random-effects model with the restricted maximum-likelihood estimator for the between-study variance (37). A random-effects model was prespecified as the primary model of analysis since we anticipated variability in the design and population characteristics of the included trials. Results for a fixed-effects model (inverse-variance method) will be presented simultaneously as a sensitivity analysis.

### **Statistical heterogeneity**

We will test for the presence of statistical heterogeneity across trial estimates using Cochran's Q test (38) and the magnitude of the between-trial heterogeneity will be quantified with the  $I^2$  metric (39). When feasible (i.e., 10 or more trials), we will investigate potential sources of statistical heterogeneity with the random-effects meta-regression analysis and subgroups analysis. Explanatory variables to be included in meta-regression models are described below.

### **Small-study and publication biases**

We will investigate the association between trial size (precision) and treatment effects in contour-enhanced funnel plots, contrasting the effect estimates on the horizontal axis against their standard

errors on the vertical axis, accompanied by a regression test for asymmetry. Furthermore, for continuous outcomes, small-study biases will be investigated by Egger's regression test, whereas for binary outcomes we will use Harbord's test (40).

## Equivalence testing

### Criteria to claim equivalence

Equivalence will be evaluated and interpreted using predefined margins of equivalence (**Fig. 1**). Upper and lower equivalence bounds were specified based on the smallest effect size of clinical importance. These values were computed from large placebo-controlled trials and validated by two rheumatologists with extensive experience treating patients with RA. Prespecified boundaries of equivalence will be applied to the primary and co-primary outcomes only.

Based on random-effects models, lower and upper confidence limits will be calculated. For a specific outcome (i.e. either ACR20 or HAQ-DI), if the two-sided 95% confidence interval for the difference in effect is completely contained within the prespecified boundaries of equivalence, biosimilars and biologics will be considered equivalent. However, the rejection of non-equivalence (for both ACR20 and HAQ-DI outcomes) will be required for biosimilars to be declared overall equivalent to biologic drugs. Secondary outcomes will be examined through standard superiority tests (two-tailed).

### Figure 1. Boundaries of equivalence (dashed lines) for a two-sided 95% confidence interval of the treatment difference.

Panel A, equivalence margins for ACR20 (primary outcome). Panel B, equivalence margins for HAQ-DI (co-primary outcome). If the summary 95% CI lies within the gray regions, the null hypothesis will be rejected, and equivalence will be claimed.

### Margins of equivalence: ACR20 criteria

For the ACR20 outcome, we prespecified equivalence margins to preserve 90% of the effects observed with biologics using the fixed-margin method (41). Specifically, we calculated the equivalence margins as:

$$\text{Equivalence margins} = e^{\pm m} \quad (1)$$

with

$$m = \log(RR^{PE-1}) \quad (2)$$

where PE stands for the preserve effect (range: 0 to 1 [100%]). Based on a network meta-analysis by Guyott et al. (42) that included 11 randomized trials with 3762 patients who were unresponsive to methotrexate, the RR under a random-effects model for ACR20 at 6 months for any biologics (adalimumab/etanercept/infliximab) vs placebo was approximately 1.80. Similar estimates were

obtained considering a combination of both methotrexate-naïve and methotrexate unresponsive patients, in which the frequentist summary RR of achieving ACR20 was 1.81 (random-effects model, 13 trials, 7087 patients) for the comparison adalimumab/etanercept/infliximab vs placebo (43). Thus, biosimilars will be considered equivalent to biologics if the 95% confidence limits for the summary RR lies within the 0.94 and 1.06 interval (**Figure 1A**).

Margins of equivalence: HAQ-DI

For HAQ-DI, which is a continuous outcome (with higher scores meaning worse function status), equivalence margins were constructed under the clinical assumption that an increase equal or larger than 0.15 units over 1 year on the 0-to-3 HAQ-DI scale is considered clinically perceptible by the patients (44). Based on trials encompassing a diversified group of RA patients that received adalimumab, etanercept or infliximab (1816 patients) (45-47), we estimated the HAQ-DI population standard deviation after a 24/26-week treatment to be approximately 0.69. Therefore, on a standardized mean difference scale, the 0.15 units difference is approximately equivalent to  $\pm 0.22$  standard deviations (**Figure 1B**).

### **Analysis of subgroups and meta-regression**

We will perform pre-specified subgroup analyses. When feasible, the following subgroup analysis will be conducted:

- Type of molecule (infliximab vs. etanercept vs. adalimumab);
- Concomitant use of synthetic disease-modifying antirheumatic drugs (yes or no);
- Sample size (average of >100 patients per group vs. <100 patients per group);
- Allocation concealment (low risk vs. high risk/unclear risk);
- Trial duration (equivalence studies) (3 months vs 6 months vs. 12 months);
- Type of design (multiple switching/switching studies vs. transitional studies)
- Funding independent of industry (yes vs no/unclear);
- Publication status (Published vs. unpublished).

The above-mentioned variables will also serve as explanatory variables in meta-regression models.

### **Assessment of Overall Certainty of Evidence**

The overall certainty of evidence will be assessed by two investigators and will be based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (48). Assessments will be conducted by outcome. Disagreements will be settled by consensus or discussion with a third reviewer. The certainty of evidence of each outcome will be graded as very low, low, moderate, or high. The following domains will be assessed:

- Study design and risk of bias
- Inconsistency

- Indirectness
- Imprecision
- Other factors (e.g., reporting bias, publication bias).

*Ranking of outcomes by their relative importance*

We have adopted the recommendations from the GRADE handbook for selecting and rating the importance of outcomes (26). Specifically, for systematic review 1 and 2, we ranked each outcome as “critical”, “important but not critical”, and “limited importance to decision-making” (**Table 2**).

**Table 2. Ranking of outcomes.**

Outcome	Rating scale 1-9* (importance of the outcome for decision-making)	Type of outcome
<b>Disease activity/Clinical response</b>		
CDAI	9 (critical)	Continuous or binary
DAS-28 ESR	8.5 (critical)	Continuous or binary
SDAI	8.5 (critical)	Continuous or binary
ACR 20	8 (critical)	Binary
ACR 50	8 (critical)	Binary
DAS-28 CRP	8 (critical)	Continuous or binary
ACR-N	7.5 (critical)	Continuous
ACR 70	6 (important but not critical)	Binary
EULAR response	5 (important but not critical)	Binary
<b>Safety</b>		
TEAE	9 (critical)	Binary
Serious TEAE	9 (critical)	Binary
Death related to treatment	9 (critical)	Binary
IRRs	9 (critical)	Binary
Hypersensitivity	9 (critical)	Binary
Active tuberculosis	9 (critical)	Binary
Serious infections	9 (critical)	Binary
Death all causes	8 (critical)	Binary
Overall Discontinuation rate	8.5 (critical)	Binary
Malignancies	8 (critical)	Binary
Fatigue	7 (critical)	Binary
<b>Function capacity/Quality of Life</b>		
HAQ-DI	8.5 (critical)	Continuous
SF-36 score - physical component summary	8 (critical)	Continuous
SF-36 score- mental component summary	8 (critical)	Continuous
<b>Structural Damage Progression</b>		
mTRSS or Sharp/van der Heijde score	5.5 (important but not critical)	Continuous
<b>Immunogenicity</b>		
Incidence of ADAs	5 (important but not critical)	Binary
NABs -positive	5 (important but not critical)	Binary

\*Based on the mean average ranking of two rheumatologists with extensive clinical experience in treating RA patients and a physical therapist with advanced training in evidence synthesis. ACR: the American College of Rheumatology; CRP: C-Reactive Protein level; HAQ-DI: Health Assessment Questionnaire - Disability Index; VAS: visual analog scale; SDAI: Simplified Disease Activity Score DA: disease activity; CDAI: Clinical Disease Activity Score DAS28-ESR: Disease Activity Score in 28 joints based on the erythrocyte sedimentation rate; DAS28-CRP: Disease Activity Score in 28 joints, four components based on C-reactive protein; ACR-N: The numeric index of the ACR response; EULAR: European League Against Rheumatism; SF-36: The Medical Outcomes Study 36-item Short-Form Health Survey; mTRSS: Sharp/van der Heijde score; TEAE: Overall Treatment Emergent Adverse Event.

The ranking was conducted through consultations with two clinical specialists (rheumatologists) and a physical therapist specialized in evidence synthesis. These professionals were invited to participate based on their clinical experience, academic background, and the lack of any conflict of interest. Briefly, before the scoping meeting, based on previous systematic reviews, we screened the list of outcomes (both primary and secondary outcomes) of 14 trials that we knew a priori that met all the eligibility criteria. Subsequently, we created an integrated list of all outcomes and categorized them into five main domains:

- Disease activity/Clinical response
- Function capacity/Quality of Life
- Structural Damage Progression
- Immunogenicity
- Safety

Through an iterative approach in a single scoping meeting, each member of the collaborative working group ranked outcomes independently. Conflicting ranking cases were discussed jointly until a consensus was reached.

## Discussion

At the time of writing this manuscript, more than 16 biosimilars of adalimumab, etanercept and infliximab had been approved in the United States, Europe, Canada, and Latin America for the treatment of RA (49–52). Our two systematic reviews will comprehensively assess the efficacy, safety, and immunogenicity of these biosimilars compared to their originator molecules in the management of RA patients and examine the clinical consequences of switching from biologics to biosimilars for RA patients.

Of note, our two systematic reviews will apply an integrated approach by combining previously accumulated data and clinical expert information, which will enable us to evaluate the equivalence between biosimilars and reference biologics for the treatment of RA with a clinically oriented perspective. Furthermore, by examining the randomized evidence on the effects of a wide range of biosimilars, we will be also able to address whether switching from reference biologics to biosimilars or vice and versa, in general, results in similar clinical benefits with acceptable immunogenicity and safety profiles. As a result, our findings hold great potential to affect not only the therapeutic regimen of RA patients that will use a DMARD for the first time but also the treatment of those that will have their treatment substituted from a biologic to a biosimilar, or from a biosimilar to a biologic.

Overall, we expect that our results will guide clinicians, researchers, decision-makers, stakeholders, and policymakers about the efficacy, safety, immunogenicity, and substitution and interchangeability of currently marketed biosimilars for the treatment of RA patients and assist health care systems to employ more efficiently the scarce existing resources.

# Abbreviations

RA: rheumatoid arthritis; DMARDs: disease-modifying antirheumatic drugs; FDA: the U.S. Food and Drug Administration; PRISMA-P: the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols guidelines; PROSPERO: the International Prospective Register of Systematic Reviews; PRISMA: Preferred reporting items for systematic review and meta-analysis; US: United States; EU: European; CENTRAL: Cochrane Central Register of Controlled Trials; LILACS: Latin American and Caribbean Health Science; RCT: randomized controlled trials; ACR20: the American College of Rheumatology 20 criteria; ACR 50: the American College of Rheumatology 50 criteria; ACR70: the American College of Rheumatology 70 criteria; VAS: visual analog scale; HAQ-DI: Health Assessment Questionnaire - Disability Index; DAS28-ESR: disease activity score in 28 joints based on the erythrocyte sedimentation rate; DAS28-CRP: disease activity score in 28 joints with four components based on C-reactive protein; ACR-N: the numeric index of the ACR response; SF-36: the Medical Outcomes Study 36-item Short-Form Health Survey; mTRSS: Sharp-Van Der Heide Modified Score Method; TEAE: treatment-emergent adverse events; IRR: infusion-related reactions ADAs: anti-drug antibodies; Nabs: neutralizing antibodies; PP: per-protocol; ITT: intention-to-treat; RR: relative risk; SMD: the standardized mean difference; PE: preserve effect; Log: logarithm; GRADE: the Grading of Recommendations Assessment, Development and Evaluation System;

# Declarations

## **Ethics approval and consent to participate**

Not applicable.

## **Consent for publication**

Not applicable.

## **Availability of data and materials**

## **Competing interests**

The authors declare that they have no competing interests.

## **Funding**

No funding to declare related to this study.

## **Authors' contributions**

BOA, TVP and PCS conceived and designed the study. BOA and MOA created the electronic search strategies. BOA and MOA will conduct the study screening and selection, extraction data, assessment of bias and overall quality of evidence. In case of disagreements, TVP will be the third reviewer. BOA, MOA and TVP customized the web platform for data extraction. ACMR, DCOA and MOA composed the panel of

experts for selecting and rating the importance of outcomes. HAOJ provided methodological advice on rheumatoid arthritis and made important intellectual contributions for the manuscript. BOA, TVP and PCS drafted the manuscript, which was edited and revised by all authors. TVP and PCS provided supervision. All authors read and approved the final version of the manuscript.

## References

1. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham III CO, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis & rheumatism*. 2010;62(9):2569-81.
2. Gabriel SE. The epidemiology of rheumatoid arthritis. *Rheum Dis Clin North Am*. 2001;27:269 –82.
3. Safiri S, Kolahi AA, Hoy D, Smith E, Bettampadi D, Mansournia MA, et al. Global, regional and national burden of rheumatoid arthritis 1990–2017: a systematic analysis of the Global Burden of Disease study 2017. *Annals of the rheumatic diseases*. 2019;78(11):1463-71.
4. Smolen JS, Landewé R, Bijlsma J, Burmester G, Chatzidionysiou K, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Annals of the rheumatic diseases*. 2017;76(6):960-77.
5. Singh JA, Saag KG, Bridges Jr SL, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis & rheumatology*. 2016;68(1):1-26.
6. Singh JA, Hossain A, Ghogomu ET, Mudano AS, Tugwell P, Wells GA. Biologic or tofacitinib monotherapy for rheumatoid arthritis in people with traditional disease-modifying anti-rheumatic drug (DMARD) failure: a Cochrane Systematic Review and network meta-analysis (NMA). *Cochrane Database of Systematic Reviews*. 2016(11).
7. Nam JL, Takase-Minegishi K, Ramiro S, Chatzidionysiou K, Smolen JS, Van Der Heijde D, et al. Efficacy of biological disease-modifying antirheumatic drugs: a systematic literature review informing the 2016 update of the EULAR recommendations for the management of rheumatoid arthritis. *Annals of the rheumatic diseases*. 2017;76(6):1113-36.
8. Daien CI, Morel J. Predictive factors of response to biological disease modifying antirheumatic drugs: towards personalized medicine. *Mediators of inflammation*. 2014:1-11.
9. Souto A, Maneiro JR, Gomez-Reino JJ. Rate of discontinuation and drug survival of biologic therapies in rheumatoid arthritis: a systematic review and meta-analysis of drug registries and health care databases. *Rheumatology*. 2016;55(3):523-34.
10. Ramiro S, Gaujoux-Viala C, Nam JL SJ, Buch M, Gossec L, Van Der Heijde D, et al. Safety of synthetic and biological DMARDs: a systematic literature review informing the 2013 update of the EULAR recommendations for management of rheumatoid arthritis. *Annals of the rheumatic diseases*. 2014;73(3):529-35.

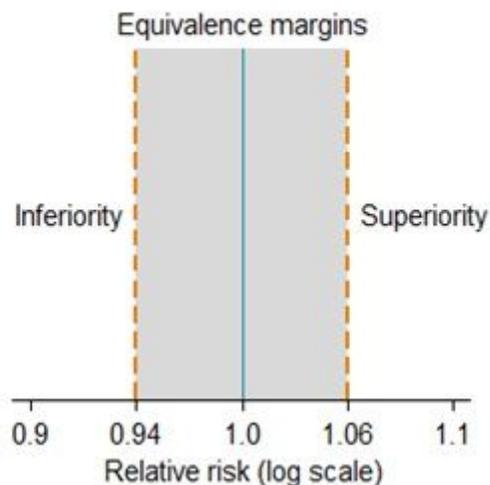
11. Hamann P, Holland R, Hyrich K, Pauling JD, Shaddick G, Nightingale A, et al. Factors Associated With Sustained Remission in Rheumatoid Arthritis in Patients Treated With Anti-Tumor Necrosis Factor. *Arthritis care & research*. 2017;69(6):783-93.
12. Stiff KM, Cline A, Feldman SR. Tracking the price of existing biologics when drugs enter the market. *Expert Review of Pharmacoeconomics & Outcomes Research*. 2019;19(4):375-7.
13. Allocati E, Gerardi C, Garattini S, Banzi R. Clinical evidence supporting the marketing authorization of biosimilars in Europe. *European Journal of Clinical Pharmacology*. 2020:1-10.
14. US Food and Drug Administration. Considerations in demonstrating interchangeability with a reference product: guidance for industry. In: Services USDoHaH, editor. Silver Spring, MD: US Food and Drug Administration; 2019.
15. Kim H, Alten R, Avedano L, Dignass A, Gomollón F, Greveson K, et al. The Future of Biosimilars: Maximizing Benefits Across Immune-Mediated Inflammatory Diseases. *Drugs*. 2020;80(2):99-113.
16. Leonard E, Wascovich M, Oskouei S, Gurz P, Carpenter D. Factors affecting health care provider knowledge and acceptance of biosimilar medicines: a systematic review. *Journal of managed care & specialty pharmacy*. 2019;25(1):102-12.
17. Sarnola K, Merikoski M, Jyrkkä J, Hämeen-Anttila K. Physicians' perceptions of the uptake of biosimilars: a systematic review. *BMJ open*. 2020;10(5):e034183.
18. Gasteiger C, Lobo M, Dalbeth N, Petrie KJ. Patients' beliefs and behaviours are associated with perceptions of safety and concerns in a hypothetical biosimilar switch. *Rheumatology International*. 2020:1-9.
19. Castañeda-Hernández G, Sandoval H, Coindreau J, Rodriguez-Davison LF, Pineda C. Barriers towards effective pharmacovigilance systems of biosimilars in rheumatology: a Latin American survey. *Pharmacoepidemiology and drug safety*. 2019;28(8):1035-44.
20. O'Callaghan J, Barry SP, Birmingham M, Morris JM, Griffin BT. Regulation of biosimilar medicines and current perspectives on interchangeability and policy. *European journal of clinical pharmacology*. 2019;75(1):1-11.
21. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews*. 2015;4(1).
22. Moher D, Liberati A, Tetzlaff J, Altman DG, Group. TP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 2009;6(7):e1000097.
23. Treadwell JR, Uhl S, Tipton K, Shamlilian T, Viswanathan M, Berkman ND, et al. Assessing equivalence and noninferiority. *Journal of clinical epidemiology*. 2012;65(11):1144-9.
24. Silvagni E, Bortoluzzi A, Carrara G, Zanetti A, Govoni M, Scirè CA. Comparative effectiveness of first-line biological monotherapy use in rheumatoid arthritis: a retrospective analysis of the RECORD-linkage On Rheumatic Diseases study on health care administrative databases. *BMJ open*. 2018;8(9):e021447.

25. Rathore AS, Bhargava A. Biosimilars in Developed Economies: Overview, Status, and Regulatory Considerations. *Regulatory Toxicology and Pharmacology*. 2020;110:104525.
26. Declerck P, Danesi R, Petersel D, I. J. The language of biosimilars: clarification, definitions, and regulatory aspects. *Drugs*. 2017;77(6):671-7.
27. Moots R, Azevedo V, Coindreau JL, Dörner T, Mahgoub E, Mysler E, et al. Switching between reference biologics and biosimilars for the treatment of rheumatology, gastroenterology, and dermatology inflammatory conditions: considerations for the clinician. *Current rheumatology reports*. 2017;19(6):37.
28. Faccin F, Tebbey P, Alexander E, Wang X, Cui L, Albuquerque T. The design of clinical trials to support the switching and alternation of biosimilars. *Expert Opinion on Biological Therapy*. 2016;16(12):1445-53.
29. Fleischmann R, Kremer J, Cush J, Schulze-Koops H, Connell CA, Bradley JD, et al. Placebo-controlled trial of tofacitinib monotherapy in rheumatoid arthritis. *N Engl J Med*. 2012;367:495-507.
30. Keystone EC, Genovese MC, Klareskog L, Hsia EC, Hall ST, Miranda PC, et al. 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. *Annals of the rheumatic diseases*. 2009;68(6):789-96.
31. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727–35.
32. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. *The Journal of rheumatology*. 2003;30(1):167-78.
33. Weir CJ, Butcher I, Assi V, Lewis SC, Murray GD, Langhorne P, et al. Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. *BMC medical research methodology*. 2018;18(1):25.
34. da Costa BR, Nüesch E, Rutjes AW, Johnston BC, Reichenbach S, Trelle S, et al. Combining follow-up and change data is valid in meta-analyses of continuous outcomes: a meta-epidemiological study. *Journal of clinical epidemiology*. 2013;66(8):847-55.
35. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0: The Cochrane Collaboration; 2011* [Available from: <https://training.cochrane.org/handbook/archive/v5.1/>].
36. Takeshima N, Sozu T, Tajika A, Ogawa Y, Hayasaka Y, Furukawa TA. Which is more generalizable, powerful and interpretable in meta-analyses, mean difference or standardized mean difference? *BMC medical research methodology*. 2014;14(1):30.
37. White IR. Multivariate random-effects meta-analysis. *The Stata Journal*. 2009;9(1):40-56.
38. Pereira TV, Patsopoulos NA, Salanti G, Ioannidis JP. Critical interpretation of Cochran's Q test depends on power and prior assumptions about heterogeneity. *Research Synthesis Methods*. 2010;1(2):149-61.

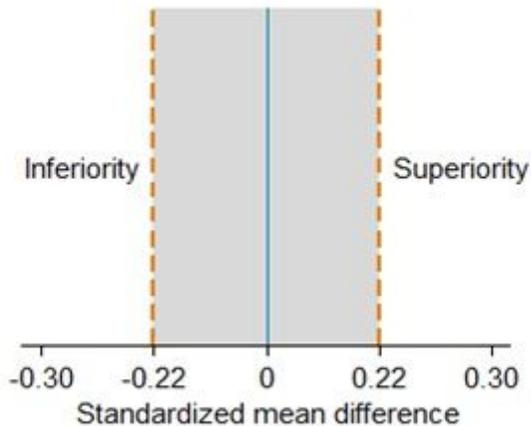
39. Borenstein M, Higgins JP, Hedges LV, Rothstein HR. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research synthesis methods*. 2017;8(1):5-18.
40. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *Bmj*. 2011;343.
41. US Food and Drug Administration. Non-inferiority clinical trials to establish effectiveness: guidance for industry. In: Services USDoHaH, editor. Silver Spring, MD: US Food and Drug Administration; 2016.
42. Guyot P, Taylor PC, Christensen R, Pericleous L, Drost P, Eijgelshoven I, et al. Indirect treatment comparison of abatacept with methotrexate versus other biologic agents for active rheumatoid arthritis despite methotrexate therapy in the United Kingdom. *The Journal of rheumatology*. 2012;39(6):1198-206.
43. Alonso-Ruiz A, Pijoan JI, Ansuategui E, Urkaregi A, Calabozo M, Quintana A. Tumor necrosis factor alpha drugs in rheumatoid arthritis: systematic review and metaanalysis of efficacy and safety. *BMC musculoskeletal disorders*. 2008;9:1.
44. Pope JE, Khanna D, Norrie D, Ouimet JM. The minimally important difference for the health assessment questionnaire in rheumatoid arthritis clinical practice is smaller than in randomized controlled trials. *The Journal of rheumatology*. 2009;36(2):254-9.
45. Maini R, St Clair EW, Breedveld F, Furst D, Kalden J, Weisman M, et al. Infliximab (chimeric anti-tumour necrosis factor  $\alpha$  monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. *The Lancet*. 1999;354(9194):1932-9.
46. Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *The Lancet*. 2004;363(9410):675-81.
47. Strand V, Rentz AM, Cifaldi MA, Chen N, Roy S, D. R. Health-related Quality of Life Outcomes of Adalimumab for Patients with Early Rheumatoid Arthritis: Results from a Randomized Multicenter Study. *The Journal of rheumatology*. 2012;39(1):63-72.
48. Schünemann H, Brožek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations: The GRADE Working Group; 2013 [updated October 2013]. Available from: <https://gdt.gradepro.org/app/handbook/handbook.html#h.1i2bwkm8zpj0>.
49. US Food and Drug Administration. FDA-Approved Drugs 2020 [Available from: <https://www.accessdata.fda.gov/scripts/cder/daf/>].
50. European Medicines Agency. Medicines: EMA; 2020 [Available from: <https://www.ema.europa.eu/en/medicines>].
51. Health Canada. Drug Product Database online query: Health Canada; 2020 [Available from: <https://health-products.canada.ca/dpd-bdpp/index-eng.jsp>].

## Figures

A



B



**Figure 1**

Boundaries of equivalence (dashed lines) for a two-sided 95% confidence interval of the treatment difference. Panel A, equivalence margins for ACR20 (primary outcome). Panel B, equivalence margins for HAQ-DI (co-primary outcome). If the summary 95% CI lies within the gray regions, the null hypothesis will be rejected, and equivalence will be claimed.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [Additionalfile2.docx](#)
- [Additionalfile3.docx](#)
- [Additionalfile4.docx](#)
- [Additionalfile5.docx](#)