The impact of patient global assessment in the definition of remission as a predictor of long-term radiographic damage in patients with rheumatoid arthritis: protocol for an individual patient data meta-analysis

Ferreira RJO[1], Welsing PMJ[2], Gossec L[3], Jacobs JWG[2], Machado PM[4], Ndosi M[5], van der Heijde D[6], da Silva JAP[7]

ACTA REUMATOL PORT. 2018;43:52-60

ABSTRACT

Background: Remission is the target for management of rheumatoid arthritis (RA) and intensification of immunosuppressive therapy is recommended for those that do not achieve this status. Patient global assessment (PGA) is the single patient reported outcome considered in the American College of Rheumatology/European League Against Rheumatism remission criteria, but its use as target has been questioned. The primary aim of this study is to assess whether excluding PGA from the definition of disease remission changes the association of disease remission with long-term radiographic damage and physical function in patients with RA.

Methods: Individual patient data meta-analysis using data from randomized controlled trials of biological and targeted synthetic agents, identified through ClinicalTrials.gov and PubMed. Different remission states will be defined: (i) 4v-remission [tender (TJC28) and swollen (SJC28) 28-joint counts both≤1, C-reactive protein (CRP)≤1 (mg/dl), and PGA≤1 (0-10 scale)], (ii) 4v-near-remission (TJC28≤1, SJC28≤1, CRP≤1, and PGA>1), (iii) non-remission (TJC28>1 or SJC28>1 or CRP>1), all mutually exclusive, and (iv) 3v-remission (TJC28≤1, SJC28≤1, CRP≤1). Likelihood ratios will be used to descriptively compare whether meeting the 3v and 4v-remission criteria in a single visit (at 6 or 12 months) predicts good outcome in the second year (1-2y). Differences in the predictive value of PGA in the definition of remission will be assessed by comparing the three mutually exclusive disease states using logistic regression analysis. Good outcome is defined primarily by radiographic damage (no deterioration in radiographic scores, whatever the instrument used in each trial), and secondarily by functional disability (Health Assessment Questionnaire consistently ≤0.5 and no deterioration), and their combination ("overall good outcome"). Additional analyses will consider longer periods over which to (concurrently) define remission status and outcome (between 1-5y and 1-10y), different cut-offs to define good radiographic outcome (change ≤0.5, ≤3 and ≤5 in radiographic score), sustained remission and the influence of treatment and other clinical factors.

Discussion: If 4v-remission and 4v-near-remission are associated with a similar probability of good outcomes, particularly regarding structural damage, the 3v-remission (excluding PGA) could be adopted as the target for immunosuppressive therapy. Patients' perspectives would remain essential, but assessed separately from disease activity, using instruments adequate to guide adjunctive therapies.

Systematic review registration: PROSPERO, CRD42017057099.

Keywords: Rheumatoid arthritis; Outcome research; Patient global assessment; Patient reported outcomes; Disease activity; Remission; Near-remission; Radiographic damage; Function; Individual patient data meta-analysis.

1. Rheumatology Department, Centro Hospitalar e Universitário de Coimbra EPE
2. Department of Rheumatology & Clinical Immunology, University Medical Center, Utrecht
3. Rheumatology Department, AP-HP, Pitié Salpêtrière Hospital
4. Centre for Rheumatology & MRC Centre for Neuromuscular Diseases, University College London
5. Centre for Health and Clinical Research, Department of Nursing and Midwifery, University of the West of England
6. Rheumatology department, Leiden University Medical Centre
7. Rheumatology department, Centro Hospitalar e Universitário de Coimbra EPE

INTRODUCTION

Disease remission or low disease activity is now a realistic therapeutic target in patients with rheumatoid arthritis (RA)[1, 2]. The American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR) proposed two alternative definitions of remission[3]: one based in Boolean criteria and another on the Simplified Disease Activity Index (SDAI). The Boolean-based definition requires that tender 28-joint count (TJC28), swollen 28-joint count (SJC28), C-Reactive Protein (CRP, in mg/dl), and patient global assessment (PGA, 0–10 scale) are all ≤1. The SDAI criterion requires that the sum of SJC28, TJC28, PGA, CRP and physician/observer global assessment [PhGA] is ≤3.3. These definitions have been recommended for use in daily care of RA[2].

PGA is the sole patient reported outcome (PRO) included in the recommended definitions of remission, having the same weight as other criteria[1-3]. Its inclusion was justified because it represents the patient's perspective and because it proved to discriminate between active and control intervention in clinical trials[3]. However, several studies[4-13] have shown that PGA is not solely influenced by RA disease activity, but also by sociodemographic features, geographic area, and cultural and ethnic aspects, reflecting the fact that PGA scores can be influenced by physical and psychological factors (including disease perception), comorbidities and fibromyalgia, among others[12].

Patients that fail only one of the four Boolean criteria have been called "near-misses"[14] or "near-remission" (designation applied when PGA is the solely criteria >1)[13]. Previous studies[4,13-15] demonstrated that the proportion of near-remission patients could vary from 14.4% up to 38.2%[13], which can represent up to four times the proportion of patients in remission[13]. Following current treatment guidelines[1,16] this state of near-remission would justify intensification of immunosuppressive treatment if based on a shared decision between patient and rheumatologist, taking structural damage, comorbidities or contraindications into account (Overarching Principles A and B)[2].

The importance of incorporating PROs in the overall management plan is indisputable. However, whether PGA conveys information that should be taken into account when considering changing immunosuppressive regimens in patients that have otherwise achieved a remission state based on TJC, SJC and CRP remains unclear. A recent study from our group[12] showed that PGA was weakly correlated with "more" objective disease activity measures (SJC28, TJC28, CRP) but the correlations were strong with pain, fatigue, function, comorbidities, depression and anxiety, and were also significant with other dimensions such as happiness and personality dimensions (weak correlations). Furthermore, it was shown that PGA correlations differ according to disease activity state, with pain, function and joint counts having stronger correlation with PGA when patients are in non-remission.

Immunosuppressive therapy, including biologic disease modifying antirheumatic drugs (bDMARDs), has been shown to improve PGA as disease activity evolves towards remission. However, in cases where PGA is being mainly driven by factors not related to RA, immunosuppressive therapy may not be able to lower PGA ≤1, despite SJC, TJC and CRP scores ≤1 having already been achieved. From this stage onwards, PGA is not dependent on disease activity and the inability to improve it further should not be interpreted as a failure of the immunosuppressive therapy.

Progression of joint damage is one of the most important outcome measures in RA, because it reflects historic disease activity, is associated with decline of physical function over time, and can be reliably assessed[17]. A recent systematic literature review (SLR)[18] investigated the clinical predictors of radiographic progression in RA, including disease activity indices and their individual components. Regarding the individual components, only SJC and acute phase reactants were associated with radiographic progression. Regarding PGA, the authors concluded that "published data for GH [patient's general health], PGA and EGA [evaluator's global assessment] are limited and do not support their use as unique tools related to progression of joint damage"[18]. The data analysed included two randomized controlled trials (RCTs) and two prospective cohorts, with 1 to 3 years of follow-up, including patients receiving conventional synthetic (cs)DMARDS. However, radiographic progression may be different in patients receiving bDMARDs. A subsequent observational study[19] with 527 patients with early RA, followed for 8 years, demonstrated that 31% of patients reaching remission according to the ACR/EULAR Boolean criteria, at 1, 2, 5, and 8 years (sustained remission) had radiographic progression (>1 unit/ /year). There was no significant contribution of PGA to the prediction of radiographic progression[19].

These observations suggest that the concept of "disease remission" should not include PGA for the

purposes of guiding immunosuppressive therapy. Such a definition of remission, i.e. excluding PGA from ACR/EULAR Boolean definition, might be designated as "remission 3 variables" or "3v-remission". Long-term longitudinal studies, looking at more objective outcomes are required to support this change in the definition of remission.

The primary aim of this study is to assess whether excluding PGA from the definition of disease remission changes the association of disease remission with long-term radiographic damage and physical function in patients with RA.

METHODS/DESIGN

This is an individual patient data (IPD) meta-analysis of published RCTs selected from a systematic literature search.

PROTOCOL AND REGISTRATION

This study protocol was registered in PROSPERO with the number CRD42017057099.[20] The results will be reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Individual Participant Data (PRISMA-IPD), and this checklist (Appendix 1) also served as guidance for writing this protocol.

Ethical approval to this study was granted by the Centro Hospitalar e Universitário de Coimbra Ethics Committee (CHUC-047-17).

ELIGIBILITY CRITERIA

Type of studies

This study will include recently published RCTs, and their long-term extensions (≥ 2 years), which evaluate the efficacy of csDMARDs, bDMARDs or targeted synthetic (ts) DMARD on radiographic damage in patients with RA.

All RCTs assessing radiographic damage assess also physical function as a secondary outcome. Studies with less than two years of follow-up will be excluded.

Participants

Both men and women with diagnosis of RA, and fulfilling the 1987 ACR criteria or the 2010 ACR-EULAR criteria for RA [21, 22], will be included.

Types of interventions

Studies testing the efficacy of any bDMARD [Tumor necrosis factor (TNF) inhibitors; Interleukin (IL) inhibitors; B-cell inhibitors; and T-cell inhibitors] or tsDMARD [janus kinase (JAK) inhibitors] will be included. All routes of drug administration will be considered. Studies testing DMARDs dose spacing or suspension will be excluded.

Types of assessments

As a minimum, studies will need to have assessed SJC28, TJC28, CRP, and PGA (in order to determine the Boolean-based criteria) at baseline and at 6 and 12 months and the radiographic damage assessment and physical function at baseline, 12 and 24 months.

IDENTIFYING STUDIES

Studies of interest were searched by one researcher (RF), between November and December of 2016, from the ClinicalTrials.gov registry. The following search strategy was used, without limits: "Rheumatoid arthritis" AND ("radiographic damage" OR "radiographic progression" OR "joint damage"). A second search was also performed in PubMed MEDLINE, for the same time-period, using also the pharmacological names of bDMARDs and tsDMARD as search terms. Additionally, local medical contacts of pharmaceutical companies were approached in order to identify possible published studies missed in previous searches and how to get access to their IPD. A summary table with the results of this search is presented in Appendix 2.

STUDY SELECTION PROCESS

Full papers of the identified studies were obtained and checked against the inclusion criteria by two researchers (RF and JAPS) independently.

DATA COLLECTION PROCESSES

Not all published studies have IPD available, which can be the case if patients did not give permission for broader research than the original study. The timing of data availability after study publication may also vary according to data holder policy[23]: while some companies make the data available immediately after the first publication, others only do it after the investigational product has been approved for use in both the United States and European Union. Thus, after study selection (Appendix 2) a research proposal was submitted to all sponsors of those trials, asking for IPD access.

DATA ITEMS

In addition to the radiographic damage score and

physical function (outcomes), and to SJC28, TJC28, CRP, and PGA (for remission definition) the following variables will be extracted from the trials: (i) patient characteristics – gender, age at baseline; (ii) clinical characteristics – disease duration at baseline, rheumatoid factor (RF) status, anti-citrullinated peptide antibody (ACPA) status, and treatment arm (iii) trial/visit information variables, namely anonymised patient identification (ID) code, visit number or sequence, and visit date. Appendix 3 provides a list off all essential variables that will be extracted from each trial.

IPD INTEGRITY

Any important issues identified when checking IPD, such as data plausibility, consistency, completeness or baseline imbalance will be reported and summarised using a PRISMA-IPD flow diagram.

RISK OF BIAS ASSESSMENT OF INDIVIDUAL STUDIES

The quality and potential bias of included studies will be assessed using the guidelines for assessing quality in prognostic studies, assigning an overall quality score per study of between 0 and 6 points according to Hayden *et al.*[24] The six topics assessed are: study participation, study attrition, prognostic factor measurement, outcome measurement, confounding measurement and account, and analysis. Of particular relevance for this study will be the assessment of radiographic damage. A list of ten possible biases in the assessment of radiographic data was described by van der Heijde[25] and this information will be collected separately for quality assessment of the radiographic outcome, but will not be used as the basis for including/excluding studies.

SPECIFICATION OF OUTCOMES AND EFFECT MEASURES

For the purposes of this study the methodology adopted by the ACR/EULAR group[3] to define good outcome in radiographic damage and function (separately and combined) will be adopted. Using the same definitions will allow direct comparisons of the results/conclusion. However, in the present study different definitions of "good radiographic outcome" will be evaluated (for more details please see "Exploration in variation effects (sensitivity analyses)" section).

Primary outcome:

a) Percentage of individuals with a good radiographic outcome during the second year of the trial (i.e. between month 12 and month 24).

"Good radiographic outcome" is defined as stable radiographic scores (change ≤0 in Sharp[26] or modified Sharp scores[27-29] during the second year of the trial).

Secondary outcomes:

b) Percentage of individuals with a good functional outcome during the second year of the trial.

"Good physical function outcome" is defined as stable and low scores of Health Assessment Questionnaire (HAQ)[30] (change ≤0 and HAQ score consistently ≤0.5 during the second year of the trial).

c) Percentage of individuals with overall good outcome during the second year of the trial.

"Overall good outcome" is considered the combination of "good radiographic outcome" and "good physical function outcome".

Additional secondary outcomes:

The above-mentioned outcomes assessed the stability between 1 and 2 years. Additional secondary outcomes will assess stability between 1 and 5 years and between 1 and 10 years after baseline for a), b) and c), in trials with such long follow-up.

Measures of association:

The principal effect measure for all outcomes will be the positive (LR+) and negative likelihood ratios (LR-).

COMPARISONS: DEFINITIONS OF REMISSION

Analyses will be based on the definition of different remission states (Figure 1), assessed at 6 months and 12 months, following the methodology adopted by the ACR/EULAR committee[3], as follows:

a) ACR/EULAR Boolean-based remission[3], also designed in this project as "4v-Remission" (i.e., TJC28≤1, SJC28≤1, CRP mg/dl ≤1, and PGA≤1/10)
b) "4v-near-remission"[13, 14], defined as TJC28≤1, SJC28≤1, CRP mg/dl ≤1, and PGA>1.
c) "Non-remission" defined as TJC28>1 or SJC28>1 or CRP mg/dl >1.

The above three definitions are mutually exclusive, i.e. each patient will be categorized in one group only.

d) "3v-remission" defined as TJC28≤1, SJC28≤1, CRP mg/dl ≤1.

All remission definitions are considered to be satisfied when they are satisfied either at 6 or 12 months follow-up.

The LR for good outcome associated with 4v and 3v--remission states will be descriptively compared. Then,

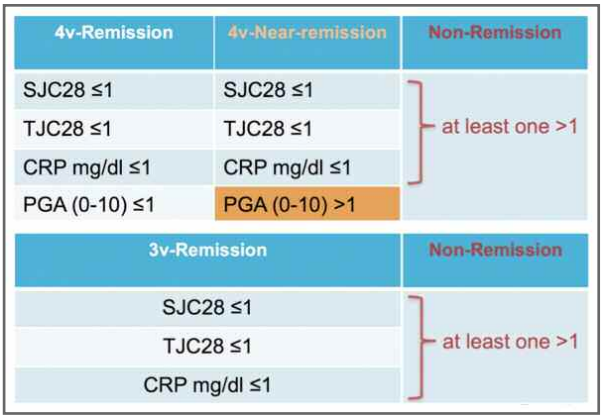

FIGURE 1. Definitions of Boolean remission considered for this study, adapted from the ACR/EULAR Boolean definition

CRP: C-reactive protein, PGA: patien global assessment; SJC28: swollen 28-joint counts; TJC28: tender 28-joint counts

an analysis will be performed for evaluating the (lack of) additional predictive value of PGA in the definition of remission, using the mutually exclusive categories ("4v-remission", "4v-near-remission" and "non-remission") (for more details please see "data analyses" section and Appendix 4).

DEFINITIONS OF SUSTAINED REMISSION

Sensitivity analyses will assess the influence of sustained remission, i.e. remission in more than a single time-point (6 or 12 months), in the prediction of good outcomes in RA. Because there is no currently uniform definition of sustained remission, the following will be tested: (i) remission at 6 and 12 months, (ii) remission at two consecutive visits among all time-points consistently available during the first year in all trials, ideally separated by 3 months (i.e. 3, 6, 9, and 12 months), as suggested by Konijn *et al.*[31], (iii) having or not ≥50% of visits in remission, and (iv) using the "Continuity Reward" (ConRew) score proposed by Boers *et al.*[32], which gives 1 point for a period in remission and a bonus (1 point more) if the subsequent period is also remission or if it is the last observation period. Because ConRew is a continuous score and the distributions are expected to be strongly right skewed[32], the 25th percentile will be considered as cut-off to dichotomize sustained versus non-sustained remission. Definitions (iii) and (iv) will consider the visits from the beginning to the end of the follow-up under consideration, i.e. not only the first year of the trial and patients will be classified to the 'highest category' for definitions ii an iv, e.g. when they satisfy both the 4v-remission and 4v--near-remission definitions at two consecutive visits, or have both a ConRew score >25th percentile based on 4v-remission and 4v-near remission, they are assigned to the 4v remission category.

DATA ANALYSIS AND SYNTHESIS

Data analysis

To guarantee privacy and security of IPD, the platforms in which the data is available require that statistics be performed via remote and secure online platforms, which impedes data download. However, all platforms commonly allow the use of SAS software, which will be used within each platform. For data synthesis Stata software, version 14 will be used. Thus the same procedures will need to be performed in each platform.

Means and standard deviations (SD) will be used to describe normally distributed continuous data, medians and interquartile ranges to describe continuous data that are not normally distributed, and frequencies and percentages will be used for categorical data. Data will be described using 95% confidence intervals (95% CI).

The number of true positive (TP), true negative (TN), false negative (FN) and false positive (FP) statistics will be extracted for each dataset. Then, sensitivity, specificity, positive and negative predictive value (PPV and NPV) will be determined for being in (sustained) remission (by 3v and 4v-remission) to predict good radiographic outcome at 2-years after baseline. The LR+ and LR- will then be calculated by the formulas: *sensitivity/(1-specificity)* and *(1-sensitivity)/specificity*, respectively. All analyses will be repeated for secondary outcomes: good function, and overall good outcome.

LRs, calculated as above, will be descriptively compared. In a second phase, participants in non-remission will be excluded from analyses and LRs for good radiographic outcome will be calculated for "4v-remission" versus "4v-near-remission", as a means to assess the predictive impact of PGA (Appendix 4, image C).

To additionally test the predictive value of PGA in the definition of remission, logistic regression will be used with the outcome of interest - radiographic stability - as dependent variable. The independent variables will include remission (a categorical variable indicating whether a patient is in "4v-remission", "4v-near-remission", or not satisfying any of these definitions, i.e. "non-remission") (Appendix 4, images D and E). The regression coefficient/odds ratio (OR) (and 95% CI) comparing the "4v-near-remission" category

with the "4v-remission" category will indicate whether there is any relevant difference between these groups and thus whether PGA has any additional predictive value for outcome.

Missing data will not be submitted to any method of data imputation, based on the comparable results of imputed and non-imputed data shown by a similar analysis[31]. The number and percentage of patients with missing values for each variable will be reported per trial.

Measures to adjust for confounders

In order to adjust for important covariates (gender, age at baseline, disease duration at baseline, RF, ACPA, radiographic damage at baseline, treatment arm), logistic regression (as above) will be used in individual studies.

Data synthesis

A two-step approach will be followed in this IPD meta-analysis. Thus, the TP, TN, FN, and FP results obtained for each trial in a first step (described above) will be used to synthetize the data in a second step. To consider the results from the mutually exclusive definitions of remission and to take into account the influence of the covariates, the OR and its standard error (SE) resulting from the logistic regression will also be synthesized, using appropriate fixed-effect and random-effects approaches, as suggested by Chang and Hoaglin[33].

As the definitions of remission will be the same over all studies as well as the definition of the outcome (as they are calculated by the authors), a Bivariate hierarchical model with random effect will be used to summarize the diagnostic association measure[34].

The I^2 of Higgins and Thompson will be calculated to quantify heterogeneity[35,36].

EXPLORATION IN VARIATION EFFECTS (SENSITIVITY ANALYSES)

In recent years a statistically significant reduction in radiographic progression during clinical studies in patients with RA has become difficult to detect due to early-escape study designs and to declining rates of progression in control-group patients[37]. For this reason, and also to be consistent with the methodology used by ACR and EULAR to establish the current definition of remission, a strict definition of good outcome was adopted for this study, i.e, a change ≤0 in radiographic progression. However, the majority of recent studies consider a cut-off ≤0.5 to define radiographic stability, to allow for a maximum change of 1 unit by one of the two readers. Therefore, a sensitivity analysis will consider change ≤0.5 units as cut-off for radiographic damage progression. Also, in order to account for inter and intra-rater variation of the radiographic score[31,38,39] and to provide information on the magnitude of structural damage, two additional cut-offs to define good radiographic outcome will be considered: ≤3 and ≤5 units of the radiographic score. It is also argued that the number of years of follow-up could affect the results of radiographic outcomes[19]. Thus, in addition to the main analysis of the 2-year outcome (the most frequently reported), outcomes at 5 and 10 years after baseline will also be assessed, i.e, 4 and 9 years' stability after 12 months.

The definition of "sustained" remission (and non-remission) based on only 1 time point or even 2 consecutive time points may not fully capture all relevant information[32]. Thus we will explore whether multiple remission and relapse periods are related to long-term radiographic progression and compare the performance of 3v and 4v-remission definitions, namely using the ConRew score[32], as explained above.

We will, finally, evaluate whether the relationship between the definition of remission and outcome is affected by treatment (mono versus combined), disease duration (early versus established), and history of previous DMARD treatments (naive versus failure/non-responders).

DISCUSSION

This study will evaluate whether the predictive value of 3v and 4v-remission states regarding the development of structural damage, are comparable. If confirmed, the inclusion of PGA in the definition of remission as treatment target should be revised (to 3v-remission), thus helping to avoid unnecessary immunosuppressive treatment escalation and associated risks. Given that more patients attain 3v than 4v-remission, it is expected that the value (long term benefit) of available therapies will be reassessed and probably recognized as higher than previously acknowledged.

If no relevant difference is observed by including or excluding PGA from the definition, the proposal towards the adoption of two separate targets for the treatment of RA of therapy (control of the disease process and control of its impact upon the patient's life) will be strongly supported. Disease impact will continue to be core to the assessment and management of the disease but it will be better served by instruments that allow

the health professionals to understand the reasons driving a high-perceived impact of the disease. Once disease control is achieved, adjunctive pharmacological and non-pharmacological therapies of different nature may be considered, based on the understanding of disease impact in the individual patient[12,13].

In case the results demonstrate that including PGA in the ACR/EULAR Boolean-based definition increases its predictive value of good outcomes, the current definition of remission is supported. This would not invalidate the need to consider a separate patient impact target[40,41]. We would argue, in any case, following available evidence, that the formulation of PGA should be standardized[42,43] and that patients should have a dedicated debriefing about this measure in order to improve its reliability[44].

An IPD meta-analysis is adequate for this study because it allows calculating a new definition of remission in RA, the 3v-remission that includes the same variables used to assess remission by current definitions. Being "new", this definition has never been published, thus not accessible through a conventional meta-analysis.

This type of meta-analysis has many advantages but also some limitations.[45-47] The main advantages of this specific study include access: (i) to large datasets of patients, (ii) with long-term outcome assessment, and (iii) rigorous data collection. The potential limitations are mainly related with the different designs of the RCTs included, namely: different inclusion criteria (e.g. patients naive versus non-responders to MTX), different treatments (e.g. patient in mono versus combined DMARD therapy), different time point assessments within the same period, or variation in radiographic scoring. To face these limitations different sub-group and sensitivity analyses are planned, that will allow guaranteeing the rigor and generalizability of the results. The inclusion of highly experienced experts in radiographic outcome assessment and in RCT design, development, and statistical analyses in this international consortium of researchers will also contribute to overcome possible difficulties. All the authors will be engaged in close critical appraisal in every step of the analyses to ensure the validity of the results and conclusions of this study. To guarantee quality control and reproducibility of the procedures the analysis's syntax will be recorded.

Two important decisions that were taken during study design are important to highlight. The first decision was regarding the number of studies to include. Usually, turning large amounts of data into actionable information allows better contributions in epidemiology[48]. The authors agreed that there was no need to include all existing RCTs testing biological and targeted synthetic agents, but including data from different data-holders would strengthen this study. The second decision was related with the fact that this study is questioning the current ACR/EULAR Boolean-based definition of remission and how strictly should this study reproduce their methodological decisions. It was decided to reproduce their analysis but performing further sensitivity analysis using other methodological options. An example was the additional cut-off for the definition of good radiographic outcomes. Another example was the definition of sustained remission: although ACR/EULAR committee[3] have used a single point in time (6 or 12 months) to define remission and despite the nonexistence of a uniform definition of sustained remission[31], this study will also use four additional definitions.

The present study considers radiographic score as primary endpoints and function (HAQ) and overall outcomes as secondary endpoints. This decision was based on the fact that HAQ: (i) is not only an outcome measure (cumulative functional deterioration over time) but also a disease activity-related measure (impact of current disease activity on function a specific point in time)[49], and (ii) is subjective[50,51], i.e. does not measure functioning of patients, but assesses their opinion on their functioning. So, the factors underlying an unjustifiably high PGA would be expected to have a similar effect upon HAQ, confounding the argument.

At the time of submission of this manuscript, all the five pharmaceutical companies (MSD/JANSSEN, Pfizer, Abbvie, Roche and UCB) contacted had already granted us access to their RCTs, which demonstrates the perceived value of this research project. These positive answers also assure that structural damage and functional outcomes after 5 and 10-years of DMARD initiation will also be possible to be compared.

ACKNOWLEDGMENTS

We wish to thank Cátia Duarte and Eduardo Santos (Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal) for critically revising the manuscripts for its intellectual content.

CORRESPONDENCE TO

José António Pereira da Silva
Centro Hospitalar Universitário de Coimbra, EPE.
Serviço de Reumatologia, Piso 7.
Praceta Professor Mota Pinto
3000-075 Coimbra. Portugal
E-mail: jdasilva@ci.uc.pt

REFERENCES

1. Smolen JS, Breedveld FC, Burmester GR, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. Ann Rheum Dis 2015; 75: 3-15.
2. Smolen JS, Landewe R, Bijlsma J, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. Ann Rheum Dis 2017; 76: 960-977.
3. Felson DT, Smolen JS, Wells G, et al. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum 2011; 63: 573-586.
4. Vermeer M, Kuper HH, van der Bijl AE, et al. The provisional ACR/EULAR definition of remission in RA: a comment on the patient global assessment criterion. Rheumatology (Oxford) 2012; 51: 1076-1080.
5. Khan NA, Spencer HJ, Abda E, et al. Determinants of discordance in patients' and physicians' rating of rheumatoid arthritis disease activity. Arthritis Care Res (Hoboken) 2012; 64: 206--214.
6. Ward MM, Guthrie LC, Alba MI. Measures of arthritis activity associated with patient-reported improvement in rheumatoid arthritis when assessed prospectively versus retrospectively. Arthritis Care Res (Hoboken) 2015; 67: 776-781.
7. Egsmose EL, Madsen OR. Interplay between patient global assessment, pain, and fatigue and influence of other clinical disease activity measures in patients with active rheumatoid arthritis. Clin Rheumatol 2015; 34: 1187-1194.
8. Sokka T, Kautiainen H, Pincus T, et al. Disparities in rheumatoid arthritis disease activity according to gross domestic product in 25 countries in the QUEST-RA database. Ann Rheum Dis 2009; 68: 1666-1672.
9. Masri KR, Shaver TS, Shahouri SH, et al. Validity and reliability problems with patient global as a component of the ACR/EULAR remission criteria as used in clinical practice. J Rheumatol 2012; 39: 1139-1145.
10. Studenic P, Radner H, Smolen JS, Aletaha D. Discrepancies between patients and physicians in their perceptions of rheumatoid arthritis disease activity. Arthritis Rheum 2012; 64: 2814--2823.
11. Inanc N, Yilmaz-Oner S, Can M, Sokka T, Direskeneli H. The role of depression, anxiety, fatigue, and fibromyalgia on the evaluation of the remission status in patients with rheumatoid arthritis. J Rheumatol 2014; 41: 1755-1760.
12. Ferreira RJO, Duarte C, Ndosi M, de Wit M, Gossec L, da Silva JAP. Suppressing inflammation in rheumatoid arthritis: Does patient global assessment blur the target? A practice–based call for a paradigm change. Arthritis Care Res (Hoboken) Online First: 23 May 2017. doi:10.1002/acr.23284.
13. Ferreira RJO, Dougados M, Kirwan J, et al. Drivers of patient global assessment in patients with rheumatoid arthritis who are close to remission: an analysis of 1588 patients Rheumatology (Oxford) 2017; 56: 1573-1578.
14. Studenic P, Smolen JS, Aletaha D. Near misses of ACR/EULAR criteria for remission: effects of patient global assessment in Boolean and index-based definitions. Ann Rheum Dis 2012; 71: 1702-1705.
15. Balogh E, Dias JM, Orr C, et al. Comparison of remission criteria in a tumour necrosis factor inhibitor treated rheumatoid arthritis longitudinal cohort: patient global health is a confounder. Arthritis Res Ther 2013; 15: R221.
16. Smolen JS, Landewe R, Breedveld FC, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. Ann Rheum Dis 2014; 73: 492-509.
17. Landewe R, van de Heijde D. Is radiographic progression a realistic outcome measure in clinical trials with early inflammatory arthritis? Clin Exp Rheumatol 2003; 21: S37-41.
18. Navarro-Compan V, Gherghe AM, Smolen JS, Aletaha D, Landewe R, van der Heijde D. Relationship between disease activity indices and their individual components and radiographic progression in RA: a systematic literature review. Rheumatology (Oxford) 2015; 54: 994-1007.
19. Svensson B, Andersson ML, Bala SV, Forslind K, Hafstrom I. Long-term sustained remission in a cohort study of patients with rheumatoid arthritis: choice of remission criteria. BMJ open 2013; 3: e003554.
20. Ferreira RJO, Machado PM, Gossec L, et al. Long-term predictive value of including patient global assessment in Boolean remission regarding radiographic damage and physical function in patients with rheumatoid arthritis: protocol for an individual patient data meta-analysis PROSPERO2017 http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017057099. Accessed in October 25th 2017.
21. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988; 31: 315--324.
22. Aletaha D, Neogi T, Silman AJ, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2010; 69: 1580-1588.
23. Committee on Strategies for Responsible Sharing of Clinical Trial; Data Board on Health Sciences Policy; Institute of Medicine. Discussion Framework for Clinical Trial Data Sharing: Guiding Principles, Elements, and Activities. Washington (DC): National Academies Press (US); 2014.
24. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med 2006; 144: 427-437.
25. van der Heijde D. Structural damage in rheumatoid arthritis as visualized through radiographs. Arthritis Res 2002; 4 Suppl 2: S29-33.
26. Sharp JT, Lidsky MD, Collins LC, Moreland J. Methods of scoring the progression of radiologic changes in rheumatoid arthritis. Correlation of radiologic, clinical and laboratory abnormalities. Arthritis Rheum 1971; 14: 706-720.
27. Ory PA. Interpreting radiographic data in rheumatoid arthritis. Ann Rheum Dis 2003; 62: 597-604.
28. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000; 27: 261-263.
29. van der Heijde DM, van Riel PL, Nuver-Zwart IH, Gribnau FW, vad de Putte LB. Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. Lancet 1989; 1: 1036-1038.
30. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980; 23: 137-145.
31. Konijn NPC, van Tuyl LHD, Boers M, et al. Do Short and Sustained Periods of American College of Rheumatology/European League Against Rheumatism Remission Predict Functional and Radiographic Outcome in Early Rheumatoid Arthritis Patients With Low Overall Damage Progression? Arthritis Care Res

(Hoboken) 2017; 69: 989-996.
32. Boers M, Berkhof J, Twisk JW, et al. A new graph and scoring system simplified analysis of changing states: disease remissions in a rheumatoid arthritis clinical trial. J Clin Epidemiol 2010; 63: 633-637.
33. Chang BH, Hoaglin DC. Meta-Analysis of Odds Ratios: Current Good Practices. Med Care 2017; 55: 328-335.
34. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10 Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 10: The Cochrane Collaboration, 2010. http://srdta.cochrane.org/ Accessed in October 25th 2017.
35. Moayyedi P. Meta-analysis: Can we mix apples and oranges? Am J Gastroenterol 2004; 99: 2297-2301.
36. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002; 21: 1539-1558.
37. Landewe RB, Connell CA, Bradley JD, et al. Is radiographic progression in modern rheumatoid arthritis trials still a robust outcome? Experience from tofacitinib clinical trials. Arthritis Res Ther 2016; 18: 212.
38. van der Heijde D, Tanaka Y, Fleischmann R, et al. Tofacitinib (CP-690,550) in patients with rheumatoid arthritis receiving methotrexate: twelve-month data from a twenty-four-month phase III randomized radiographic study. Arthritis Rheum 2013; 65: 559-570.
39. Fleischmann RM, Huizinga TW, Kavanaugh AF, et al. Efficacy of tofacitinib monotherapy in methotrexate-naive patients with early or established rheumatoid arthritis. RMD open 2016; 2: e000262.
40. van Tuyl LHD, Boers M. Rheumatoid arthritis: Remission - keeping the patient experience front and centre. Nat Rev Rheumatol 2017; 13: 573-574.
41. Ferreira RJO, Duarte C, Ndosi M, de Wit M, Gossec L, da Silva JAP. Controversy: The role of patient global assessment (PGA) in the definition of remission in rheumatoid arthritis. Nature Reviews Rheumatology (in press).
42. Nikiphorou E, Radner H, Chatzidionysiou K, et al. Patient global assessment in measuring disease activity in rheumatoid arthritis: a review of the literature. Arthritis Res Ther 2016; 18: 251.
43. Eugénio G, Ferreira R, Silva C, Medeiros C, Silva J, Duarte C. THU0145 Impact of different formulations of "patient global assessment" on remission classification by disease activity indices in rheumatoid arthritis [abstract]. Ann Rheum Dis 2017; 76: 255-256.
44. Henriques M, Duarte C, Ndosi M, Marques A, Silva Jd, Ferreira R. "It can't be zero": a qualitative study of patients' perspective on patient global assessment in rheumatoid arthritis [abstract]. Ann Rheum Dis 2017; 76(Suppl 2): 112-112.
45. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ 2010; 340: c221.
46. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002; 25: 76-97.
47. Stewart LA, Tierney JF. Reviews of individual patient data. 2011. In: Cochrane Handbook for Systematic Reviews of Interventions. Chichester: John Wiley & Sons Ltd.
48. European Commission. Directorate-General for Health and Consumers. Unit D3 eHealth and Health Technology Assessment. The Use of Big Data in Public Health Policy and Research: Background information document. Brussels, 2014.
49. Welsing PM, van Gestel AM, Swinkels HL, Kiemeney LA, van Riel PL. The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. Arthritis Rheum 2001; 44: 2009-2017.
50. Wolfe F. A reappraisal of HAQ disability in rheumatoid arthritis. Arthritis Rheum 2000; 43: 2751-2761.
51. van den Ende CH, Hazes JM, Le Cessie S, Breedveld FC, Dijkmans BA. Discordance between objective and subjective assessment of functional ability of patients with rheumatoid arthritis. Br J Rheumatol 1995; 34: 951-955.

PRISMA-IPD CHECKLIST OF ITEMS TO INCLUDE WHEN REPORTING A SYSTEMATIC REVIEW AND META-ANALYSIS OF INDIVIDUAL PARTICIPANT DATA (IPD)

PRISMA-IPD Section/topic	Item No	Checklist item	Reported on page
Title			
Title	1	Identify the report as a systematic review and meta-analysis of individual participant data.	
Abstract			
Structured summary	2	Provide a structured summary including as applicable:	
		Background: state research question and main objectives, with information on participants, interventions, comparators and outcomes.	
		Methods: report eligibility criteria; data sources including dates of last bibliographic search or elicitation, noting that IPD were sought; methods of assessing risk of bias.	
		Results: provide number and type of studies and participants identified and number (%) obtained; summary effect estimates for main outcomes (benefits and harms) with confidence intervals and measures of statistical heterogeneity. Describe the direction and size of summary effects in terms meaningful to those who would put findings into practice.	
		Discussion: state main strengths and limitations of the evidence, general interpretation of the results and any important implications.	
		Other: report primary funding source, registration number and registry name for the systematic review and IPD meta-analysis.	
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of the questions being addressed with reference, as applicable, to participants, interventions, comparisons, outcomes and study design (PICOS). Include any hypotheses that relate to particular types of participant-level subgroups.	
Methods			
Protocol and registration	5	Indicate if a protocol exists and where it can be accessed. If available, provide registration information including registration number and registry name. Provide publication details, if applicable.	
Eligibility criteria	6	Specify inclusion and exclusion criteria including those relating to participants, interventions, comparisons, outcomes, study design and characteristics (e.g. years when conducted, required minimum follow-up). Note whether these were applied at the study or individual level i.e. whether eligible participants were included (and ineligible participants excluded) from a study that included a wider population than specified by the review inclusion criteria. The rationale for criteria should be stated.	
Identifying studies – information sources	7	Describe all methods of identifying published and unpublished studies including, as applicable: which bibliographic databases were searched with dates of coverage; details of any hand searching including of conference proceedings; use of study registers and agency or company databases; contact with the original research team and experts in the field; open adverts and surveys. Give the date of last search or elicitation.	

continues on the next page

CONTINUATION

PRISMA-IPD Section/topic	Item No	Checklist item	Reported on page
Identifying studies – search	8	Present the full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
Study selection processes	9	State the process for determining which studies were eligible for inclusion.	
Data collection processes	10	Describe how IPD were requested, collected and managed, including any processes for querying and confirming data with investigators. If IPD were not sought from any eligible study, the reason for this should be stated (for each such study).	
		If applicable, describe how any studies for which IPD were not available were dealt with. This should include whether, how and what aggregate data were sought or extracted from study reports and publications (such as extracting data independently in duplicate) and any processes for obtaining and confirming these data with investigators.	
Data items	11	Describe how the information and variables to be collected were chosen. List and define all study level and participant level data that were sought, including baseline and follow-up information. If applicable, describe methods of standardising or translating variables within the IPD datasets to ensure common scales or measurements across studies.	
IPD integrity	A1	Describe what aspects of IPD were subject to data checking (such as sequence generation, data consistency and completeness, baseline imbalance) and how this was done.	
Risk of bias assessment in individual studies.	12	Describe methods used to assess risk of bias in the individual studies and whether this was applied separately for each outcome. If applicable, describe how findings of IPD checking were used to inform the assessment. Report if and how risk of bias assessment was used in any data synthesis.	
Specification of outcomes and effect measures	13	State all treatment comparisons of interests. State all outcomes addressed and define them in detail. State whether they were pre-specified for the review and, if applicable, whether they were primary/main or secondary/additional outcomes. Give the principal measures of effect (such as risk ratio, hazard ratio, difference in means) used for each outcome.	
Synthesis methods	14	Describe the meta-analysis methods used to synthesise IPD. Specify any statistical methods and models used. Issues should include (but are not restricted to):	
		• Use of a one-stage or two-stage approach.	
		• How effect estimates were generated separately within each study and combined across studies (where applicable).	
		• Specification of one-stage models (where applicable) including how clustering of patients within studies was accounted for.	
		• Use of fixed or random effects models and any other model assumptions, such as proportional hazards.	
		• How (summary) survival curves were generated (where applicable).	
		• Methods for quantifying statistical heterogeneity (such as I2 and $\tau 2$).	
		• How studies providing IPD and not providing IPD were analysed together (where applicable).	
		• How missing data within the IPD were dealt with (where applicable).	

continues on the next page

CONTINUATION

PRISMA-IPD Section/topic	Item No	Checklist item	Reported on page
Exploration of variation in effects	A2	If applicable, describe any methods used to explore variation in effects by study or participant level characteristics (such as estimation of interactions between effect and covariates). State all participant-level characteristics that were analysed as potential effect modifiers, and whether these were pre-specified.	
Risk of bias across studies	15	Specify any assessment of risk of bias relating to the accumulated body of evidence, including any pertaining to not obtaining IPD for particular studies, outcomes or other variables.	
Additional analyses	16	Describe methods of any additional analyses, including sensitivity analyses. State which of these were pre-specified.	
Results			
Study selection and IPD obtained	17	Give numbers of studies screened, assessed for eligibility, and included in the systematic review with reasons for exclusions at each stage. Indicate the number of studies and participants for which IPD were sought and for which IPD were obtained. For those studies where IPD were not available, give the numbers of studies and participants for which aggregate data were available. Report reasons for non-availability of IPD. Include a flow diagram.	
Study characteristics	18	For each study, present information on key study and participant characteristics (such as description of interventions, numbers of participants, demographic data, unavailability of outcomes, funding source, and if applicable duration of follow-up). Provide (main) citations for each study. Where applicable, also report similar study characteristics for any studies not providing IPD.	
IPD integrity	A3	Report any important issues identified in checking IPD or state that there were none.	
Risk of bias within studies	19	Present data on risk of bias assessments. If applicable, describe whether data checking led to the up-weighting or down-weighting of these assessments. Consider how any potential bias impacts on the robustness of meta-analysis conclusions.	
Results of individual studies	20	For each comparison and for each main outcome (benefit or harm), for each individual study report the number of eligible participants for which data were obtained and show simple summary data for each intervention group (including, where applicable, the number of events), effect estimates and confidence intervals. These may be tabulated or included on a forest plot.	
Results of syntheses	21	Present summary effects for each meta-analysis undertaken, including confidence intervals and measures of statistical heterogeneity. State whether the analysis was pre-specified, and report the numbers of studies and participants and, where applicable, the number of events on which it is based.	
		When exploring variation in effects due to patient or study characteristics, present summary interaction estimates for each characteristic examined, including confidence intervals and measures of statistical heterogeneity. State whether the analysis was pre-specified. State whether any interaction is consistent across trials.	
		Provide a description of the direction and size of effect in terms meaningful to those who would put findings into practice.	
Risk of bias across studies	22	Present results of any assessment of risk of bias relating to the accumulated body of evidence, including any pertaining to the availability and representativeness of available studies, outcomes or other variables.	

continues on the next page

CONTINUATION

PRISMA-IPD Section/topic	Item No	Checklist item	Reported on page
Additional analyses	23	Give results of any additional analyses (e.g. sensitivity analyses). If applicable, this should also include any analyses that incorporate aggregate data for studies that do not have IPD. If applicable, summarise the main meta-analysis results following the inclusion or exclusion of studies for which IPD were not available.	
Discussion			
Summary of evidence	24	Summarise the main findings, including the strength of evidence for each main outcome.	
Strengths and limitations	25	Discuss any important strengths and limitations of the evidence including the benefits of access to IPD and any limitations arising from IPD that were not available.	
Conclusions	26	Provide a general interpretation of the findings in the context of other evidence.	
Implications	A4	Consider relevance to key groups (such as policy makers, service providers and service users). Consider implications for future research.	
Funding			
Funding	27	Describe sources of funding and other support (such as supply of IPD), and the role in the systematic review of those providing such support.	

A1 – A3 denote new items that are additional to standard PRISMA items. A4 has been created as a result of re-arranging content of the standard PRISMA statement to suit the way that systematic review IPD meta-analyses are reported.

APPENDIX 2. SUMMARY TABLE OF ELIGIBLE STUDIES FOR THIS IPD META-ANALYSIS ALREADY IDENTIFIED

Trial name – Trial registry	Drug – Company	Sample	Follow-Up Duration	Damage	Function	Population
PREMIER - NCT00195663	Adalimumab - Abbvie	799/ 697/ 452	10y	mTSS	HAQ	Naïve MTX
DE019 - NCT00195702	Adalimumab - Abbvie	619/ 202/ 327	10y	mTSS	HAQ	Non-responders MTX
TEAR - NCT00259610	Etanercept - Pfizer	755/476	2y	mTSS	mHAQ	Non-responders MTX
COMET - NCT00195494	Etanercept - Pfizer	542/398	2y	mTSS	HAQ	Naïve MTX
CAMEO - NCT00654368	Etanercept - Pfizer	205	2Y	mTSS	HAQ	Non-responders MTX
PRIZE - NCT00913458	Etanercept - Pfizer	193/131	1,5/2y	mTSS	HAQ	Naïve MTX
TEMPO - NCT00393471	Etanercept - Pfizer	682	3y	mTSS	HAQ	Failed DMARD (not MTX)
PRESERVE - NCT00565409	Etanercept - Pfizer	600	2y (88w)	mTSS	HAQ	Non-responders MTX
ERA - NCT00356590	Etanercept - Pfizer	632/512 ~300	2y 2 + 3y	Sharp	HAQ	Naïve MTX
GO BEFORE - NCT00264537	Golimumab - MSD	637/422	2+3Y	mTSS	HAQ	Naïve MTX
GO FORWARD - NCT00264550	Golimumab - MSD	444/313	5Y	mTSS	HAQ	Non-responders MTX
GO FURTHER - NCT00973479	Golimumab - MSD	592/486	2Y (100w)	mTSS	HAQ	Non-responders MTX
ATTRACT - NCT00269867	Infliximab - MSD	428/340	2y (102w)	mTSS	HAQ	Non-responders MTX
ASPIRE - NCT00236028	Infliximab - MSD	1049	1y	mTSS	HAQ	Non-responders MTX
LITHE - NCT00106535	Tocilizumab - Roche	1190/ 1149	5y	GmTSS	HAQ	Non-responders MTX
FUNCTION - NCT01007435	Tocilizumab - Roche	1157	2y	mTSS	HAQ	Non-responders MTX
BREVACTA - NCT01232569	Tocilizumab - Roche	656/ 314	2Y	mTSS	HAQ	Non-responders MTX
ACT-RAY - NCT00810199	Tocilizumab - Roche	512/423	2Y	GmTSS	HAQ	Non-responders MTX
SAMURAI - NCT00144508	Tocilizumab - Roche	306/241	1y 3y	mTSS	mHAQ	Non-responders MTX
SURPRISE - NCT01120366	Tocilizumab - Roche	226	2Y	mTSS	HAQ	Non-responders MTX
REFLEX - NCT00468546/ NCT02097745	Rituximab - Roche	517 184	2Y 5Y	GmTSS	HAQ	Inadequate response to anti-TNF
IMAGE - NCT00299104	Rituximab - Roche	776/606	2Y	GmTSS	HAQ	Naïve MTX
RAPID 1 - NCT00152386	Certolizumab - UCB	982/847	3y (148w)	mTSS	HAQ	Non-responders MTX
RAPID 2 - NCT00160602/ NCT00175877	Certolizumab - UCB	619/567	2,5y (128w)	mTSS	HAQ	Non-responders MTX
C-OPERA - NCT01451203	Certolizumab - UCB	316/184 131	1y 2y	mTSS	HAQ	Naïve MTX
ORAL START - NCT01039688	Tofacitinib - Pfizer	958/956	2y	mTSS	HAQ	Naïve MTX
ORAL SCAN - NCT00847613	Tofacitinib - Pfizer	797	2y	mTSS	HAQ	Non-responders MTX

DMARD - disease-modifying antirheumatic drugs; GmTSS - Genant-modified Total Sharp Score; HAQ - Health Assessment Questionnaire; mHAQ – modified Health Assessment questionnaire; mTSS - modified Total Sharp Score; MTX - Metotrexate; TNF - tumor necrosis factor

APPENDIX 3. LIST OF VARIABLES REQUIRED AND DESIRABLE FOR ANALYSES

- Remission definition data (minimum at baseline, 6 and 12 months; desired also at 3 and 9 months, 2 years, others)
  - SJC28
  - TJC28
  - CRP (in mg/dl or mg/L, but clearly indicated)
  - PGA
- Outcome data (minimum at baseline, 12 and 24 months; desired 5 and 10 years)
  - Radiographic score
  - HAQ
- Patient characteristics (all at baseline only)
  - gender
  - age at baseline
- Clinical characteristics (all at baseline only)
  - disease duration at baseline
  - RF
  - ACPA (not essential)
  - Treatment arm
- Trial/visit information
  - anonymised patient ID code (at baseline only)
  - visit number or sequence
  - visit date

APPENDIX 4. EXAMPLES OF PATIENT'S CLASSIFICATION IN DIFFERENT REMISSION STATE DEFINITIONS (DICHOTOMIC AND CATEGORICAL) AND CONSIDERING SINGLE OR MULTIPLE TIME POINTS (SUSTAINED REMISSION) FOR ITS ASSESSMENT

A. Example of sub-group determination to compare Likelihood Ratios between 3v and 4v-remission, considering 1 point in time during the first year (6 or 12 months).

	6 months					12-months						
Patient	TJC ≤1	SJC ≤1	CRP ≤1	PGA ≤1	Remission state	TJC ≤1	SJC ≤1	CRP ≤1	PGA ≤1	Remission state	4v-rem at 6 or 12 months?	3v-rem at 6 or 12 months?
1	✔	✔	✔	✗	3v-Rem	✔	✔	✔	✗	3v-Rem	No	Yes
2	✔	✔	✔	✗	3v-Rem	✔	✔	✔	✔	4v & 3v-Rem	Yes	Yes
3	✔	✔	✔	✔	4v & 3v-Rem	✔	✔	✔	✔	4v & 3v-Rem	Yes	Yes
4	✔	✔	✔	✔	4v & 3v-Rem	✔	✔	✔	✗	3v-Rem	Yes	Yes
5	✔	✔	✔	✔	4v & 3v-Rem	✔	✔	✔	✔	4v & 3v-Rem	Yes	Yes
6	✔	✔	✔	✗	3v-Rem	✗	✔	✔	✗	Non-Rem	No	Yes
7	✔	✗	✔	✔	Non-Rem	✔	✔	✔	✔	4v & 3v-Rem	Yes	Yes
8	✗	✗	✔	✗	Non-Rem	✔	✔	✔	✗	3v-Rem	No	Yes
9	✗	✗	✔	✗	Non-Rem	✔	✔	✔	✔	4v-Rem	Yes	Yes
10	✔	✔	✗	✗	Non-Rem	✗	✗	✗	✗	Non-Rem	No	No
11	✗	✗	✔	✔	Non-Rem	✔	✔	✗	✔	Non-Rem	No	No
12	✔	✗	✔	✗	Non-Rem	✗	✔	✔	✗	Non-Rem	No	No
13	✔	✔	✗	✗	Non-Rem	✗	✔	✔	✗	Non-Rem	No	No

Likelihood Ratios will be obtained from the following comparisons for this example:
a) Patients in 4v-Rem vs patient in Non 4v-Rem = 6 vs 7
b) Patients in 3v-Rem vs patient in Non 3v-Rem = 9 vs 4

B. Example of sub-group determination to compare Likelihood Ratios between 3v and 4v-remission, considering sustained remission during the first year (…, 6, …, and 12 months*).

Patient	6 months					12-months					4v-rem sustained rem?	3v-rem sustained rem?
	TJC ≤1	SJC ≤1	CRP ≤1	PGA ≤1	Remission state	TJC ≤1	SJC ≤1	CRP ≤1	PGA ≤1	Remission state		
1	✔	✔	✔	✗	3v-Rem	✔	✔	✔	✗	3v-Rem	No	Yes
2	✔	✔	✔	✗	3v-Rem	✔	✔	✔	✔	4v & 3v-Rem	No	Yes
3	✔	✔	✔	✔	4v & 3v-Rem	✔	✔	✔	✔	4v & 3v-Rem	Yes	Yes
4	✔	✔	✔	✔	4v & 3v-Rem	✔	✔	✔	✗	3v-Rem	No	Yes
5	✔	✔	✔	✔	4v & 3v-Rem	✔	✔	✔	✔	4v & 3v-Rem	Yes	Yes
6	✔	✔	✔	✗	3v-Rem	✗	✔	✔	✗	Non-Rem	No	No
7	✔	✗	✔	✔	Non-Rem	✔	✔	✔	✔	4v & 3v-Rem	No	No
8	✗	✗	✔	✗	Non-Rem	✔	✔	✔	✗	3v-Rem	No	No
9	✗	✗	✔	✗	Non-Rem	✔	✔	✔	✔	4v-Rem	No	No
10	✔	✔	✗	✗	Non-Rem	✗	✗	✗	✗	Non-Rem	No	No
11	✗	✗	✔	✔	Non-Rem	✔	✔	✗	✔	Non-Rem	No	No
12	✔	✗	✔	✗	Non-Rem	✗	✔	✔	✗	Non-Rem	No	No
13	✔	✔	✗	✗	Non-Rem	✗	✔	✔	✗	Non-Rem	No	No

Likelihood Ratios will be obtained from the following comparisons for this example :
a) Patients in sustained 4v-Rem vs patient in Non sustained 4v-Rem = 2 vs 11
b) Patients in sustained 3v-Rem vs patient in Non sustained 3v-Rem = 5 vs 7

* Only 2 time points are presented in this example (6 and 12 months), however, if they were more (e.g. 3, 6, 9 and 12 months) the same principle of decision would be applied: to be in 4v-rem a patient need to present the criteria in all considered visits, the same for Near-rem. Patients not classified in sustained 4v-rem neither in sustained near-rem would be classified in non sustained rem

C. Example of sub-group determination to compare Likelihood Ratios between 4v and 3v-non-4v-remission, considering 1 point in time during the first year (6 or 12 months).

Patient	6 months					12-months					4v-rem at 6 or 12 months?	3v-rem at 6 or 12 months?
	TJC ≤1	SJC ≤1	CRP ≤1	PGA ≤1	Remission state	TJC ≤1	SJC ≤1	CRP ≤1	PGA ≤1	Remission state		
1	✔	✔	✔	✗	3v-Rem	✔	✔	✔	✗	3v-Rem	No	Yes
2	✔	✔	✔	✗	3v-Rem	✔	✔	✔	✔	4v & 3v-Rem	Yes	Yes
3	✔	✔	✔	✔	4v & 3v-Rem	✔	✔	✔	✔	4v & 3v-Rem	Yes	Yes
4	✔	✔	✔	✔	4v & 3v-Rem	✔	✔	✔	✗	3v-Rem	Yes	Yes
5	✔	✔	✔	✔	4v & 3v-Rem	✔	✔	✔	✔	4v & 3v-Rem	Yes	Yes
6	✔	✔	✔	✗	3v-Rem	✗	✔	✔	✗	Non-Rem	No	Yes
7	✔	✗	✔	✔	Non-Rem	✔	✔	✔	✔	4v & 3v-Rem	Yes	Yes
8	✗	✗	✔	✗	Non-Rem	✔	✔	✔	✗	3v-Rem	No	Yes
9	✗	✗	✔	✗	Non-Rem	✔	✔	✔	✔	4v-Rem	Yes	Yes
10	✔	✔	✗	✗	Non-Rem	✗	✗	✗	✗	Non-Rem	No	No
11	✗	✗	✔	✔	Non-Rem	✔	✔	✗	✔	Non-Rem	No	No
12	✔	✗	✔	✗	Non-Rem	✗	✔	✔	✗	Non-Rem	No	No
13	✔	✔	✗	✗	Non-Rem	✗	✔	✔	✗	Non-Rem	No	No

All patient in non-3v rem are excluded

Likelihood Ratios will be obtained from the following comparisons for this example:
a) Patients in 4v-Rem vs patients in 3v Rem & Non 4v-Rem = 6 vs 3

D. Example of patient's classification in 4v-remission, near-remission and non-remission, considering 1 point in time during the first year (6 or 12 months).

Patient	6 months					12-months					"Best" remission state at 6 or 12 months?
	TJC ≤1	SJC ≤1	CRP ≤1	PGA ≤1	Remission state	TJC ≤1	SJC ≤1	CRP ≤1	PGA ≤1	Remission state	
1	✔	✔	✔	✗	Near-Rem	✔	✔	✔	✗	Near-Rem	Near-Rem
2	✔	✔	✔	✗	Near-Rem	✔	✔	✔	✔	4v-Rem	4v-Rem
3	✔	✔	✔	✔	4v-Rem	✔	✔	✔	✔	4v-Rem	4v-Rem
4	✔	✔	✔	✔	4v-Rem	✔	✔	✔	✗	Near-Rem	4v-Rem
5	✔	✔	✔	✔	4v-Rem	✔	✔	✔	✔	4v-Rem	4v-Rem
6	✔	✔	✔	✗	Near-Rem	✗	✔	✔	✗	Non-Rem	Near-Rem
7	✔	✗	✔	✔	Non-Rem	✔	✔	✔	✔	4v-Rem	4v-Rem
8	✗	✗	✔	✗	Non-Rem	✔	✔	✔	✗	Near-Rem	Near-Rem
9	✗	✗	✔	✗	Non-Rem	✔	✔	✔	✔	4v-Rem	4v-Rem
10	✔	✔	✗	✗	Non-Rem	✗	✗	✗	✗	Non-Rem	Non Rem
11	✗	✗	✔	✔	Non-Rem	✔	✔	✗	✔	Non-Rem	Non Rem
12	✔	✗	✔	✗	Non-Rem	✗	✔	✔	✗	Non-Rem	Non Rem
13	✔	✔	✗	✗	Non-Rem	✗	✔	✔	✗	Non-Rem	Non Rem

E. Example of patient's classification in 4v-remission, near-remission and non-remission, considering sustained remission during the first year (…, 6, ..., and 12 months*).

Patient	6 months					12-months					"Best" sustained remission state at … 6, …, and 12 months?
	TJC ≤1	SJC ≤1	CRP ≤1	PGA ≤1	Remission state	TJC ≤1	SJC ≤1	CRP ≤1	PGA ≤1	Remission state	
1	✔	✔	✔	✗	Near-Rem	✔	✔	✔	✗	Near-Rem	Sust. Near-Rem
2	✔	✔	✔	✗	Near-Rem	✔	✔	✔	✔	4v-Rem	Sust. Near-Rem
3	✔	✔	✔	✔	4v-Rem	✔	✔	✔	✔	4v-Rem	Sust. 4v-Rem
4	✔	✔	✔	✔	4v-Rem	✔	✔	✔	✗	Near-Rem	Sust. Near-Rem
5	✔	✔	✔	✔	4v-Rem	✔	✔	✔	✔	4v-Rem	Sust. 4v-Rem
6	✔	✔	✔	✗	Near-Rem	✗	✔	✔	✗	Non-Rem	Non Sust. Rem
7	✔	✗	✔	✔	Non-Rem	✔	✔	✔	✔	4v-Rem	Non Sust. Rem
8	✗	✗	✔	✗	Non-Rem	✔	✔	✔	✗	Near-Rem	Non Sust. Rem
9	✗	✗	✔	✗	Non-Rem	✔	✔	✔	✔	4v-Rem	Non Sust. Rem
10	✔	✔	✗	✗	Non-Rem	✗	✗	✗	✗	Non-Rem	Non Sust. Rem
11	✗	✗	✔	✔	Non-Rem	✔	✔	✗	✔	Non-Rem	Non Sust. Rem
12	✔	✗	✔	✗	Non-Rem	✗	✔	✔	✗	Non-Rem	Non Sust. Rem
13	✔	✔	✗	✗	Non-Rem	✗	✔	✔	✗	Non-Rem	Non Sust. Rem

* Only 2 time points are presented in this example (6 and 12 months), however, if they were more (e.g. 3, 6, 9 and 12 months) the same principle of decision would be applied: to be in 4v-rem a patient need to present the criteria in all considered visits, the same for Near-rem. Patients not classified in sustained 4v-rem neither in sustained near-rem would be classified in non sustained rem