Revisiting the Use of Remission Criteria for Rheumatoid Arthritis by Excluding Patient Global Assessment: An Individual Meta-Analysis of 5792 Patients

Ricardo J O Ferreira, Paco M J Welsing, Johannes W G Jacobs, Laure Gossec, Mwidimi Ndosi, Pedro M Machado, Désirée Heijde, Jose a P da Silva

Revisiting the use of remission criteria for rheumatoid arthritis by excluding patient global assessment: an individual meta-analysis of 5792 patients

Authors:

Ricardo J. O. Ferreira, PhD[1,2*], Paco M. J. Welsing, PhD[3*], Johannes W. G. Jacobs, PhD [4], Laure Gossec, PhD[5,6], Mwidimi Ndosi, PhD[7], Pedro M. Machado, PhD[8,9,10], Désirée van der Heijde, PhD[11], José A. P. da Silva, PhD[1, 12]

* Joint first authors

1 – Rheumatology department – Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal

2 – Health Sciences Research Unit: Nursing (UICISA: E), Nursing School of Coimbra, Coimbra, Portugal; rferreira@reumahuc.org

3 – Department of Rheumatology & Clinical Immunology – University Medical Center, Utrecht, The Netherlands; P.M.J.Welsing@umcutrecht.nl

4 – Department of Rheumatology & Clinical Immunology – University Medical Center, Utrecht, The Netherlands; J.W.G.Jacobs@umcutrecht.nl

5 – Sorbonne Université – Institut Pierre Louis d'Epidémiologie et de Santé Publique, INSERM, Paris France

6 – Rheumatology department – Pitié Salpêtrière hospital, AP-HP, Paris, France. laure.gossec@gmail.com

7 – Faculty of Health and Applied Sciences – University of the West of England, Bristol, UK; mwidimi.ndosi@uwe.ac.uk

8 – Centre for Rheumatology & MRC Centre for Neuromuscular Diseases – University College London, London, UK

9 – Department of Rheumatology, University College London Hospitals NHS Foundation Trust, London, UK

10 – Department of Rheumatology – Northwick Park Hospital, London North West University Healthcare NHS Trust, London, UK; p.machado@ucl.ac.uk

11 – Rheumatology department – Leiden University Medical Centre, Leiden, The Netherlands; mail@dvanderheijde.nl

12 – Clínica Universitária de Reumatologia, and i-CBR Coimbra Institute for Clinical and Biological Research, Faculty of Medicine - University of Coimbra, - Portugal; jdasilva@chuc.min-saude.pt

Corresponding author: José António P. da Silva, MD, PhD

Rheumatology Department, Centro Hospitalar Universitário de Coimbra, EPE.

Praceta Professor Mota Pinto, 3000-075 Coimbra. Portugal

Phone: 00351 239400547 Fax: 00351 239400587 e-mail: jdasilva@chuc.min-saude.pt

Word count: 5325

Abstract

Objectives: To determine the impact of excluding patient global assessment (PGA) from the ACR/EULAR Boolean remission criteria, upon prediction of radiographic and functional outcome of RA.

Methods: Meta-analyses using individual patient data from RCTs testing the efficacy of biological agents on radiographic and functional outcomes at ≥2 years. Remission states were defined by 4 variants of the ACR/EULAR Boolean definition: (i) tender and swollen 28-joint counts (TJC28/SJC28), C-reactive protein (CRP, mg/dl), and PGA (0-10=worst) all≤1 (4V-remission), (ii) the same, except PGA>1 (4V-near-remission), (iii) 3V-remission (i and ii combined; similar to 4V, but without PGA), and (iv) non-remission (TJC28>1 and/or SJC28>1 and/or CRP>1). The most stringent class achieved at 6 or 12 months was considered. Good radiographic (GRO) and functional outcome (GFO) were defined as no worsening (i.e. change in modified Total Sharp score ≤0.5 units and ≤0.0 HAQ-DI points, respectively, during the second year. The pooled probabilities of GRO and GFO for the different definitions of remission were estimated and compared.

Results: Individual patient data (n=5,792) from eleven trials were analysed. 4V-remission was achieved by 23% of patients and 4V-near-remission by 19%. The probability of GRO in the 4V-near-remission group was numerically, but non-significantly, lower than that in the 4V-remission (78 vs 81%) and significantly higher than that for non-remission (72%; difference=6%, 95%CI:2-10%). Applying 3V-remission could have prevented therapy escalation in 19% of all participants, at the cost of an additional 6.1%, 4.0%, and 0.7% of patients having ΔmTSS>0.0, >0.5, and >5 units over 2 years, respectively. The probability of GFO (assessed in 8 trials) in 4V-near-remission (67%, 95%CI:63-71%) was significantly lower than in 4V-remission (78%, 74-81%) and similar to non-remission (69%, 66-72%).

Conclusion: 4V-near-remission and 3V-remission have similar validity as the original 4V-remission definition in predicting GRO, despite expected worse prediction of GFO, while potentially reducing the risk of overtreatment. This supports further exploration of 3V-remission as the target for immunosuppressive therapy complemented by patient-oriented targets.

Keywords: Rheumatoid arthritis, outcomes research, patient global assessment, patient reported outcomes, disease activity, remission, near-remission, radiographic damage, individual patient data meta-analysis.

KEY MESSAGES

What is already known about this subject?

- Few previous studies compared the prediction of good structural and functional outcomes between patients who fulfilled all four criteria of the current ACR/EULAR Boolean-based definition of remission ("4V-remission") versus those who attained only three ("3V-remission"), i.e. excluding patient global assessment (PGA). No significant differences were found but the two groups of patients evaluated significantly overlap.

What does this study add?

- This was the first study comparing these outcomes between patients achieving 4V-remission (23%) and those missing this status due solely to PGA above 1/10 (4V-near-remission) (19%). It is based on individual patient data meta-analysis of 11 recent clinical trials in RA (5,792 patients).
- The rate of good radiographic outcome (≤0.5 units progression over the second year) was numerically higher in patients in 4V-remission (81%; 95%CI 74 to 87%) than in those in 4V-near-remission (78%; 95%CI: 69 to 86%), but the difference is not statistically significant.
- In this population, if a 'treat-to-remission' strategy had been applied, the 3V-remission definition would have prevented therapy escalation in 19% of all patients, at the cost of an additional 6.1%, 4.0%, and 0.7% of patients having ΔmTSS>0.0, >0.5 and >5 units over 2 years, respectively.

How might this impact on clinical practice or future developments?

- These results suggest that the use of 3V-remission as the target for immunosuppressive therapy, together with a separate assessment of disease impact upon patient's lives, a dual target approach, deserves further consideration and research.

INTRODUCTION

Disease remission has become the guiding target in the management of rheumatoid arthritis (RA), as it conveys the best possible outcomes.[1] Current treatment recommendations advise that remission (or at least low disease activity) should be attained as soon and as consistently as possible, and changes in treatment should be considered when this does not happen.[2, 3]

The most influential and authoritative definition of remission was published in 2011 under the auspices of the American College of Rheumatology (ACR), the European League Against Rheumatology (EULAR) and the Outcome Measures in Rheumatology (OMERACT) groups.[4] A Boolean-based definition was endorsed: and requires that scores of tender and swollen 28-joint counts (TJC28 and SJC28), C-Reactive Protein (CRP, in mg/dl), and patient global assessment of disease activity (PGA, 0–10 scale) are all ≤1.[4]

The inclusion of PGA in the definitions of remission in RA was justified because it added predictive value for later good radiographic and functional outcomes, while conveying the much-needed patient's perspective.[4]

Despite this, the inclusion of PGA remains controversial.[5-9] Using the definitions above, studies in different clinical practice cohorts,[10-15] have reported that as many as 10%[13] to 38%[14] of all patients with RA, do not reach remission solely due to a PGA score >1, a state that has become designated as "4V-near-remission".[14, 16] Moreover, it has been demonstrated that PGA bears little relationship with markers of the disease process, which drives structural damage, rather reflecting pain, fatigue and function.[9, 17, 18] This is especially evident when analyses are restricted to the lower levels of disease activity, in the range where the definition of remission has a decisive impact on whether to maintain or to escalate immunosuppressive treatment. According to this perspective, patients in 4V-near-remission would not benefit from additional immunosuppression, as this cannot be expected to improve their condition or foster remission,[9, 17] and are exposed by current recommendations to the risk of overtreatment and unjustified side-effects.[19]

These observations have led to the suggestion that the patients' interest would be better served by the adoption of two separate complementary targets: the first focused on remission of the inflammatory process, guided by an instrument without PGA; the second focused only on patient-reported impact measures.[9, 16, 20] However, this proposal would not be sustainable if, as suggested in the original ACR/EULAR/OMEARCT paper, removing PGA from the Boolean-based remission significantly diminishes its ability to predict good radiographic and functional outcome.[4] A systematic literature review (SLR) indicated that, among the individual components included in the definitions of remission, only swollen joints and acute phase reactants are associated with radiographic progression.[21] Two other studies, using data from a clinical cohort[13] and from clinical trials,[22] compared the prediction of good radiographic outcome by "4V-remission" versus "3V-remission" (without PGA) achieved in RA patients: no significant differences were observed, but the two groups were not mutually exclusive. No study has ever compared the radiographic outcomes between the 4V-remission and 4V-near-remission groups.

The primary aim of this study was to compare 4V-near-remission and 4V-remission regarding their association with radiographic damage progression. Secondarily, we aimed to explore the impact of using 3V- instead of 4V-remission in patients with RA, both in terms of prevalence of remission and association with structural damage progression and functional impairment.

METHODS

Design and study selection

This was an individual patient data meta-analysis of published randomized controlled trials (RCTs) selected through a systematic literature review. The study protocol was registered in PROSPERO with the number CRD42017057099[23] and published elsewhere.[24]

RCTs were included if they tested the efficacy of biological disease-modifying antirheumatic drugs (bDMARDs) on ≥2-year radiographic outcomes, in patients fulfilling the 1987 ACR or the

2010 ACR-EULAR criteria for RA.[25, 26] Information on the processes of identifying and selecting studies, as well collecting data are reported in the protocol.[24]

Risk of bias assessment of individual studies

Studies selected for retrieval were assessed by two independent reviewers (RF and MN) for methodological validity prior to inclusion in this review, using the "Risk of Bias 2" tool.[27] Any disagreements between the reviewers were resolved through discussion, or with a third reviewer (JAPS). The full protocols of the studies were consulted, and their authors contacted to request missing or additional data for clarification, where required.

Specification of outcomes

Primary outcome

The primary outcome of this study was the percentage of individuals with a good radiographic outcome (GRO) during the second year of the trial (i.e. between month 12 and month 24), defined as: a change (Δ) ≤0.5 units in the van der Heijde modified-total Sharp score (mTSS).[28]

This ≤0.5 cut-off is preferred[29-31] over the one used in the ACR/EULAR pivotal publication (≤0 cut-off), because 0.5 is the optimal cut-off if the average of two readers is used,[32] as it allows to the very minimum difference of 1 unit out of 448 between the two readers.

Secondary outcomes

Two secondary endpoint cut-offs were used to define good radiographic outcome during the second year of the trial:

i. ΔmTSS≤5 units, a higher, frequently used rate (sometimes referred to as clinically non-relevant radiographic progression);
ii. ΔmTSS≤0 units, to allow comparisons with the results obtained in the ACR/EULAR study.[4]

Also as secondary outcome we studied the percentage of individuals with a good functional outcome (GFO) during the second year of the trial (i.e. between month 12 and month 24), defined as no worsening i.e. a change (Δ) ≤0.0 units in the Health Assessment Questionnaire – Disability Index (HAQ-DI). This definition has been preferred over the one used in the ACR/EULAR pivotal publication (Δ HAQ≤0.0 AND HAQ≤0.5 at both time points), because this is believed to be too strict, representing a better outcome even than expected for general population.[4, 33] Despite this consideration, this definition of GFO was also tested to allow comparison with the original ACR/EULAR paper.

Comparisons: mutually and non-mutually exclusive definitions of remission

Analyses were based on different definitions of remission states, assessed at two time points, 6 months and 12 months, following the methodology adopted by the ACR/EULAR committee,[4] as follows:

a) ACR/EULAR Boolean-based remission,[4] also designated in this study as "4V-Remission" (i.e., TJC28≤1, SJC28≤1, CRP≤1 mg/dl, and PGA≤1/10)

b) "4V-near-remission",[11, 14] defined as TJC28≤1, SJC28≤1, CRP≤1 mg/dl, and PGA>1.

c) "Non-remission" defined as TJC28>1 and/or SJC28>1 and/or CRP>1 mg/dl, irrespective of PGA value.

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

d) "3V-remission" defined as TJC28≤1, SJC28≤1, and CRP≤1 mg/dl. This is a combination of 4V-remission and 4V-near-remission - patients classified in 4V-remission also meet the 3V-remission criteria (Figure 1).

All definitions of remission were considered fulfilled if they were achieved at 6 OR 12 months' follow-up and patients were classified according to the most stringent definition they satisfied (for instance, if a patient was in 4V-near-remission at 6 months and in 4V-remission at 12 months, he/she was classified as in 4V-remission).

Data analysis and synthesis

Data analysis

All "primary" analyses were performed with SAS software (v.9.3), within the online secure platforms. For each trial we determined the number of patients with GRO in each definition group (4V-remission, 4V-near-remission, 3V remission and non-remission). The rates of true positive (TP) i.e. remission and GRO, true negative (TN) i.e. non-remission and not-GRO, false negative (FN) i.e. non-remission and GRO, and false positive (FP) i.e. remission and not-GRO cases were also determined for all definitions. The percentage of patients with accurate prediction of having and not having GRO were also determined (sum of TP and TN) for the 4V- and 3V-remission. Missing data was not substituted. Similar analyses were performed for the secondary outcomes.

Meta-analysis

Frequency of remission status and outcomes

The frequency/proportion of each remission state observed in each of the trials were meta-analysed, irrespective of the treatment arm. The same procedure was used to determine the pooled prevalence of GRO and GFO according to remission status.

Primary analysis: likelihood of achieving GRO for 4V-near-remission compared to 4V-remission and to non-remission

From our hypothesis that PGA might lead to false negative rating of remission when using the 4V-remission definition, we aimed to analyse the value of 3V-remission definition, excluding PGA. Direct comparison of 4V-remission and 3V-remission however is not possible, given the overlap between the two states (see Figure 1). Therefore, for each trial we determined the differences in the proportion/chance (Δ proportion) of GRO (ΔmTSS≤0.5) between 4V-near-remission and 4V-remission, mutually exclusive states, and then pooled these differences with the random effect model to obtain an overall estimate of the difference (with 95%CI). We also compared this between 4V-near-remission and non-remission states. The Risk Ration or Relative Risk (RR, 95%CI) for GRO between these groups were also calculated.

Secondary analyses: The likelihood of achieving each of the secondary outcomes for 4V-near-remission compared to 4V-remission and to non-remission was assessed using similar methods for the different definitions.

Sensitivity analyses

Different sensitivity analyses were performed regarding radiographic progression. The first was to explore the likelihood of GRO between remission states after excluding the seemingly outlier trials.

The second was a multivariate analysis. Multivariate logistic regressions were performed in each trial to explain GRO (dependent variable) using the mutually exclusive remission states as independent variables, adjusted for important covariates at baseline: gender, age, disease duration (except for three trials due to >50% of missing data in this covariate), rheumatoid factor status, level of radiographic damage, and treatment arm. The OR obtained in each trial and its 95%CI and standard error were meta-analysed to obtain the pooled OR of GRO comparing different mutually-exclusive remission states. However, we hypothesise that this covariate adjustment may constitute an overcorrection, because patients in remission are 'naturally' different from patients not in remission regarding these prognostic factors. For this reason, these sensitivity analyses are presented cautiously and only in supplementary material.

The third was to clarify the value of PGA as a predictor of radiographic damage progression, selecting only the patients in 4V-near-remission (in 8 of the 11 trials, 796 patients, due to restrictions in accessing the data). We used Poisson regression models with 2y mTSS as dependent variable and PGA as independent variable. To assess the specific, independent impact of PGA, we corrected for SJC28, TJC28 and CRP, determined as the mean of the observation at 6 and 12 months, by also introducing them as independent variables, together with baseline mTSS. To allow the combined analysis the different variables, we standardized their values using z-scores. A meta-analysis was then performed to obtain pooled rate ratios (RR with 95% CI) per variable.

The last was to explore the proportion of patients in 3v-remission (8 trials; 1,937 patients) who have radiographic damage progression ≥0.5 and those who have radiographic progression ≥5 during year 2, according to PGA score ≤1 versus >1 at 6 and 12 months).

Likelihood of reaching good radiographic and functional outcomes with 4V-remission compared to 3V-remission

If the null hypothesis of this study (the chance of GRO in 4V-near-remission group are similar to the 4V-remission group) is not rejected, the current 4V-remission and the proposed 3V-remission can be compared in terms of their positive (LR+) and negative likelihood ratios (LR-) of GRO per remission group. The TP, TN, FN, and FP values were used to synthesize these measures. Similar procedures were performed regarding GFO.

All meta-analyses were performed with the OpenMeta[Analyst] software,[34] using the DerSimonian-Laird random-effect method[35] and the Arcsine transformed proportion.[36] The STATA software (v.14) was used only to determine OR adjusted to covariates (sensitivity analyses). The I^2 of Higgins and Thompson was calculated to quantify heterogeneity.[37]

RESULTS

Studies and participants

From a total of 27 identified studies, we were granted access to 17 through secure online platforms, but only 11 trials reported radiographic damage progression during the second year, thus allowing inclusion in the final analyses. Reasons for the non-inclusion of 16 out of the 27 trials initially identified are described in Figure 2 and Supplementary Table S1. The critical appraisal results for each of the 11 RCTs are summarized in Supplementary Figure S1 (low risk of bias in all items assessed for all the trials). We had access to data from 100% of the randomized patients in 9 out of the 11 trials and from 93% of patients in the remaining two, resulting in a total sample of 8,114 patients. Most trials tested anti-TNFα therapies (n=9), and included patients with insufficient response to MTX (n=7) and with established disease (>2 years) (n=9) – Supplementary Table S2. The mean (SD) DAS28CRP3v ranged from 4.7 (0.9) to 5.3 (0.8) at baseline. The van der Heijde mTSS was used as the scoring method of radiographic damage progression in 10 of the trials. The remaining used the Genant method. The mean mTSS at baseline ranged from 5.9 (14.5) to 69.0 (55.8) (Supplementary Table S2).

Altogether, 2322 patients (29%) were excluded from the final analyses (Supplementary Table S3). The main reason for exclusion was the lack of data on radiographic outcome (71% of all cases). Those excluded from these analyses were older (1.3 years on average), reported higher PGA and HAQ and had more active disease according to Physician's global assessment. Regarding disease status at 6 or 12 months, 305 of the excluded patients had no data and the remaining 2017 had lower rates of 4V-remission and higher rates of non-remission, compared with those included.

Frequency of remission status, radiographic and functional outcomes

A total of 5,792 (71%) patients had information on both the remission definition and on the primary outcome (radiographic progression) (Table 1). Pooled meta-analytic frequency (95% CI) of 4V-remission at 6 OR 12 months was 23.0% (18.0 to 28.0%), while for 4V-near-remission was 18.9% (15.4 to 22.1%), considering all treatment arms together (Table 1).

Good radiographic outcome was observed in 74.1% (66.2 to 82.0%) of all patients using the primary cut-off (ΔmTSS≤0.5), and by 94.6% (92.9 to 96.4%) using ΔmTSS≤5 (Table 1). Good functional outcome, which could only be assessed in 8 RCTs (3,904 patients), was observed in 70.6% (66.7 to 73.5%) of all patients using the elected cut-off (ΔHAQ-DI≤0.0), and by 31.1% (24.9 to 37.2%) using ΔHAQ-DI≤0.0 AND HAQ-DI≤0.5 (Table 1).

Likelihood of reaching good radiographic outcome for patients in 4V-near-remission compared to patients in 4V-remission and to patients in non-remission

Overall, the proportion of GRO for the primary score (ΔmTSS≤0.5) was high (71.8 to 81.1%) for the three mutually-exclusive remission categories (Table 2). The proportion of patients with GRO did not differ significantly between those in 4V-near-remission and 4V-remission: -2.9% (95%CI: -7.3 to +1.5%). Patients in 4V-near-remission had a significantly higher chance of achieving GRO compared to patients in non-remission (+6.2%; 95%CI: 2.3 to 10.1%). Results for these comparisons are shown in Table 2 and Figure 3. Similar observations were made for GRO defined as ΔmTSS≤5 (Table 2). None of the differences was statistically significant when ΔmTSS≤0 was used (Table 2).

We performed a sensitivity analysis by excluding the three apparent outliers in Figure 3 (the DE019, GO-FURHTER, and TEMPO trials) which confirmed no significant difference in the meta-analytic RRs (ΔmTSS≤0.5) between 4V-remission and 4V-near-remission (RR=0.99; 95%CI 0.95 to 1.03).

Likelihood of reaching good functional outcome for patients in 4V-near-remission compared to patients in 4V-remission and to patients in non-remission

Overall, the proportion of GFO for the elected outcome (ΔHAQ-DI≤0.0) was high (68.8 to 77.6%) for the three mutually exclusive remission categories (Table 2). The proportion of patients with GFO was significantly lower in 4V-near-remission than 4V-remission: -11.0%

(95%CI: -16.3 to -5.7%). Patients in 4V-near-remission had a similar chance of achieving GFO compared to patients in non-remission (-2.2%; 95%CI: -6.8 to +2.4%). The differences between 4V-near-remission and 4V-remission were more striking for the GFO defined as ΔHAQ-DI≤0 AND HAQ-DI≤0.5: -39.6% (95%CI: -48.4 to -30.9%). The difference between 4V-near-remission and non-remission was non-significant (+1.7%; 95%CI: -7.4 to +10.8).

Comparison of the 4V-remission and the proposed 3V-remission regarding prediction accuracy for radiographic and functional outcome

Having shown that the difference in the probability of GRO between 4V-remission and 4V-near-remission, was neither statistically nor clinically relevant,[38] we were allowed to evaluate the difference between the 4V-remission and 3V-remission (the latter combining the 4V-near-remission and 4V-remission) groups (Table 3). The results indicated that the likelihood ratio of having GRO (ΔmTSS≤0.5) was higher for patients in 4V-remission compared to 4V-non-remission (LR+=1.36, 1.15 to 1.61) than between patients in 3V-remission vs 3V-non-remission (LR+=1.26; 1.13 to 1.41), although there was a large overlap in 95%CIs. Conversely, the likelihood of having GRO in the absence of remission was significantly smaller for the 3V-remission (LR-=0.86; 0.79 to 0.94) and non-significant for the 4V-remission (LR-=0.92; 0.81 to 1.04) vs their counterparts (Table 3).

The same comparisons were made regarding functional outcomes (Table 3). The likelihood ratio of having GFO (ΔHAQ≤0.0) was significantly higher for patients in 4V-remission compared to in 4V-non-remission (LR+=1.34, 1.16 to 1.54), while it was not significantly different between patients in 3V-remission vs 3V-non-remission (LR+=1.08; 0.99 to 1.17). Contrariwise, the likelihood of having GFO in the absence of remission was not significantly different from that for either the 3V-remission (LR-=0.94; 0.88 to 1.02) or the 4V-remission (LR-=0.90; 0.79 to 1.02) vs their comparator groups (Table 3).

The proportion of patients whose prediction of GRO was accurate (= TP + TN) was, overall, quite low for both definitions of remission (≤53%). It was, however, higher for the 3V-remission definition than for the 4V-remission definition: 6.5%, 10.6%, and17.2% higher at ΔmTSS≤0..0, ≤0.5, and ΔmTSS≤5, respectively (See Figure 4). As expected, the improved accuracy of the 3V-remission is a result of a substantially lower percentage of FN, i.e. patients without remission who do not have radiographic progression, at the cost of a much smaller increase in the percentage of FP, i.e. the patients with remission who do have progression.

Regarding the elected definition of GFO, the proportion accurately predicted with the 3V definition (50.3%; 46.0 to 54.6) was significantly higher than with the 4V definition (43.8%; 40.9 to 46.6). The percentage accurately predicted was much higher for the alternative definition of GFO, the statistically significant difference being favourable for the 4V definition.

Figure 5 presents a “clinical eye’s” summary of good/bad radiographic outcomes observed according to the current and the proposed (3V) Boolean-based definitions of remission (95%CI and I^2 statistics are presented in Supplementary Table S4). Overall, 73.3% (95%CI: 63.9% to 81.8%) of the patients in non-4V-remission still had GRO (ΔmTSS≤0.5), and the same was observed for 71.8% (95%CI: 62.1% to 80.5%) of those in non-3V-remission. The percentages of GRO increase to 81.1% (95%CI: 74.4% to 86.9%) and 79.6% (95%CI: 72.2% to 86.1%) among those in 4V and 3V-remission, respectively. None of these differences were statistically significant.

The overall proportion of patients achieving 3V-remission was almost double of those reaching 4V-remission (41.9% vs 23.0%).

Sensitivity analyses

Adjustment to co-factors. The models adjusted for co-factors for the same comparisons showed even smaller differences between 4V-near-remission and 4V-remission categories regarding the prediction of good radiographic outcomes (Supplementary Table S5 and S6).

Exploration of radiographic damage in 4V-near-remission. Within the subgroup of patients in 4V-near-remission, PGA (at 6 and 12 months) is not a statistically significant predictor of radiographic progression over 2y (RR= 1.05 per SD unit increase, 95%CI: 0.93 to 1.16); similarly, non-significant results were obtained for SJC28 and TJC28 (both 0 vs 1 in this subgroup): RR= 1.09; 95%CI 0.90 to 1.27, and RR=0.86; 95%CI 0.68 to 1.04, respectively. Only CRP was a (borderline) statistically significant predictor of radiological progression (RR = 1.06, 95%CI 1.00 to 1.12).

Radiographic damage progression according to PGA. In the subgroup of patients reaching 3V-remission a ∆mTSS>5 units, was observed in 2.3% (95%CI: 1.0 to 4.3%) of patients scoring PGA>1 and in 1.3% (0.6 to 2.3%) of those with PGA <1. The corresponding values for ∆mTSS>0.5 units were 18.4% (13.8 to 23.5%) and 15.2% (9.9 to 21.4%), respectively. (Supplementary Table S7).

DISCUSSION

This is the first study assessing the prevalence of 4V-near-remission in RCTs and the first comparing radiographic damage progression between patients in 4V-near-remission and in 4V-remission. The pooled rate of 4V-near-remission was almost the same of 4V-remission (19% vs 23%). These mutually exclusive groups did not differ significantly in terms of subsequent radiographic damage accrual. Patients in 4V-near-remission had a significantly better radiographic outcome than those in non-remission.

These observations legitimised the next step in our analyses: to explore the implications of choosing between the 3V and the 4V definitions of remission. The odds of good structural outcome were slightly higher for the 4V-remission, but without statistical, or, in our view, clinical significance. The 3V-remission showed a better performance in terms of true estimations of significant damage (i.e. sum of TP and TN estimations). If a ‘treat-to-remission’ strategy had been applied in this population, the 3V-remission definition would have prevented therapy escalation in 19% of all participants, when compared to the 4V-remission. This would occur at the cost of having an excess of 6.1% of patients having a ΔmTSS>0.0, 4.0% of patients having a ΔmTSS>0.5 and of 0.7% having ΔmTSS>5 units. These trade-offs may be differently valued by different observers. Our proposal to use the 3V-remission definition is also rooted in solid clinical common sense: a (major) part of patients who fail remission solely because of PGA is not be expected to benefit from additional immunosuppressive therapy, as PGA does not reflect disease activity in these patients. However, clinical judgement is needed as to decide in individual patients whether the PGA level > 1 indicates residual disease activity that might be successfully treated with more intensive RA treatment, or reflects another cause, for which more intensive RA treatment would be unnecessary and potentially harmful. Guiding definitions and recommendations should always be aligned with good clinical wisdom.

The data also emphasizes that all remission concepts have a relatively poor predictive value regarding radiographic damage, as shown by low LRs (although better in 4V-remission) and predictive accuracies below 53% (better in 3V-remission). This reflects the fact that 73% of patients in non-4V-remission had good radiographic outcomes and 19% of those in 4V-remission still presented radiographic progression (ΔmTSS>0.5).

4V-remission was associated with significantly higher rates of GFO (77.6%), compared to 4V-near-remission (66.9%); this latter rate is similar to that observed in non-remission (68.8%). The differences were more marked in favour of a 4V-remisision if the definition of GFO adopted by the ACR/EULAR committee was used (4V-remission=60.5%, 4V-near-remission=22.5%, and non-remission=21.2%). Positive likelihood ratios also favoured 4V-remission, while negative LRs did not reach significance in favour of 4V-near-remission. The predictive accuracy of 3V-remission for the elected functional outcome was numerically better than for 4V-remission, nearly reaching statistical significance.

The results regarding functional outcome demand a critical appraisal. Overall, PGA and HAQ-DI are correlated to the level r = 0.5 to 0.7. In higher disease activity states, both PGA and HAQ-DI predominantly reflect disease activity. In remission, they are expected to remain correlated, even if one assumes (as we do) that neither of them substantially reflects inflammation at this stage, because they are essentially determined by similar subjective factors and comorbidities [9, 14, 17, 39] It follows that, irrespective of disease activity, PGA is bound to predict HAQ-DI, and this obviously questions the use of HAQ-DI to assess the use of PGA, especially in a definition of remission, if it is intended to guide decisions on immunosuppressive therapy. The current results confirm this interpretation: How else could we coherently explain that, also in our study, 4V-remission is associated with significantly higher prevalence of GFO than 4V-near-remission if these two conditions share similar levels of SJC28, TJC28 and CRP (all ≤1) and similar levels of radiographic progression? The only difference is PGA.

The robustness of this work is supported by (i) the use of individual patient data, allowing uniform analyses procedures, (ii) the availability of data collected under stringent RCT conditions, (iii) the inclusion of over 5,700 patients, and (iv) the use of both crude and adjusted statistical analyses. This study also has potential limitations and biases. The definition of remission was based only on two independent time-points (6 OR 12 months) and used to predict radiographic progression over the following year. Although this was also the methodology used by the ACR/EULAR group,[4] it is recognized that alternative ways exist to quantify sustained remission, which might be useful both in understanding the construct of remission and investigating its relationship with structural damage accrual.[4] Good outcome was assessed only within the second year after randomization. Although this is the efficacy endpoint used in most trials, longer follow-up assessment could provide different results.[40] When 3V-remission is agreed to be an acceptable endpoint for evaluating disease modifying treatment in RA, the ability of the 3V-remission definition to detect differences between (effective) treatments, i.e. its responsiveness, should be established and compared to that of 4V-remission and other established trial endpoints in RA. Patients with missing data, excluded from the analysis, had higher PGA and HAQ-DI scores and more active disease at 6 and 12 months, but they were not significantly different with regards to other factors recognised as relevant for radiographic outcome. The exclusion of these patients might have changed the relationship between disease activity status and the outcomes under consideration in an unknown direction. It should be noted that we did not analyse within trial arms and used the data of clinical trials as in observational studies, therefore discarding the effects of randomization. As patients fulfilled inclusion criteria for RCTs, generalizability of our results is limited to patients with high disease activity starting treatment. In 7 out of the 11 RCTs, joint assessments were performed by independent assessors, and the 4 other studies did not use an independent joint assessor. We do not know whether this may have affected the (interpretation of the) results of our study in any way. Finally, some changes to the published protocol for this study need to be disclosed, namely the use of ΔmTSS≤0.5 units as the primary outcome instead of the ≤0 cut-off, for the reasons outlined in the methods section.

The most relevant implications of this study for clinical practice and research relate to the most appropriate definition of remission and its use as the guiding target for therapy. Our results demonstrate that patients in 4V-near-remission do not differ significantly from those in 4V-remission in terms of radiographic damage accrual, while they can be clearly separated from those in non-remission. This supports the aggregation of the first two groups, i.e. the proposed 3V-remission definition. Contrary to ACR/EULAR,[4] but in line with previous and current evidence,[13, 21, 22, 41] our results demonstrated that the 3V-remission definition does not significantly diminish the ability to predict structural damage, while it may significantly reduce the risk of overtreatment, but this should be validated in clinical settings.[19, 20] The implications of these observations should be further tested in the remission definitions based on composite indices SDAI and CDAI, as also endorsed by ACR/EULAR.

The ACR/EULAR committee also addressed the 3V-definition and reached the opposite conclusion.[4] This may be explained by differences in methodology and reasoning. First, ACR/EULAR tested one single and very strict cut-off to define good radiographic outcome (ΔmTSS≤0), which is, in our view, excessively stringent, as it does not even allow for a difference of one unit in change score in the total of 448 joints assessed by the 2 radiograph assessors, which is averaged to 0.5. Both cut-offs are well below the smallest detectable change within one subject: 2-3 units according to an OMERACT expert panel.[38] However, in our study, the ΔmTSS≤0 was the one with more favourable results for the 4V compared to the 3V-remission in terms of GRO prediction, predictive accuracy, and rate of FN, but not in LR, for which the ΔmTSS ≤0.5 was more favourable. While considering these issues, one should take into account that ΔmTSS =1 has been estimated to justify a decrease of the HAQ score of only 0.01.[42] Second, the ACR/EULAR committee limited their analysis to 4V vs 3V, which significantly overlap, thus "diluting" the characteristics of a very unique group of patients: 4V-near-remission. Also, the number of patients analysed by ACR/EULAR was much lower. Furthermore, the decision of the ACR/EULAR committee was, seemingly, strongly influenced by the much better prediction of good functional and "overall" good outcomes for the 4V- versus the 3V-remission. This position was recently reaffirmed.[22] The reasons why we disagree with this approach are presented above. Furthermore, the ACR/EULAR study analysed primarily the methotrexate-alone treatment groups of three trials, while we included all arms in each of eleven trials. This may explain why our likelihood ratios of GRO between 4V-remission and non-remission are much lower than the ACR/EULAR study, given that inhibition of radiographic damage by bDMARDs has been demonstrated even in the absence of remission, thus reducing the predictive accuracy of disease activity for radiographic damage.[43-45] However, we performed a sensitivity analysis, using data from patients in the monotherapy bDMARD arms (in 9 RCTs), which showed that bDMARDs indeed reduce structural damage, and result in GRO in the majority, but not universally. Altogether, 28% of all patients exposed to bDMARDs monotherapy presented ΔmTSS≥0.5 (11 to 57% in the individual trials; data not shown). In summary, we believe that our approach is valid and provides a better representation of current clinical practice. However, it will not fit contexts where access to bDMARDs is severely limited. Finally, the selection of tools by the ACR/EULAR committee was "based (...) on the need to include patient-reported outcomes", among other factors.[4] PGA was selected because it is associated with better prediction of the combination of radiographic and functional outcome.[4] While this is valid in the overall spectrum of disease activity, this argument is no longer true when the disease process is under control (SJC28, TJC28 and PCR ≤1) as demonstrated in this study and elsewhere.[17] It has been proposed to raise the cut-off value of PGA [22, 46, 47] but this is at best a partial solution: we previously found that among 4,381 international patients in 3v-remission, 63% scored PGA>1, but still 44% scored it >2, 32% >3 and 0.6% scored PGA as high as 10.[17] In addition, PGA at low disease activity states is essentially determined by subjective factors and comorbidities,[9, 17, 18] in contrast to e.g. swollen joint counts and CRP. The current study shows that PGA has no significant relationship with radiographic damage progression, both by comparing the 4V and 3V remission groups and by analysing the relationship between the 2 parameters within the specific group of patients in 4V-near-remission. These observations support our view to leave it out of the treatment target definition used to control inflammation (biological remission).

It has been recognized that treating to target often leaves room for improvement.[48] For patients with active disease, there is little doubt that controlling the disease is the most important means to improve the patient’s condition, both at short and long-term. Once low disease activity or remission is achieved, a persistently high disease impact should become the guiding target: after a diligent search for remaining (undetected) disease activity, it needs to be analysed and understood so as to choose the best adjunctive intervention, such as analgesia, rehabilitation or anti-depressive therapy, among other pharmacological and non-pharmacological therapies.[49] PGA score is not appropriate for this purpose, and more analytic instruments, such as the Patient Reported Outcome Measurement Information System (PROMIS),[50] the RA Impact of Disease (RAID) score[51, 52] or the RA Flare Questionnaire[53] are required.

Overall, these results support the proposal that the 3V definition of remission in parallel with a separate evaluation of the patient's perspective, i.e. the dual target strategy, deserves consideration. The first target aims to control of inflammation (biological remission) and the other one to control of disease impact (symptom remission), guided by clinically informative PROMs.[9, 16, 20] Pursuing and achieving the first is an important contribution, but no guarantee that the second will be fulfilled. Further research, specifically regarding adjuvant interventions required to achieve effective control of disease impact endured by patients in biological remission designed to bring patients from 4V-near-remission into full remission is warranted to validate the concept of dual-target. Improving symptoms and signs of RA, both short and long term is the major goal of treatment and it deserves being highlighted by an independent treatment target.

Competing interests

RJOF reports a research grant from AbbVie, and speaker fees from Sanofi Genzyme, Amgen, MSD, and UCB Pharma. PMJW: no conflicts of interest. JWGJ reports a research grant from Roche. LG reports a research grant from Lilly, Mylan, Pfizer, and Sandoz, and speaker fees from AbbVie, Amgen, Biogen, Celgene, Janssen, Lilly, MSD, Novartis, Pfizer, Sandoz, Sanofi-Aventis, and UCB Pharma. MN reports a research grant from Bristol Myers Squibb, and speaker fees from Janssen, and Pfizer. PMM reports speaker fees from AbbVie, Celgene, Janssen, Lilly, MSD, BMS, Novartis, Pfizer, Roche and UCB Pharma. DvdH is Director of Imaging Rheumatology bv, and reports speaker fees from AbbVie, Amgen, Astellas, AstraZeneca, BMS, Boehringer Ingelheim, Celgene, Cyxone, Daiichi, Eisai, Eli-Lilly, Galapagos, Gilead, Glaxo-Smith-Kline, Janssen, Merck, Novartis, Pfizer, Regeneron, Roche, Sanofi, Takeda, UCB Pharma. JAPS reports a research grant from Pfizer and AbbVie, and speaker fees from Pfizer, AbbVie, Roche, Lilly, Novartis.

Contributorship

All authors designed the study and protocol, which was firstly drafted by RJOF and JAPS. RJOF and PMJW performed the data analyses. RJOF and JAPS wrote the initial draft of the manuscript, which was critically revised and refined by all authors. All authors formally approved the final manuscript.

Acknowledgments

We would like to acknowledge the invaluable support provided from Jos van der Velden (SAS Portugal), who assisted us with the use of SAS software and access to the SAS Clinical Trial Data Transparency Portal. We also acknowledge the support from Adam LaMana (SAS International) and from the personal from "data sharing" teams from Pfizer, AbbVie, Roche, UCB and YODA.

We also would like to acknowledge the support of Eduardo Santos (Coimbra, Portugal) in performing the meta-analyses.

Funding info

This manuscript is based on research using data from data contributors AbbVie, Pfizer & UCB that has been made available through Vivli, Inc. Vivli has not contributed to or approved, and is not in any way responsible for, the contents of this publication. This study was also supported by CSDR (ClinicalStudyDataRequest), which has an agreement with Roche Inc. (Project #1808), who provided the authors with access to the data but did not sponsor or influence this analysis. Data was also obtained from the Yale University Open Data Access Project (YODA Project #2017-1451), which has an agreement with JANSSEN RESEARCH & DEVELOPMENT, L.L.C.. The interpretation and reporting of research using this data are solely the responsibility of the authors and does not necessarily represent the official views of the Yale University Open Data Access Project or JANSSEN RESEARCH & DEVELOPMENT, L.L.C..

Pedro M. Machado is supported by the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre (BRC). The views expressed are those of the authors and not necessarily those of the (UK) National Health Service (NHS), the NIHR, or the (UK) Department of Health, or any other governmental, academic, commercial or industrial institution.

Ricardo Ferreira was supported by a grant from ARCo – Associação de Reumatologia de Coimbra, a non-profit association of health professionals.

Ethical approval information

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

Data sharing statement

Data may be obtained from a third party and are not publicly available. Data have been provided by the respective sponsors of the trials. Any requests for individual patient level data will have to be addressed to these sponsors directly.

Patient and Public Involvement

We did not directly include PPI in this study, but the main concepts and research questions formulated and answered prior to this study were developed with PPI (co-authors in the publications).

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 RBM, Bijlsma JWJ, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann Rheum Dis. 2020;10.1136/annrheumdis-2019-216655.
3. Singh JA, Saag KG, Bridges SL, et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Care Res (Hoboken). 2016;68:1-25.
4. 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. Ann Rheum Dis. 2011;70:404-13.
5. van Tuyl LH, Boers M. Patient's global assessment of disease activity: what are we measuring? Arthritis Rheum. 2012;64:2811-3.
6. 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.
7. van Tuyl LHD, Boers M. Rheumatoid arthritis: Remission - keeping the patient experience front and centre. Nat Rev Rheumatol. 2017;13:573-4.
8. Ferreira RJO, Duarte C, Ndosi M, et al. The controversy of using PGA to define remission in RA. Nat Rev Rheumatol. 2018;14:245.
9. Ferreira RJO, Duarte C, Ndosi M, et al. Suppressing Inflammation in Rheumatoid Arthritis: Does Patient Global Assessment Blur the Target? A Practice-Based Call for a Paradigm Change. Arthritis Care Res (Hoboken). 2018;70:369-78.
10. 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;51:1076-80.
11. 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;71:1702-5.
12. 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.
13. Svensson B, Andersson ML, Bala SV, et al. Long-term sustained remission in a cohort study of patients with rheumatoid arthritis: choice of remission criteria. BMJ Open. 2013;3:e003554.
14. 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; 2017;56:1573-8.

15. Gossec L, Kirwan JR, de Wit M, et al. Phrasing of the patient global assessment in the rheumatoid arthritis ACR/EULAR remission criteria: an analysis of 967 patients from two databases of early and established rheumatoid arthritis patients. Clin Rheumatol. 2018;37:1503-10.
16. Ferreira RJO, Ndosi M, de Wit M, et al. Dual target strategy: a proposal to mitigate the risk of overtreatment and enhance patient satisfaction in rheumatoid arthritis. Ann Rheum Dis. Ann Rheum Dis. 2019;78:e109.
17. Ferreira RJO, Carvalho PD, Ndosi M, et al. Impact of patient global assessment on achieving remission in patients with rheumatoid arthritis: a multinational study using the METEOR database. Arthritis Care Res (Hoboken). 2019;71:1317–25.
18. Ward MM, Guthrie LC, Dasgupta A. Direct and indirect determinants of the patient global assessment in rheumatoid arthritis: Differences by level of disease activity. Arthritis Care & Res (Hoboken). 2017;69:323-9.
19. Landewe RBM. Overdiagnosis and overtreatment in rheumatology: a little caution is in order. Ann Rheum Dis. 2018; 77:1394-6.
20. Landewe RBM. Response to: 'Dual target strategy: a proposal to mitigate the risk of overtreatment and enhance patient satisfaction in rheumatoid arthritis' by Ferreira et al. Ann Rheum Dis. 2019 Oct;78(10):e110
21. Navarro-Compan V, Gherghe AM, Smolen JS, et al. Relationship between disease activity indices and their individual components and radiographic progression in RA: a systematic literature review. Rheumatology (Oxford). 2015;54:994-1007.
22. Studenic P, Felson D, de Wit M, et al. Testing different thresholds for patient global assessment in defining remission for rheumatoid arthritis: are the current ACR/EULAR Boolean criteria optimal? Ann Rheum Dis. 2020 Apr;79(4):445-452.
23. 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 [Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017057099.
24. Ferreira RJO, Welsing PMJ, Gossec L, et al. 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. Acta Reumatol Port. 2018;43:52-60.
25. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988;31:315-24.

26. 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;69:1580-8.
27. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
28. Ory PA. Interpreting radiographic data in rheumatoid arthritis. A Ann Rheum Dis. 2003;62:597-604.
29. van der Heijde D, Schiff M, Tanaka Y, et al. Low rates of radiographic progression of structural joint damage over 2 years of baricitinib treatment in patients with rheumatoid arthritis. RMD Open. 2019;5:e000898.
30. Smolen JS, Pedersen R, Jones H, et al. Impact of flare on radiographic progression after etanercept continuation, tapering or withdrawal in patients with rheumatoid arthritis. Rheumatology (Oxford). 2020;59:153-64.
31. 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.
32. van der Heijde D, Simon L, Smolen J, et al. How to report radiographic data in randomized clinical trials in rheumatoid arthritis: guidelines from a roundtable discussion. Arthritis Rheum. 2002;47:215-8.
33. Sokka T, Kautiainen H, Hannonen P, et al. Changes in Health Assessment Questionnaire disability scores over five years in patients with rheumatoid arthritis compared with the general population. Arthritis Rheum. 2006;54:3113-8.
34. Wallace BC, Dahabreh IJ, Trikalinos TA, et al. Closing the Gap between Methodologists and End-Users: R as a Computational Back-End. Journal of Statistical Software. 2012;49:1-15.
35. Macaskill P, Gatsonis C, Deeks JJ, et al. 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.
36. Barendregt JJ, Doi SA, Lee YY, et al. Meta-analysis of prevalence. J Epidemiol Community Health. 2013;67:974-8.
37. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21:1539-58.
38. Bruynesteyn K, van der Heijde D, Boers M, et al. Determination of the minimal clinically important difference in rheumatoid arthritis joint damage of the Sharp/van der Heijde and Larsen/Scott scoring methods by clinical experts and comparison with the smallest detectable difference. Arthritis Rheum. 2002;46:913-20.

39. Carvalho PD, Ferreira RJO, Landewe R, et al. Association of seventeen definitions of remission with functional status in a large clinical practice cohort of patients with rheumatoid arthritis. J Rheumatol. 2020;47:20-7.
40. Landewe R, Ostergaard M, Keystone EC, et al. Analysis of integrated radiographic data from two long-term, open-label extension studies of adalimumab for the treatment of rheumatoid arthritis. Arthritis Care Res (Hoboken). 2015;67:180-6.
41. Ferreira RJO, Fautrel B, Saraux A, et al. Patient global assessment and radiographic progression in early arthritis: 3-year results from the ESPOIR cohort. Arthritis Care & Res (Hoboken). 2020 Apr 27. doi: 10.1002/acr.24237.
42. Smolen JS, Aletaha D, Grisar JC, et al. Estimation of a numerical value for joint damage-related physical disability in rheumatoid arthritis clinical trials. Ann Rheum Dis. 2010;69:1058-64.
43. Smolen JS, Han C, Bala M, et al. Evidence of radiographic benefit of treatment with infliximab plus methotrexate in rheumatoid arthritis patients who had no clinical improvement: a detailed subanalysis of data from the anti-tumor necrosis factor trial in rheumatoid arthritis with concomitant therapy study. Arthritis Rheum. 2005;52:1020-30.
44. Smolen JS, Han C, van der Heijde DM, et al. Radiographic changes in rheumatoid arthritis patients attaining different disease activity states with methotrexate monotherapy and infliximab plus methotrexate: the impacts of remission and tumour necrosis factor blockade. Ann Rheum Dis. 2009;68:823-7.
45. Landewe R, van der Heijde D, Klareskog L, et al. Disconnect between inflammation and joint destruction after treatment with etanercept plus methotrexate: results from the trial of etanercept and methotrexate with radiographic and patient outcomes. Arthritis Rheum. 2006;54:3119-25.
46. 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;39:1139-45.
47. Wells GA, Boers M, Shea B, et al. Minimal disease activity for rheumatoid arthritis: a preliminary definition. J Rheumatol. 2005;32:2016-24.
48. Schoemaker CG, de Wit MP. Treat to target from the patient perspective is bowling for a perfect strike. Arthritis Rheumatol. 2020 Aug 2. doi: 10.1002/art.41461.
49. Santos EJF, Duarte C, Marques A, et al. Effectiveness of non-pharmacological and non-surgical interventions for rheumatoid arthritis: an umbrella review. JBI Database System Rev Implement Rep. 2019;17:1494-531.
50. Bartlett SJ, Orbai AM, Duncan T, et al. Reliability and Validity of Selected PROMIS Measures in People with Rheumatoid Arthritis. PLoS ONE. 2015;10:e0138543.

51. Gossec L, Paternotte S, Aanerud GJ, et al. Finalisation and validation of the rheumatoid arthritis impact of disease score, a patient-derived composite measure of impact of rheumatoid arthritis: a EULAR initiative. Ann Rheum Dis. 2011;70:935-42.
52. Ferreira RJO, Gossec L, Duarte C, et al. The Portuguese Rheumatoid Arthritis Impact of Disease (RAID) score and its measurement equivalence in three countries: validation study using Rasch Models. Qual Life Res. 2018;27:2909-21.
53. Bartlett SJ, Barbic SP, Bykerk VP, et al. Content and Construct Validity, Reliability, and Responsiveness of the Rheumatoid Arthritis Flare Questionnaire: OMERACT 2016 Workshop Report. J Rheumatol. 2017;44:1536-43.

Table 1 Frequency of remission and good radiographic outcome in the included studies

Trial (year)	n [a]	Remission at 6 OR 12 months, n (%)			Good Radiographic outcome from 12 to 24 months [b], n (%)			Good functional outcome from 12 to 24 months, n (%)		
		4V-remission	4V-near-remission	Non-remission	ΔmTSS≤0	ΔmTSS≤0.5	ΔmTSS≤5	n total	ΔHAQ-DI≤0[c]	ΔHAQ-DI≤0 AND HAQ-DI ≤0.5
DE019 (2004)	425	68 (16.0)	45 (10.6)	312 (73.4)	245 (57.6)	297 (69.9)	397 (93.4)	398	281 (70.6)	114 (28.6)
TEMPO (2004)	442	113 (25.6)	91 (20.6)	238 (53.8)	282 (63.8)	330 (74.7)	423 (95.7)	421	300 (71.3)	152 (36.1)
COMET (2008)	344	102 (29.7)	107 (31.1)	135 (39.2)	250 (72.7)	289 (84.0)	329 (95.6)	324	237 (73.1)	138 (42.6)
RAPID 1 (2008)	650	177 (27.2)	143 (22.0)	330 (50.8)	424 (65.2)	508 (78.2)	636 (97.7)	642	420 (65.4)	135 (21.0)
RAPID 2 (2009)	417	51 (12.2)	81 (19.4)	285 (68.4)	286 (68.6)	324 (77.7)	398 (95.4)	435	290 (66.7)	79 (18.2)
GO-FORWARD (2010)	352	86 (24.4)	74 (21.0)	192 (54.6)	200 (56.8)	228 (64.8)	304 (86.4)	358	na	105 (29.3)
GO-BEFORE (2011)	499	117 (23.5)	80 (16.0)	302 (60.5)	403 (80.8)	446 (89.4)	493 (98.8)	507	na	187 (36.9)
LITHE (2011)	796	146 (18.3)	174 (21.9)	476 (59.8)	558 (70.1)	640 (80.4)	790 (99.2)	550	369 (67.1)	123 (22.4)
DE013 (2013)	540	156 (28.9)	50 (9.3)	334 (61.8)	286 (53.0)	351 (65.0)	483 (89.4)	518	383 (73.9)	249 (48.1)
GO-FURTHER (2014)	483	54 (11.2)	89 (18.4)	340 (70.4)	151 (31.3)	191 (39.5)	405 (83.9)	493	na	94 (19.1)
FUNCTION (2016)	844	308 (36.5)	151 (17.9)	385 (45.6)	713 (84.5)	766 (90.8)	840 (99.5)	616	470 (76.3)	250 (40.6)
Total n	5,792	1,378	1,085	3,329	3,798	4,370	5,498	5,262[c]	2,750	1,626
Meta-analytic %		23.0	18.9	58.1	64.1	74.1	94.6		70.6	31.1
(95% CI)		(18.0 to 28.0)	(15.4 to 22.1)	(52.0 to 64.1)	(54.9 to 73.2)	(66.2 to 82.0)	(92.9 to 96.4)		(67.7 to 73.5)	(24.9 to 37.2)

a. Number of patients with information both on remission status and on radiographic outcome

b. All trials used van der Heijde mTSS (0 to 448) except the LITHE trial, in which the Genant mTSS (0 to 202) was used instead.

c. Not possible to be determined in the three golimumab trials due to changes that occurred in the research environment and statistical software available since the initial data analyses (thus, n=3,904)

Legend: 4V-remission = SJC28, TJC28, CRP (in mg/dl), and PGA (0-10), all ≤1; 4V-near-remission= SJC28, TJC28, CRP (in mg/dl) ≤1 and PGA (0-10)>1; Non-remission = SJC28 >1 OR TJC28>1 OR CRP (in mg/dl)>1 at 6 OR 12 months of follow-up in all cases; ΔmTSS = change in the modified Total Sharp Score during the second year of follow-up

Table 2: Pooled outcomes[a] and measures of association between remission categories and good radiographic and good functional outcomes, during the second year of follow-up.

	Good Radiographic Outcome (GRO) defined as ∆**mTSS≤0.5**		
	4V-remission (n=1,378)	***4V-near-remission (n=1,085)***	***Non-remission (n=3,329)***
Percentage GRO (95% CI)	81.1 (74.4 to 86.9)	78.2 (69.5 to 85.8)	71.8 (62.1 to 80.5)
	4V-near-remission vs 4V-remission		***4V-near-remission vs Non-remission***
∆ percentage GRO (95% CI)	−2.9 (−7.3 to 1.5)		6.2 (2.3 to 10.1)
Relative Risk GRO (95% CI)	0.98 (0.94 to 1.02)		1.07 (1.02 to1.12)
	Good Radiographic Outcome (GRO) defined as ∆**mTSS≤0**		
	4V-remission	***4V-near-remission***	***Non-remission***
Percentage GRO (95% CI)	71.5 (63.5 to 78.8)	64.1 (54.6 to 73.2)	62.2 (51.5 to 72.4)
	4V-near-remission vs 4V-remission		***4V-near-remission vs Non-remission***
∆ percentage GRO (95% CI)	-7.7 (-16.6 to 1.1)		1.7 (-8.1 to 11.5)
Relative Risk GRO (95% CI)	0.91 (0.82 to 1.02)		1.04 (0.94 to 1.16)
	Good Radiographic Outcome (GRO) defined as ∆**mTSS≤5**		
	4V-remission	***4V-near-remission***	***Non-remission***
Percentage GRO (95% CI)	97.5 (95.4 to 98.9)	96.1 (92.5 to 98.5)	94.2 (90.2 to 97.2)
	4V-near-remission vs 4V-remission		***4V-near-remission vs Non-remission***
∆ percentage GRO (95% CI)	-2.5 (-7.5 to 2.6)		4.1 (0.7 to 7.6)
Relative Risk GRO (95% CI)	99.9 (0.97 to 1.01)		1.01 (1.00 to 1.02)
	Good Functional Outcome (GFO) defined as **ΔHAQ-DI≤0**		
	4V-remission (n=1,041)	***4V-near-remission (n=758)***	***Non-remission (n=2,105)***
Percentage GFO (95% CI)	77.6 (74.3 to 80.8)	66.9 (62.6 to 71.2)	68.8 (66.0 to 71.7)
	4V-near-remission vs 4V-remission		***4V-near-remission vs Non-remission***
∆ percentage GFO (95% CI)	-11.0 (-16.3 to -5.7)		-2.2 (-6.8 to 2.4)
Relative Risk GFO (95% CI)	0.87 (0.81 to 0.94)		0.98 (0.92 to 1.04)
	Good Functional Outcome (GFO) defined as **ΔHAQ-DI≤0 AND HAQ-DI≤0.5**		
	4V-remission (n=1,305)	***4V-near-remission (n=1,003)***	***Non-remission (n=2,954)***
Percentage GFO (95% CI)	60.2 (53.3 to 67.0)	22.5 (15.9 to 29.1)	21.2 (16.1 to 26.3)

	4V-near-remission vs 4V-remission	*4V-near-remission vs Non-remission*
∆ percentage GFO (95% CI)	−39.6 (−48.4 to -30.9)	1.7 (-7.4 to 10.8)
Relative Risk GFO (95% CI)	0.37 (0.30 to 0.46)	1.12 (0.82 to 1.53)

Legend: 4V-remission = SJC28, TJC28, CRP (in mg/dl), and PGA (0-10), all ≤1; 4V-near-remission= SJC28, TJC28, CRP (in mg/dl) ≤1 and PGA (0-10)>1; Non-remission = SJC28 >1 OR TJC28>1 OR CRP (in mg/dl)>1, irrespective of PGA value; at 6 OR 12 months of follow-up in all cases; ∆mTSS = change in the modified Total Sharp Score during the second year of follow-up; GRO = Good Radiographic Outcome.

a. Determined by meta-analyses: for each trial, we calculated the differences in the proportion/chance (∆ proportion) of GRO or GFO between 4V-near-remission and 4V-remission states and between 4V-near-remission and non-remission states; then, we pooled these differences with a random effects model to obtain an overall estimate of the difference (with 95%CI).

Table 3. Meta-analyses of good outcomes likelihood ratios for the 4V- and 3V-remission status.

Good Outcome [a]	4V-Remission (versus non-4V)		I^2 LR+ LR-	3V-Remission (versus non-3V)		I^2 LR+ LR-
	LR+ (95% CI)	LR- (95% CI)		LR+ (95% CI)	LR- (95% CI)	
ΔmTSS≤ 0.5	1.36 (1.15 to 1.61)	0.92 (0.81 to 1.04)	38% 0%	1.26 (1.13 to 1.41)	0.86 (0.79 to 0.94)	40% 3%
ΔmTSS≤ 0	1.32 (1.17 to 1.50)	0.91 (0.82 to 1.02)	19% 0%	1.20 (1.12 to 1.29)	0.87 (0.81 to 0.93)	0% 0%
ΔmTSS≤ 5	1.40 (0.88 to 2.23)	1.01 (0.76 to 1.33)	56% 0%	1.33 (1.03 to 1.71)	0.92 (0.77 to 1.10)	40% 0%
ΔHAQ-DI ≤0	1.34 (1.16 to 1.54)	0.90 (0.79 to 1.02)	18% 0%	1.08 (0.99 to 1.17)	0.94 (0.88 to 1.02)	17% 0%
ΔHAQ-DI ≤0 AND HAQ-DI ≤0.5	3.35 (2.78 to 4.03)	0.60 (0.52 to 0.68)	72% 45%	1.82 (1.59 to 2.07)	0.55 (0.47 to 0.65)	80% 87%

a. n=5,792 for ΔmTSS, n=3,904 for ΔHAQ-DI≤0 and n= 5,262 for ΔHAQ-DI≤0 AND HAQ-DI≤0.5,

Legend: 4V-remission = SJC28, TJC28, CRP (in mg/dl), and PGA (0-10), all ≤1; 3V-remission= SJC28, TJC28, CRP (in mg/dl) ≤1; Non-remission = SJC28 >1 OR TJC28>1 OR CRP (in mg/dl)>1, irrespective of PGA value; at 6 OR 12 months of follow-up in all cases ΔmTSS = change in the modified Total Sharp during the second year of follow-up. LR+ = Positive Likelihood Ratio; LR- = Negative Likelihood Ratio. I^2: heterogeneity index.

FIGURE CAPTIONS

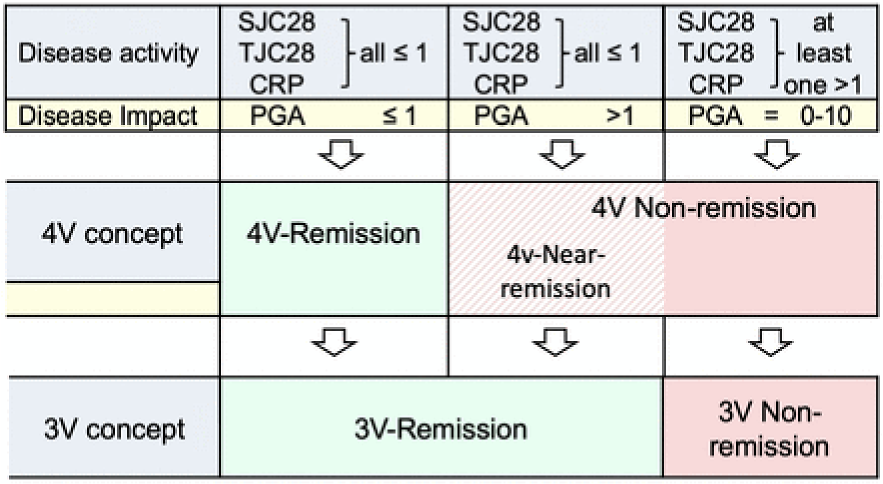

Figure 1 – Definitions of remission tested in the study

Legend: SJC28 = swollen 28-joint count, range 0-28; TJC28 = tender 28-joint count, range 0-28; CRP = C-reactive protein, mg/dl; PGA = patient global assessment, range 0-10 = worst.

Footnote: In general, in no remission states, disease-modifying antirheumatic drug (DMARD) therapy will be intensified, while at remission states, DMARD therapy will be unchanged or tapered. The no remission/4V-near-remission state (hatched) has a risk of overtreatment, if DMARD therapy is intensified.

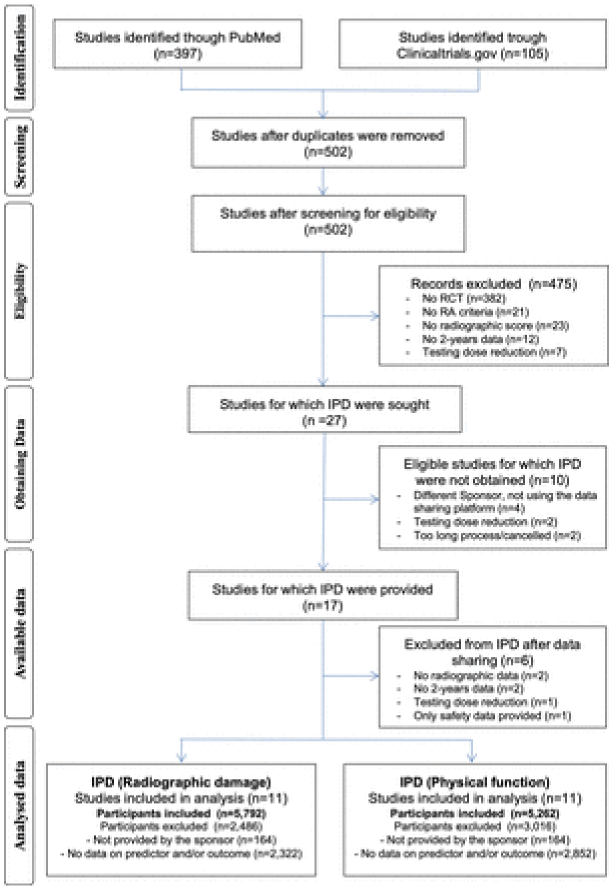

Figure 2 - Flowchart with the process of study identification and data access

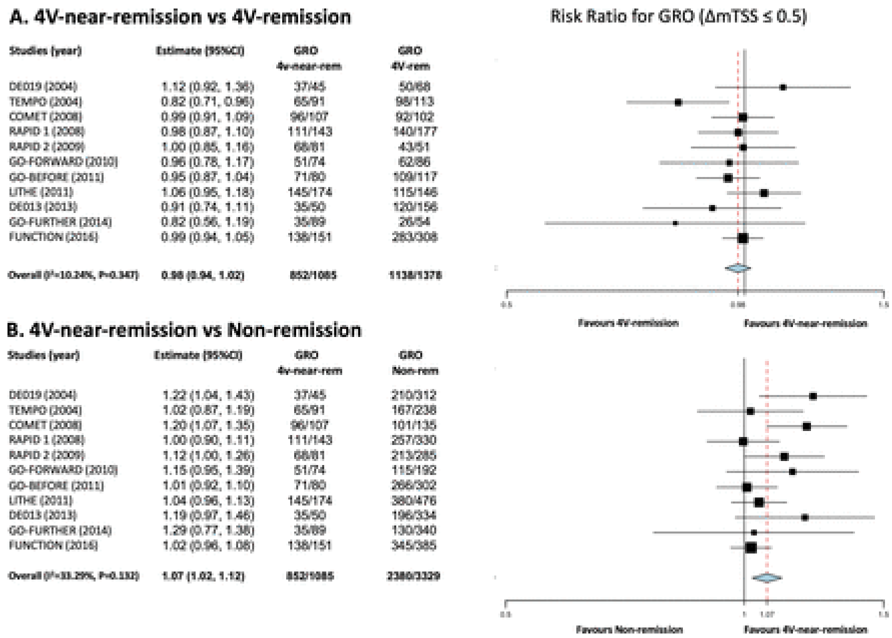

Figure 3 – Meta-analyses of risk ratio of obtaining good radiographic outcome (ΔmTSS≤0.5 units); 4V-near-remission vs 4V-remission and vs Non-remission.

Legend: 4V-remission = SJC28, TJC28, CRP (in mg/dl), and PGA (0-10), all ≤1; 4V-near-remission= SJC28, TJC28, CRP (in mg/dl) ≤1 and PGA (0-10)>1; Non-remission = SJC28 >1 and/or TJC28>1 and/or CRP (in mg/dl)>1, irrespective of PGA value; at 6 OR 12 months of follow-up in all cases; ΔmTSS = change in the modified Total Sharp Score during the second year of follow-up.

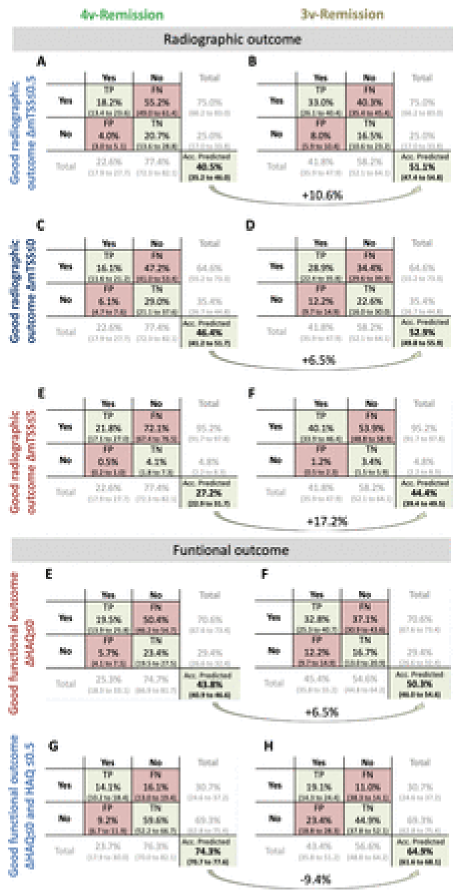

Figure 4 - Pooled meta-analytic prediction accuracy of 4V- and 3V-remission status for the good radiographic and functional outcomes

Footnote: The sum of the meta-analytic percentages of TP, FN, FP, and TN is slightly less than 100% due to error estimation when multi-category (k>2) prevalence is estimated.[35] All meta-analyses used double arcsine transformation as the preferred method to correct this situation.[35]

The panels from A to F include 5,792 analysed patients (11 RCTs), E and F include 3,904 (8 RCTs), and G and H 5,262 analysed patients (11 RCTs).

Legend: 4V-remission = SJC28, TJC28, CRP (in mg/dl), and PGA (0-10), all ≤1; 3V-remission= SJC28, TJC28, CRP (in mg/dl) ≤1; ΔmTSS = change in the modified Total Sharp Score from 12 months to 24 months. TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative; Accurately predicted = TP + TN. Between brackets is the pooled 95% confidence interval.

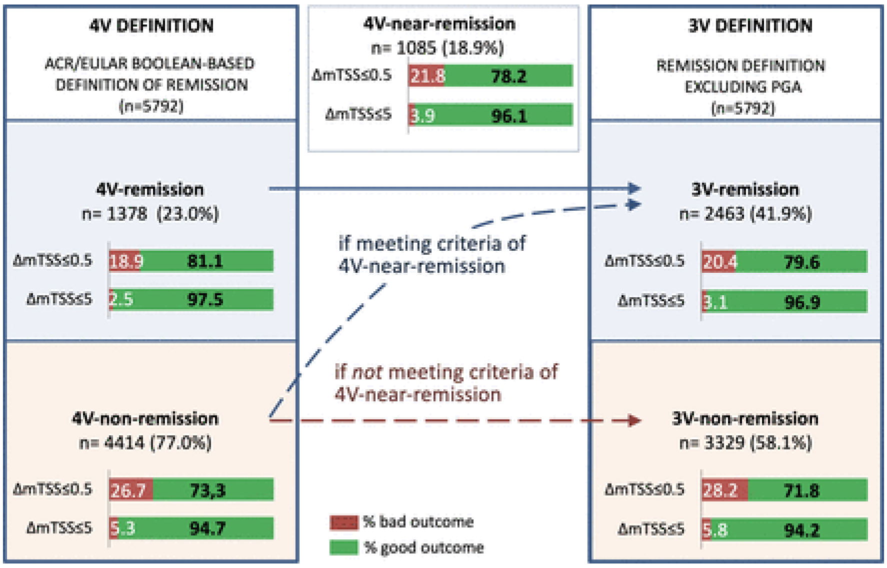

Figure 5 – Reclassification of remission status and respective radiographic outcomes (n=5,792). Percentages were calculated through meta-analyses.

Footnote: Excluding PGA from the remission of remission (3V-remission) almost duplicated the percentage of patients in remission but showed only a slight increase in the rate of bad outcome when compared with 4V-remission. The radiographic outcome in the group of patients who had no overt signs of inflammation but who presented with high PGA (4V-near-remission) was also not statistically different from patient in 4V-remission.

Legend: 4V-remission = SJC28, TJC28, CRP (in mg/dl), and PGA (0-10), all ≤1; 4V-near-remission= SJC28, TJC28, CRP (in mg/dl) ≤1 and PGA (0-10)>1; Non-remission = SJC28 >1 AND/OR TJC28>1 AND/OR CRP (in mg/dl)>1, irrespective of PGA value; 3V-remission= SJC28, TJC28, CRP (in mg/dl) ≤1; All definitions as observed at 6 OR 12 months. ΔmTSS = change in the modified Total Sharp Score during the second year of follow-up.

Note: Confidence intervals and I^2 statistics of pooled radiographic outcomes can be found in Supplementary Table S4.

Supplementary file for Ferreira RJO et al. "Revisiting the use of remission criteria for rheumatoid arthritis by excluding patient global assessment: an individual data meta-analysis of 5792 patients"

Supplementary Table S1 – Overview of data requested, obtained and used, with reasons for non-inclusion

Platform used	Trials requested	Data obtained	Data used	Reasons for not being provided (if known) or for not being used
Abbvie's own platform[1]	PREMIER - NCT00195663	yes	yes	
	DE019 - NCT00195702	yes	yes	
Pfizer's own platform[2]	TEAR - NCT00259610	no	no	The Sponsor was University of Alabama at Birmingham. Data not available.
	COMET - NCT00195494	yes	yes	
	CAMEO - NCT00654368	no	no	The Sponsor was Amgen. Data not available.
	PRIZE - NCT00913458	no	no	It was a tapering trial (dose reduction)
	TEMPO - NCT00393471	yes	yes	
	PRESERVE - NCT00565409	yes	no	It was a tapering trial (dose reduction)
	ERA - NCT00356590	no	no	The Sponsor was Amgen. Data not available.
	ORAL START - NCT01039688	no	no	Delays in the process, which was eventually cancelled.
	ORAL SCAN - NCT00847613	no	no	Delays in the process, which was eventually cancelled.
The YODA project[3] – data from Johnson & Johnson	GO BEFORE - NCT00264537	yes	yes	
	GO FORWARD - NCT00264550	yes	yes	
	GO FURTHER - NCT00973479	yes	yes	
	ATTRACT - NCT00269867	yes	no	The radiographic data was provided too late (after change in the platform has occurred)
ClinicalStudy-DataRequest[4] – data from Roche	ASPIRE - NCT00236028	no	no	The Sponsor was Centocor. Data not available in the Sharing Data Platform
	LITHE - NCT00106535	yes	yes	
	FUNCTION - NCT01007435	yes	yes	
	BREVACTA - NCT01232569	yes	no	Only 1y data was provided
	ACT-RAY - NCT00810199	no	no	Tested discontinuation of therapy in the 2nd year of trial.
	SAMURAI - NCT00144508	yes	no	Only 1y data was provided
	SURPRISE - NCT01120366	no	no	The Sponsor was "SURPRISE Study Group". Data not available in the Sharing Data Platform. Linked with ACT-RAY study.

Supplementary file for Ferreira RJO et al. "Revisiting the use of remission criteria for rheumatoid arthritis by excluding patient global assessment: an individual data meta-analysis of 5792 patients"

ClinicalStudy-DataRequest[5]– data from UCB	REFLEX - NCT00468546/ NCT02097745	yes	no	The protocol did include week 104 assessment of radiographic score. Also, very difficult to match visit date with visit number.
	IMAGE - NCT00299104	yes	No	Only safety data was provided
	RAPID 1 - NCT00152386	yes	yes	
	RAPID 2 - NCT00160602/ NCT00175877	yes	yes	
	C-OPERA - NCT01451203	no	no	The Sponsor was Astellas. Data not available in the Sharing Data Platform.

1 – https://www.abbvie.com/our-science/clinical-trials/clinical-trials-data-and-information-sharing/data-and-information-sharing-with-qualified-researchers.html - meanwhile transitioned to Vivli (https://vivli.org/)

2 - https://www.pfizer.com/science/clinical-trials/trial-data-and-results/data-requests - meanwhile transitioned to Vivli (https://vivli.org/)

3 - https://yoda.yale.edu/ meanwhile transitioned to Microsoft Online.

4 - https://www.clinicalstudydatarequest.com/Default.aspx

5 - https://www.clinicalstudydatarequest.com/Default.aspx, meanwhile transitioned to Vivli (https://vivli.org/)

Supplementary file for Ferreira RJO et al. "Revisiting the use of remission criteria for rheumatoid arthritis by excluding patient global assessment: an individual data meta-analysis of 5792 patients"

Supplementary Figure S1 – Risk of bias assessment of the 11 RCTs

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
DE019	+	+	+	+	+	+	+
TEMPO	+	+	+	+	+	+	+
COMET	+	+	+	+	+	+	+
RAPID 1	+	+	+	+	+	+	+
RAPID 2	+	+	+	+	+	+	+
GO-FORWARD	+	+	+	+	+	+	+
GO-BEFORE	+	+	+	+	+	+	+
LITHE	+	+	+	+	+	+	+
DE013	+	+	+	+	+	+	+
GO-FURTHER	+	+	+	+	+	+	+
FUNCTION	+	+	+	+	+	+	+

Footnote: To appraise the quality of the trials we assessed different papers[1-17] resulting from the same trial (e.g. reporting outcomes for different timepoints). We also assessed the full protocols provided by the sponsors and requested additional information to individual authors when needed.

References:

1 Keystone EC, Kavanaugh AF, Sharp JT, *et al.* Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum* 2004;50:1400-11.

2 Keystone EC, van der Heijde D, Kavanaugh A*, et al.* Clinical, functional, and radiographic benefits of longterm adalimumab plus methotrexate: final 10-year data in longstanding rheumatoid arthritis. *J Rheumatol* 2013;40:1487-97.

3 Breedveld FC, Weisman MH, Kavanaugh AF*, et al.* The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis Rheum* 2006;54:26-37.
4 Klareskog L, van der Heijde D, de Jager JP*, 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. *Lancet* 2004;363:675-81.
5 van der Heijde D, Klareskog L, Rodriguez-Valverde V*, et al.* Comparison of etanercept and methotrexate, alone and combined, in the treatment of rheumatoid arthritis: two-year clinical and radiographic results from the TEMPO study, a double-blind, randomized trial. *Arthritis Rheum* 2006;54:1063-74.
6 Emery P, Breedveld FC, Hall S*, et al.* Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial. *Lancet* 2008;372:375-82.
7 Keystone E, Heijde D, Mason D, Jr.*, et al.* Certolizumab pegol plus methotrexate is significantly more effective than placebo plus methotrexate in active rheumatoid arthritis: findings of a fifty-two-week, phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. *Arthritis Rheum* 2008;58:3319-29.
8 Keystone EC, Combe B, Smolen J*, et al.* Sustained efficacy of certolizumab pegol added to methotrexate in the treatment of rheumatoid arthritis: 2-year results from the RAPID 1 trial. *Rheumatology (Oxford)* 2012;51:1628-38.
9 Smolen J, Landewe RB, Mease P*, et al.* Efficacy and safety of certolizumab pegol plus methotrexate in active rheumatoid arthritis: the RAPID 2 study. A randomised controlled trial. *Ann Rheum Dis* 2009;68:797-804.
10 Keystone E, Genovese MC, Klareskog L*, et al.* Golimumab in patients with active rheumatoid arthritis despite methotrexate therapy: 52-week results of the GO-FORWARD study. *Ann Rheum Dis* 2010;69:1129-35.
11 Keystone EC, Genovese MC, Hall S*, et al.* Golimumab in patients with active rheumatoid arthritis despite methotrexate therapy: results through 2 years of the GO-FORWARD study extension. *J Rheumatol* 2013;40:1097-103.
12 Emery P, Fleischmann RM, Doyle MK*, et al.* Golimumab, a human anti-tumor necrosis factor monoclonal antibody, injected subcutaneously every 4 weeks in patients with active rheumatoid arthritis who had never taken methotrexate: 1-year and 2-year clinical, radiologic, and physical function findings of a phase III, multicenter, randomized, double-blind, placebo-controlled study. *Arthritis Care Res (Hoboken)* 2013;65:1732-42.
13 Bingham CO, 3rd, Mendelsohn AM, Kim L*, et al.* Maintenance of Clinical and Radiographic Benefit With Intravenous Golimumab Therapy in Patients With Active Rheumatoid Arthritis Despite Methotrexate Therapy: Week-112 Efficacy and Safety Results of the Open-Label Long-Term Extension of a Phase III, Double-Blind, Randomized, Placebo-Controlled Trial. *Arthritis Care Res (Hoboken)* 2015;67:1627-36.
14 Weinblatt ME, Westhovens R, Mendelsohn AM*, et al.* Radiographic benefit and maintenance of clinical benefit with intravenous golimumab therapy in patients with active rheumatoid arthritis despite methotrexate therapy: results up to 1 year of the phase 3, randomised, multicentre, double blind, placebo controlled GO-FURTHER trial. *Ann Rheum Dis* 2014;73:2152-9.
15 Kremer JM, Blanco R, Brzosko M*, et al.* Tocilizumab inhibits structural joint damage in rheumatoid arthritis patients with inadequate responses to methotrexate: results from the double-blind treatment phase of a randomized placebo-controlled trial of tocilizumab safety and prevention of structural joint damage at one year. *Arthritis Rheum* 2011;63:609-21.
16 Fleischmann RM, Halland AM, Brzosko M*, et al.* Tocilizumab inhibits structural joint damage and improves physical function in patients with rheumatoid arthritis and inadequate responses to methotrexate: LITHE study 2-year results. *J Rheumatol* 2013;40:113-26.
17 Burmester GR, Rigby WF, van Vollenhoven RF*, et al.* Tocilizumab combination therapy or monotherapy or methotrexate monotherapy in methotrexate-naive patients with early rheumatoid arthritis: 2-year clinical and radiographic results from the randomised, placebo-controlled FUNCTION trial. *Ann Rheum Dis* 2017;76:1279-84.

Supplementary file for Ferreira RJO et al. "Revisiting the use of remission criteria for rheumatoid arthritis by excluding patient global assessment: an individual data meta-analysis of 5792 patients"

Supplementary Table S2 - Baseline characteristics of the population samples of the studies (all placebo-controlled)

Trial name (Year of publication)	DE019 (2004)	TEMPO (2004)	COMET (2008)	RAPID 1 (2008)	RAPID 2 (2009)	GO FORWARD (2010)	GO BEFORE (2011)	LITHE (2011)	DE013 (2013)	GO FURTHER (2014)	FUNCTION (2016)
Biologic agent	Adalimumab	Etanercept	Etanercept	Certolizumab	Certolizumab	Golimumab	Golimumab	Tocilizumab	Adalimumab	Golimumab	Tocilizumab
Inclusion criteria	MTX-IR	csDMARD-IR [a]	MTX-naive	MTX-IR	MTX-IR	MTX-IR	MTX-naive	MTX-IR	MTX-naive	MTX-IR	MTX-IR
No. patients randomized	619	686	542	982	619	444	637	1196	799	592	1162
No patients available for this IPD study	619	684	542	857	582	444	637	1196	799	592	1162
No.(%) patients with predictors and outcome at 2y	425 (68.6)	442 (64.6)	344 (63.5)	650 (75.8)	417 (71.6)	352 (79.3)	499 (78.3)	796 (66.6)	540 (67.6)	483 (81.6)	844 (60.3)
Demographics [b]											
Female (%)	74.8	75.6	73.5	82.3	80.8	81.2	83.6	83.2	73.7	81.4	79.6
Mean age (yrs)	55.4 (12.0)	51.9 (12.5)	51.7 (13.7)	51.8 (11.5)	50.8 (11.5)	50.2 (11.0)	49.9 (12.0)	51.9 (11.9)	52.2 (13.4)	51.4 (11.8)	49.9 (12.9)
Mean RA duration (yrs)	10.8 (9.0)	6.3 (5.0)	7.4 (5.4)	6.3 (4.3)	6.0 (4.1)	6.4 (6.5) [b]	2.5 (3.8) [b]	9.5 (7.8)	**0.7 (0.8)**	4.3 (4.9) [b]	**0.5 (0.5)**
RF positive (%)	84.9	66.1	95.8	83.5	77.0	83.0	81.0	82.0	85.4	90.9	90.6
Disease activity measures											
Mean DAS28CRP3v	4.9 (0.7)	5.3 (0.8)	4.9 (0.9)	5.3 (0.7)	5.2 (0.7)	4.6 (0.8)	4.7 (0.9)	4.8 (1.0)	5.3 (0.8)	4.9 (0.8)	4.9 (0.9)
Mean CRP (mg/dl)	1.8 (1.9)	2.7 (3.1)	3.6 (3.6)	2.5 (2.7)	2.4 (2.5)	1.7 (2.1)	2.4 (3.0)	2.2 (2.5)	3.8 (3.9)	2.5 (2.5)	2.5 (2.9)
Mean TJC28	14.5 (6.4)	18.0 (6.7)	13.8 (7.1)	17.7 (6.1)	17.9 (6.4)	13.5 (7.3)	14.2 (7.3)	14.7 (7.6)	16.8 (6.3)	14.8 (6.4)	15.8 (7.3)
Mean SJC28	13.2 (5.5)	15.0 (5.8)	12.0 (6.2)	14.8 (5.4)	14.3 (5.6)	9.8 (5.6)	10.4 (6.0)	11.5 (6.2)	14.4 (5.7)	10.9 (5.2)	11.7 (6.0)
Mean PGA (cm)	5.2 (2.2)	6.9 (1.7) [c]	6.5 (1.9) [c]	6.3 (1.9)	6.0 (2.1)	5.4 (2.4)	6.0 (2.3)	5.7 (2.4)	6.4 (2.4)	6.5 (1.8)	6.5 (2.2)
Mean (PhGA) (cm)	6.1 (1.7)	6.6 (1.5) [c]	6.5 (1.5) [c]	6.3 (1.5)	6.4 (1.4)	5.7 (1.7)	6.2 (1.7)	5.7 (2.2)	6.5 (1.8)	6.2 (1.6)	6.3 (1.8)
Functional status											
Mean baseline score	1.4 (1.4)	1.7 (0.6)	1.6 (1.6)	1.6 (0.6)	1.6 (0.6)	1.4 (0.8)	1.5 (0.6)	1.4 (0.6)	1.5 (0.6)	1.6 (0.7)	1.5 (0.7)
Radiographic scores [d]											
Mean baseline score	69.0 (55.8)	35.7 (49.7)	**8.4 (16.2)**	48.2 (57.1)	34.2 (46.3)	34.5 (48.9)	15.8 (28.9)	30.3 (30.9)	19.9 (21.0)	49.6 (55.7)	**5.9 (14.5)**

a. Other than Methotrexate

b. There was no imputation of missing data. The most frequent missing result was disease duration (11.9%; except for the three Golimumab trials for which this variable was missing in >50% of patients) followed by rheumatoid factor status (2.1%).

c. Assessed with numeric rating scale (0 to 10) and not with visual analogue scale (0 to 10cm)

d. All trials used Sharp van der Heijde mTSS (0 to 448) except in the LITHE trial, in which Genant mTSS (0 to 202) was used instead.

Ferreira RJO, *et al. Ann Rheum Dis* 2020;0:1–11. doi: 10.1136/annrheumdis-2020-217171

Supplementary file for Ferreira RJO et al. "Revisiting the use of remission criteria for rheumatoid arthritis by excluding patient global assessment: an individual data meta-analysis of 5792 patients"

Legend: MTX - Methotrexate, IR- Insufficient responder, IPD - Individual patient data, RA, rheumatoid arthritis, RF, Rheumatoid Factor, DAS28CRP3v, Disease Activity Score with 28-joint counts, using c-Reactive protein and 3 variables; CRP, C-Reactive Protein, TJC28, Tender 28-joint counts; SJC28, Swollen 28-joint counts; PGA, Patient Global Assessment of disease activity; PhGA, Physician Global Assessment of disease activity.

Supplementary file for Ferreira RJO et al. "Revisiting the use of remission criteria for rheumatoid arthritis by excluding patient global assessment: an individual data meta-analysis of 5792 patients"

Supplementary Table S3 - Comparison (through meta-analysis) between patients included and excluded from analyses.

	Included (n=5,792) [a]		Excluded (n=2,322)		Difference (95%CI)[b]
	n	Estimate (95%CI)	n	Estimate (95%CI)	
Baselines features					
Female (%)	5,792	79% (77 to 81)	2,322	79% (76 to 81)	0% (-2.2 to 2.6)
Age (yrs)	5,792	51.4 (50.5 to 52.4)	2,320	52.8 (51.6 to 54.0)	**1.3 (0.5 to 2.0)**
Disease duration (yrs)	5,102	5.6 (4.5 to 6.7)	2,086	5.6 (4.5 to 6.7)	0.03 (-0.18 to 0.23)
RF positive (%)	5,666	84% (80 to 88)	2,212	82% (76 to 87)	4% (-0.3 to 7.5)
DAS28CRP3V	5,781	4.98 (4.83 to 5.13)	2,260	5.03 (4.88 to 5.18)	0.04 (-0.01 to 0.08)
CRP (mg/dl)	5,781	2.54 (2.23 to 2.85)	2,262	2.73 (2.30 to 3.16)	0.11 (-0.03 to 0.25)
TJC28	5,792	15.6 (14.6 to 16.6)	2,268	16.0 (15.1 to 16.9)	0.29 (-0.03 to 0.63)
SJC28	5,792	12.6 (11.5 to 13.6)	2,268	12.7 (11.7 to 13.8)	0.11 (-0.17 to 0.39)
PGA (cm)	5,775	6.1 (5.8 to 6.4)	2,252	6.5 (6.2 to 7.8)	**0.34 (0.21 to 0.47)**
PhGA (cm)	5,771	6.2 (6.1 to 6.4)	2,257	6.4 (6.2 to 6.6)	**0.16 (0.07 to 0.25)**
mTSS	5,792	30.0 (20.8 to 39.2)	1,451[c]	29.6 (19.8 to 39.5)	-0.03 (-1.47 to 1.41)
HAQ-DI [d]	4,392	1.54 (1.46 to 1.61)	1,897	1.66 (1.59 to 1.67)	**0.12 (0.09 to 0.16)**
Randomization arm % (95% CI) [e]					
Placebo	349	69% (60 to 77)	161	31% (23 to 40)	
MTX mono	935	64% (57 to 70)	524	36% (30 to 43)	P=0.68[f]
bDMARD mono	1,684	70% (65 to 74)	683	30% (26 to 35)	
DMARD and MTX	1,543	71% (69 to 74)	628	28% (25 to 31)	
Remission at 6 OR 12 months, % (95% CI)					
4V-remission	5,792	23 (18 to 28)	2,017	11 (9 to 13)	**-16 (-21 to -11.3)**
4V-near-remission	5,792	19 (15 to 22)	2,017	12 (9 to 16)	**-8 (-6 to -11)**
Non-remission	5,792	58 (52 to 64)	2,017	76 (72 to 81)	**20 (16 to 24)**

In **bold** are presented the differences of which the 95%CI do not include zero, in general indicating statistical significance

Legend: 95%CI, 95% Confidence Interval, HAQ-DI, Health Assessment Questionnaire – Disability Index, mTSS, modified Total Sharp Score, RF, Rheumatoid Factor, DAS28CRP3v, Disease Activity Score with 28-joint counts, using c-Reactive protein and 3 variables; CRP, C-Reactive Protein, TJC28, Tender 28-joint counts; SJC28, Swollen 28-joint counts; PGA, Patient Global Assessment of disease activity; PhGA, Physician Global Assessment of disease activity. 4V-remission = SJC28, TJC28, CRP (in mg/dl), and PGA (0-10), all ≤1; 4V-near-remission= SJC28, TJC28, CRP (in mg/dl) ≤1 and PGA (0-10)>1; Non-remission = SJC28 >1 OR TJC28>1 OR CRP (in mg/dl)>1, irrespective of PGA value; at 6 OR 12 months of follow-up in all cases.

a. There was no imputation of missing data. The most frequently missing result was disease duration (11.9%), followed by rheumatoid factor status (2.1%).

b. The difference in percentages/mean may not match exactly with (raw) arithmetic difference because all estimates were determined using meta-analyses with double arcsine transformation.(1)

c. The number of missing patients was not possible to be determined in GO-FURTHER trial

d. Not possible to be determined in the three golimumab trials due to changes that occurred in the research environment and statistical software available since the initial data analyses.

e. Not possible to be determined in the three golimumab trials due to changes that occurred in the research environment and statistical software available since the initial data analyses; Placebo arm in 3 trials, MTX arm in 5 trials, bDMARD mono in 6 trials, bDMARD and MTX arm in 6 trials. When tested, the different dosages of bDMARD were considered in the same group

f. The distributions of included and excluded patients per randomization treatment arm are not statistically significant, according to Pearson's Chi Square test

Supplementary file for Ferreira RJO et al. "Revisiting the use of remission criteria for rheumatoid arthritis by excluding patient global assessment: an individual data meta-analysis of 5792 patients"

Supplementary Table S4 - Pooled meta-analytic frequency of radiographic outcomes (with 95%CI) and heterogeneity statistics for each remission definition (n=5,792). This table provides complementary information to Figure 5 in the article.

∆mTSS cut-off	Remission Definition	% Good Outcome				% Bad Outcome			
		Pooled	95%CI Lower	95%CI Higher	I^2	Pooled	95%CI Lower	95%CI Higher	I^2
≤0.5	4V-rem.	81.1	74.4	86.9	88.6	18.9	13.1	25.6	88.6
	Non-4V-rem.	73.3	63.9	81.8	97.7	26.7	18.2	36.1	97.9
	4V-near-rem.	78.2	69.5	85.8	90.8	21.8	14.2	30.5	90.8
	3V-rem.	79.6	72.2	86.1	94.7	20.4	13.9	27.8	94.7
	Non-3V-rem.	71.8	62.1	80.5	97.2	28.2	19.5	37.9	97.2
≤5	4V-rem.	97.5	95.4	98.9	76.2	2.5	1.1	4.6	76.2
	Non-4V-rem.	94.7	90.8	97.6	96.2	5.3	2.4	9.2	92.2
	4V-near-rem.	96.1	92.5	98.5	85.0	3.9	1.5	7.5	85.0
	3V-rem.	96.9	94.2	98.8	90.7	3.1	1.2	5.8	90.7
	Non-3V-rem.	94.2	90.2	97.2	94.8	5.8	2.8	9.8	94.8

Legend: rem.: remission. 4V-remission = SJC28, TJC28, CRP (in mg/dl), and PGA (0-10), all ≤1; 4V-near-remission= SJC28, TJC28, CRP (in mg/dl) ≤1 and PGA (0-10)>1; Non-remission = SJC28 >1 AND/OR TJC28>1 AND/OR CRP (in mg/dl)>1, irrespective of PGA value; 3V-remission= SJC28, TJC28, CRP (in mg/dl) ≤1; All definitions as observed at 6 OR 12 months. ∆mTSS = change in the modified Total Sharp Score during the second year of follow-up.

Supplementary Table S5 - Meta-analyses of the adjusted[a] odds ratios to compare the predictive value of good radiographic and good functional outcomes between patients in 4V-remission and in 4V-near-remission status (at 6 OR 12 months)

Good Radiographic Outcome (from 12 to 24 months)	No. studies (participants)	4V-near-remission	4V-remission	I^2
		(Reference)	OR (95% CI)	
ΔmTSS ≤ 0.5	11 (5,653)	1.00	0.97 (0.69 to 1.23)	0%
ΔmTSS ≤ 0	11 (5,653)	1.00	1.06 (0.81 to 1.30)	0%
ΔmTSS ≤ 5	7 (3,109)[b]	1.00	0.85 (0.02 to 2.19)	0%
ΔHAQ-DI≤ 0	8 (3,696)	1.00	1.28 (0.94 to 2.05)	0%
ΔHAQ-DI ≤0 AND HAQ-DI ≤0.5	11 (5,049)	1.00	3.47 (2.36 to 4.91)	33%

a. Model adjusted to age at baseline, gender, rheumatoid factor, disease duration (except for GOBEFORE, GOFORWARD, and GOFURTHER trials as these had missing data>50%) radiographic damage at baseline, and treatment arm were included as possible confounders.
b. Without GOBEFORE, LITHE, FUNCTION, and RAPID2 trials due to invalid data obtained from logistic regressions.

Legend: 4V-remission = SJC28, TJC28, CRP (in mg/dl), and PGA (0-10), all ≤1; 4V-near-remission= SJC28, TJC28, CRP (in mg/dl) ≤1 and PGA (0-10)>1; All definitions as observed at 6 OR 12 months. ΔmTSS = change in the modified Total Sharp Score during the second year of follow-up; OR= Odds Ratio.

Supplementary file for Ferreira RJO et al. "Revisiting the use of remission criteria for rheumatoid arthritis by excluding patient global assessment: an individual data meta-analysis of 5792 patients"

Supplementary Table S6 - Meta-analyses of the adjusted[a] odds ratios to descriptively compare the predictive value of good outcomes between patients in 4V-remission and in 3V-remission status (6 OR 12 months)

Definition of Good Outcome (from 12 to 24 months)	No. studies (participants)	4V-remission (Reference)	Non-remission OR (95% CI)	I^2	3V-remission (Reference)	Non-remission OR (95% CI)	I^2
ΔmTSS ≤ 0.5	11 (5,653)	1.00	0.66 (0.50 to 0.85)	34%	1.00	0.64 (0.54 to 0.77)	0%
ΔmTSS ≤ 0	11 (5,653)	1.00	0.68 (0.54 to 0.84)	40%	1.00	0.73 (0.64 to 0.83)	0%
ΔmTSS ≤ 5	8 (3,607)[b]	1.00	0.22 (0.05 to 0.44)	0%	1.00	0.79 (0.47 to 1.12)	0%
ΔHAQ-DI≤ 0	8 (3,696)	1.00	0.63 (0.51 to 0.76)	0%	1.00	0.72 (0.60 to 0.85)	0%
ΔHAQ-DI≤ 0 AND HAQ-DI≤ 0.5	11 (5,049)	1.00	0.17 (0.13 to 0.22)	51%	1.00	0.30 (0.24 to 0.37)	40%

a. adjusted analysis to: age at baseline, gender, rheumatoid factor, disease duration (except for GOBEFORE, GOFORWARD, and GOFURTHER trials as these had missing data>50%) radiographic damage at baseline, and treatment arm were included as possible confounders.

b – Without LITHE, FUNCTION, RAPID2 trials due to invalid data obtained from logistic regressions.

Legend: 4V-remission = SJC28, TJC28, CRP (in mg/dl), and PGA (0-10), all ≤1; 3V-remission= SJC28, TJC28, CRP (in mg/dl) ≤1; Non-remission = SJC28 >1 AND/OR TJC28>1 AND/OR CRP (in mg/dl)>1, irrespective of PGA value; All definitions as observed at 6 OR 12 months. ΔmTSS = change in the modified Total Sharp Score during the second year of follow-up; OR = Odds Ratio.

Supplementary file for Ferreira RJO et al. "Revisiting the use of remission criteria for rheumatoid arthritis by excluding patient global assessment: an individual data meta-analysis of 5792 patients"

Supplementary Table S7. Proportion of patients in 3v-remission who have radiographic damage progression ≥0.5 and ≥5 according to PGA[1] score ≤1 OR >1.

Trial	N total	Patients with ΔmTSS>**0.5** AND				p-value[2]	Patients with ΔmTSS>**5** AND				p-value[2]
		PGA**≤1**		PGA**>1**			PGA≤1		PGA>1		
		n	%	n	%		n	%	n	%	
DE019	114	13/43	30.2	14/71	19.7	0.26	1/43	2.3	5/71	7.0	0.40
TEMPO	204	8/63	12.7	33/141	23.4	0.09	2/63	3.2	5/141	3.6	1.0
COMET	200	3/45	6.7	18/155	11.6	0.42	0/45	0	6/155	3.9	0.34
RAPID1	316	26/128	20.3	41/188	21.8	0.78	2/128	1.6	4/188	2.1	1.0
RAPID2	129	4/29	13.8	17/100	17.0	0.78	0/29	0	1/100	1.0	1.0
LITHE	313	18/92	19.6	41/221	18.6	0.87	1/92	1.1	2/221	0.9	1.0
DE013	204	23/122	18.8	26/82	31.7	**0.04**	3/122	2.5	5/82	6.1	0.27
FUNCTION	457	11/179	6.1	27/278	9.7	0.22	0/179	0	0/278	0	na
Pooled prevalence (95%CI)		15.2% (9.9 to 21.4)		18.4% (13.8 to 23.5)		--	1.3% (0.6 to 2.3)		2.3% (1.0 to 4.3%)		--

1. Mean values at both 6 and 12 months.

2. Using Fisher's Exact test (2X2 contingency tables)

NOTE: The % presented in the grey columns complement with the percentage of patients who did not progress in the same sub-group of PGA values. For instance, for the DE019 trial: 13 out of the 43 (30.2%) who had a PGA≤1 presented a damage progression >0.5 units and, thus, the remaining 30 patients (69.8%) presented a damage progression ≤0.5.