

CLINICAL SCIENCE

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 ^{1,2}, Paco M J Welsing,³ Johannes W G Jacobs ³,
Laure Gossec ^{4,5}, Mwidimi Ndosi ⁶, Pedro M Machado ^{7,8,9},
Désirée van der Heijde ¹⁰, Jose A P Da Silva ^{1,11}

Handling editor Josef S Smolen

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/annrheumdis-2020-217171>).

For numbered affiliations see end of article.

Correspondence to

Professor Jose A P Da Silva, Rheumatology, Centro Hospitalar e Universitario de Coimbra EPE, 3000-076 Coimbra, Portugal; jdasilva@ci.uc.pt

RJOF and PMJW are joint first authors.

Received 17 February 2020
Revised 29 August 2020
Accepted 3 September 2020
Published Online First 6 October 2020



► <http://dx.doi.org/10.1136/annrheumdis-2020-218802>



© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Ferreira RJO, Welsing PMJ, Jacobs JWG, et al. *Ann Rheum Dis* 2021;**80**:293–303.

ABSTRACT

Objectives To determine the impact of excluding patient global assessment (PGA) from the American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) Boolean remission criteria, on prediction of radiographic and functional outcome of rheumatoid arthritis (RA).

Methods Meta-analyses using individual patient data from randomised controlled trials 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); (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 (ie, change in modified total Sharp score (Δ mTSS) ≤ 0.5 units and ≤ 0.0 Health Assessment Questionnaire–Disability Index 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=5792) from 11 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% to 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 Δ mTSS > 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% to 71%) was significantly lower than in 4V-remission (78%, 74% to 81%) and similar to non-remission (69%, 66% to 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.

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 American College of Rheumatology/European League Against Rheumatism Boolean-based definition of remission ('4V-remission') versus those who attained only three ('3V-remission'), that is, 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 rheumatoid arthritis (5792 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 a change in modified total Sharp score > 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 on 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 Rheumatism (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, 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 mentioned previously, 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/OMERACT paper, removing PGA from the Boolean-based remission significantly diminishes its ability to predict good radiographic and functional outcome.⁴ A systematic literature review 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 cohort¹³ and from clinical trials,²² compared the prediction of good radiographic outcome by '4V-remission' versus '3V-remission' (without PGA) achieved in patients with RA: 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-remission 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 randomised controlled trials (RCTs) selected through a systematic literature review. The study protocol was registered in PROSPERO with the number CRD42017057099²³ 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 (RJOF 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 (ie, 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:

1. Δ mTSS ≤ 5 units, a higher, frequently used rate (sometimes referred to as clinically non-relevant radiographic progression);
2. Δ mTSS ≤ 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 (ie, between month 12 and month 24), defined as no worsening, that is, 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

| Disease activity | SJC28 TJC28 CRP } all ≤ 1 | SJC28 TJC28 CRP } all ≤ 1 | SJC28 TJC28 CRP } at least one > 1 |
|------------------|--------------------------------------|--------------------------------------|--|
| Disease Impact | PGA ≤ 1 | PGA > 1 | PGA = 0-10 |
| | ⇓ | ⇓ | ⇓ |
| 4V concept | 4V-Remission | 4V-Near-remission | 4V Non-remission |
| | ⇓ | ⇓ | ⇓ |
| 3V concept | 3V-Remission | | 3V Non-remission |

Figure 1 Definitions of remission tested in the study. *Legend:* CRP, C reactive protein, mg/dL; PGA, patient global assessment, range 0–10=worst; SJC28, swollen 28-joint count, range 0–28; TJC28, tender 28-joint count, range 0–28. *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.

the methodology adopted by the ACR/EULAR committee,⁴ as follows:

1. ACR/EULAR Boolean-based remission,⁴ also designated in this study as ‘4V-Remission’ (ie, TJC28 ≤ 1 , SJC28 ≤ 1 , CRP ≤ 1 mg/dL and PGA $\leq 1/10$).
2. ‘4V-near-remission’,^{11 14} defined as TJC28 ≤ 1 , SJC28 ≤ 1 , CRP ≤ 1 mg/dL and PGA > 1 .
3. ‘Non-remission’ defined as TJC28 > 1 and/or SJC28 > 1 and/or CRP > 1 mg/dL, irrespective of PGA value.

The three definitions are mutually exclusive, that is, each patient was categorised in one group only.

4. ‘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), that is, remission and GRO; true negative (TN), that is, non-remission and not-GRO; false negative (FN), that is, non-remission and GRO; and false positive (FP), that is, 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-remission and 3V-remission. Missing data were 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 with 4V-remission and with 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 (Δ mTSS ≤ 0.5) between 4V-near-remission and 4V-remission, mutually exclusive states, and then pooled these differences with the random-effects 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 ratio 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 with 4V-remission and with 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 SE 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 online supplemental 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 standardised 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; 1937 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.

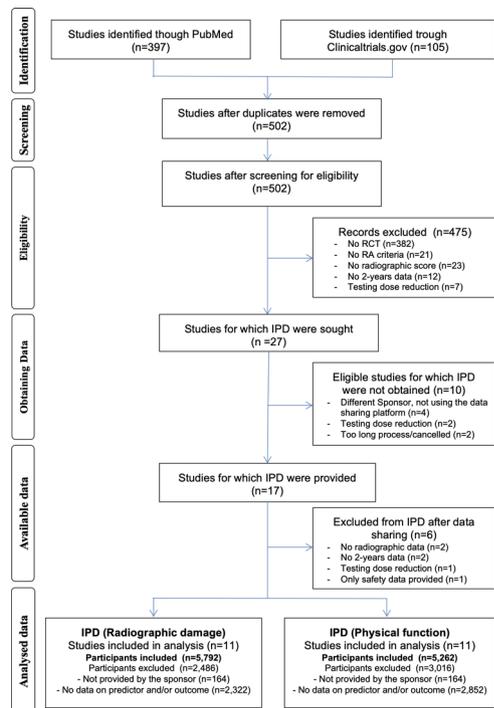


Figure 2 Flowchart with the process of study identification and data access. IPD, individual patient data; RA, rheumatoid arthritis; RCT, randomised controlled trial.

Likelihood of reaching good radiographic and functional outcomes with 4V-remission compared with 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 synthesise these measures. Similar procedures were performed regarding GFO.

All meta-analyses were performed with the OpenMeta[Analyst] software,³⁴ using the DerSimonian-Laird random-effects method³⁵ and the arcsine-transformed proportion.³⁶ 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 online supplemental table S1. The critical appraisal results for each of the 11 RCTs are summarised in online supplemental figure S1 (low risk of bias in all items assessed for all the trials). We had access to data from 100% of the randomised patients in 9 out of the 11 trials and from 93% of patients in the remaining two, resulting in a total sample of 8114 patients. Most trials tested anti-TNF α therapies (n=9), and included patients with insufficient response to methotrexate (n=7) and with established disease (>2 years) (n=9)—online supplemental table S2. The mean (SD) DAS28CRP_v 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) (online supplemental table S2).

Altogether, 2322 patients (29%) were excluded from the final analyses (online supplemental 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 5792 (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, it 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 (Δ mTSS ≤ 0.5) and by 94.6% (92.9% to 96.4%) using Δ mTSS ≤ 5 (table 1). Good functional outcome, which could only be assessed in eight RCTs (3904 patients), was observed in 70.6% (66.7% to 73.5%) of all patients using the elected cut-off (Δ HAQ-DI ≤ 0.0), and by 31.1% (24.9 to 37.2%) using Δ HAQ-DI ≤ 0.0 and HAQ-DI ≤ 0.5 (table 1).

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

Overall, the proportion of GRO for the primary score (Δ mTSS ≤ 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 with 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 Δ mTSS ≤ 5 (table 2). None of the differences was statistically significant when Δ mTSS ≤ 0 was used (table 2).

We performed a sensitivity analysis by excluding the three apparent outliers in figure 3 (the DE019, GO-FURTHER and TEMPO trials) which confirmed no significant difference in the meta-analytic RRs (Δ mTSS ≤ 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 with patients in 4V-remission and with patients in non-remission

Overall, the proportion of GFO for the elected outcome (Δ HAQ-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 with patients in non-remission (-2.2%; 95% CI

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

| Trial (year) | n* | Remission at 6 OR 12 months, n (%) | | | Good radiographic outcome from 12 to 24 months†, 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† | ΔHAQ-DI ≤0 HAQ-DI ≤0.5 |
| DEO19 (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 2 (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) |
| DEO3 (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 | 5792 | 1378 | 1085 | 3329 | 3798 | 4370 | 5498 | 5262‡ | 2750 | 1626 |
| Meta-analytic % (95% CI) | | 23.0 (18.0 to 28.0) | 18.9 (15.4 to 22.1) | 58.1 (52.0 to 64.1) | 64.1 (54.9 to 73.2) | 74.1 (66.2 to 82.0) | 94.6 (92.9 to 96.4) | | 70.6 (67.7 to 73.5) | 31.1 (24.9 to 37.2) |

4V-remission=SJC28, TJC28, CRP (mg/dL) and PGA (0–10); all ≤1; 4V-near-remission=SJC28, TJC28, CRP (mg/dL) ≤1 and PGA (0–10) >1; non-remission=SJC28 >1 or TJC28 >1 or CRP (mg/dL) >1 at 6 or 12 months of follow-up in all cases.

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

†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.

‡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=3904).

CRP, C reactive protein; HAQ-DI, Health Assessment Questionnaire–Disability Index; ΔmTSS, change in the modified total Sharp score during the second year of follow-up; PGA, patient global assessment; SJC28, swollen 28-joint count; TJC28, tender 28-joint count.

–6.8% to +2.4%). The differences between 4V-near-remission and 4V-remission were more striking for the GFO defined as Δ HAQ-DI \leq 0 and HAQ-DI \leq 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 (Δ mTSS \leq 0.5) was higher for patients in 4V-remission compared with 4V-non-remission (LR+=1.36, 1.15 to 1.61) than between patients in 3V-remission versus 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) versus their counterparts (table 3).

The same comparisons were made regarding functional outcomes (table 3). The likelihood ratio of having GFO (Δ HAQ \leq 0.0) was significantly higher for patients in 4V-remission compared with 4V-non-remission (LR+=1.34; 1.16 to 1.54), while it was not significantly different between patients in 3V-remission versus 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) versus 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 (\leq 53%). It was, however, higher for the 3V-remission definition than for the 4V-remission definition: 6.5%, 10.6% and 17.2% higher at Δ mTSS \leq 0.0, \leq 0.5 and Δ mTSS \leq 5, respectively (see figure 4). As expected, the improved accuracy of the 3V-remission is a result of a substantially lower percentage of FN, that is, patients without remission who do not have radiographic progression, at the cost of a much smaller increase in the percentage of FP, that is, 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 I² statistics are presented in online supplemental table S4). Overall, 73.3% (95% CI 63.9% to 81.8%) of the patients in non-4V-remission still had GRO (Δ mTSS \leq 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-remission and 3V-remission, respectively. None of these differences were statistically significant.

Table 2 Pooled outcomes* 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 Δ mTSS ≤ 0.5 | | | |
|---|--|---|---------------------------|
| | 4V-remission (n=1378) | 4V-near-remission (n=1085) | Non-remission (n=3329) |
| 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) |
| Δ percentage GRO (95% CI) | 4V-near-remission vs 4V-remission -2.9 (-7.3 to 1.5) | 4V-near-remission vs Non-remission 6.2 (2.3 to 10.1) | |
| Relative risk GRO (95% CI) | 0.98 (0.94 to 1.02) | 1.07 (1.02 to 1.12) | |
| Good radiographic outcome (GRO) defined as Δ 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) |
| Δ percentage GRO (95% CI) | 4V-near-remission vs 4V-remission -7.7 (-16.6 to 1.1) | 4V-near-remission vs non-remission 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 Δ 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) |
| Δ percentage GRO (95% CI) | 4V-near-remission vs 4V-remission -2.5 (-7.5 to 2.6) | 4V-near-remission vs Non-remission 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 Δ HAQ-DI ≤ 0 | | | |
| | 4V-remission (n=1041) | 4V-near-remission (n=758) | Non-remission (n=2105) |
| 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) |
| Δ percentage GFO (95% CI) | 4V-near-remission vs 4V-remission -11.0 (-16.3 to -5.7) | 4V-near-remission vs Non-remission -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 Δ HAQ-DI ≤ 0 and HAQ-DI ≤ 0.5 | | | |
| | 4V-remission (n=1305) | 4V-near-remission (n=1003) | Non-remission (n=2954) |
| 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) |
| Δ percentage GFO (95% CI) | 4V-near-remission vs 4V-remission -39.6 (-48.4 to -30.9) | 4V-near-remission vs Non-remission 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) | |

4V-remission=SJC28, TJC28, CRP (mg/dL) and PGA (0–10), all ≤ 1 ; 4V-near-remission=SJC28, TJC28, CRP (mg/dL) ≤ 1 and PGA (0–10) > 1 ; non-remission=SJC28 > 1 or TJC28 > 1 or CRP (mg/dL) > 1 , irrespective of PGA value; at 6 or 12 months of follow-up in all cases.

*Determined by meta-analyses: for each trial, we calculated the differences in the proportion/change (Δ 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).

CRP, C reactive protein; Δ HAQ-DI, change in Health Assessment Questionnaire–Disability Index; PGA, patient global assessment; SJC28, swollen 28-joint count; TJC28, tender 28-joint count.

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 (online supplemental tables 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 2 years (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 Δ 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 Δ mTSS > 0.5 units were 18.4% (13.8% to 23.5%) and 15.2% (9.9% to 21.4%), respectively (online supplemental table S7).

DISCUSSION

This is the first study assessing the prevalence of 4V-near-remission in RCTs and the first comparing radiographic damage

A. 4V-near-remission vs 4V-remission

| Studies (year) | Estimate (95%CI) | GRO 4v-near-rem | GRO 4V-rem |
|---|-------------------|--------------------|---------------|
| DE019 (2004) | 1.12 (0.92, 1.36) | 37/45 | 50/68 |
| TEMPO (2004) | 0.82 (0.71, 0.96) | 65/91 | 98/113 |
| COMET (2008) | 0.99 (0.91, 1.09) | 96/107 | 92/102 |
| RAPID 1 (2008) | 0.98 (0.87, 1.10) | 111/143 | 140/177 |
| RAPID 2 (2009) | 1.00 (0.85, 1.16) | 68/81 | 43/51 |
| GO-FORWARD (2010) | 0.96 (0.78, 1.17) | 51/74 | 62/86 |
| GO-BEFORE (2011) | 0.95 (0.87, 1.04) | 71/80 | 109/117 |
| LITHE (2011) | 1.06 (0.95, 1.18) | 145/174 | 115/146 |
| DE013 (2013) | 0.91 (0.74, 1.11) | 35/50 | 120/156 |
| GO-FURTHER (2014) | 0.82 (0.56, 1.19) | 35/89 | 26/54 |
| FUNCTION (2016) | 0.99 (0.94, 1.05) | 138/151 | 283/308 |
| Overall (I ² =10.24%, P=0.347) | 0.98 (0.94, 1.02) | 852/1085 | 1138/1378 |

B. 4V-near-remission vs Non-remission

| Studies (year) | Estimate (95%CI) | GRO 4v-near-rem | GRO Non-rem |
|---|-------------------|--------------------|----------------|
| DE019 (2004) | 1.22 (1.04, 1.43) | 37/45 | 210/312 |
| TEMPO (2004) | 1.02 (0.87, 1.19) | 65/91 | 167/238 |
| COMET (2008) | 1.20 (1.07, 1.35) | 96/107 | 101/135 |
| RAPID 1 (2008) | 1.00 (0.90, 1.11) | 111/143 | 257/330 |
| RAPID 2 (2009) | 1.12 (1.00, 1.26) | 68/81 | 213/285 |
| GO-FORWARD (2010) | 1.15 (0.95, 1.39) | 51/74 | 115/192 |
| GO-BEFORE (2011) | 1.01 (0.92, 1.10) | 71/80 | 266/302 |
| LITHE (2011) | 1.04 (0.96, 1.13) | 145/174 | 380/476 |
| DE013 (2013) | 1.19 (0.97, 1.46) | 35/50 | 196/334 |
| GO-FURTHER (2014) | 1.29 (0.77, 1.38) | 35/89 | 130/340 |
| FUNCTION (2016) | 1.02 (0.96, 1.08) | 138/151 | 345/385 |
| Overall (I ² =33.29%, P=0.132) | 1.07 (1.02, 1.12) | 852/1085 | 2380/3329 |

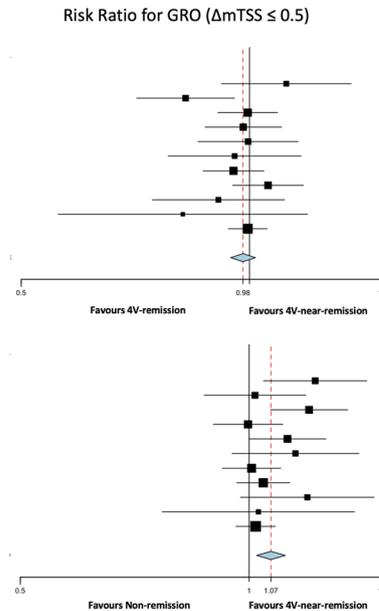


Figure 3 Meta-analyses of risk ratio of obtaining good radiographic outcome ($\Delta mTSS \leq 0.5$ units); 4V-near-remission vs 4V-remission and vs non-remission. *Legend:* 4V-remission=SJC28, TJC28, CRP (mg/dL) and PGA (0–10), all ≤ 1 ; 4V-near-remission=SJC28, TJC28 and CRP (mg/dL) ≤ 1 and PGA (0–10) > 1 ; non-remission=SJC28 > 1 and/or TJC28 > 1 and/or CRP (mg/dL) > 1 , irrespective of PGA value; at 6 or 12 months of follow-up in all cases. CRP, C reactive protein; $\Delta mTSS$, change in the modified total Sharp score during the second year of follow-up; GRO, good radiographic outcome; PGA, patient global assessment; SJC28/TJC28, swollen/tender 28-joint counts.

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 (ie, 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 with the 4V-remission. This would occur at the cost of having an excess

of 6.1% of patients having a $\Delta mTSS > 0.0$, 4.0% of patients having a $\Delta mTSS > 0.5$ and of 0.7% having $\Delta mTSS > 5$ 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 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 emphasises that all remission concepts have a relatively poor predictive value regarding radiographic damage, as shown by low LR_s (although better in 4V-remission) and

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

| Good outcome* | 4V-Remission (vs non-4V) | | I ² LR+ LR– | 3V-Remission (vs non-3V) | | I ² LR+ LR– |
|--|--------------------------|---------------------|---------------------------|--------------------------|---------------------|---------------------------|
| | LR+ (95% CI) | LR– (95% CI) | | LR+ (95% CI) | LR– (95% CI) | |
| $\Delta mTSS \leq 0.5$ | 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% |
| $\Delta mTSS \leq 0$ | 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% |
| $\Delta mTSS \leq 5$ | 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% |
| $\Delta HAQ-DI \leq 0$ | 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% |
| $\Delta HAQ-DI \leq 0$ and $HAQ-DI \leq 0.5$ | 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% |

4V-remission=SJC28, TJC28, CRP (mg/dL) and PGA (0–10), all ≤ 1 ; 3V-remission=SJC28, TJC28 and CRP (mg/dL) ≤ 1 ; non-remission=SJC28 > 1 or TJC28 > 1 or CRP (mg/dL) > 1 , irrespective of PGA value; at 6 or 12 months of follow-up in all cases.

*n=5792 for $\Delta mTSS$, n=3904 for $\Delta HAQ-DI \leq 0$ and n=5262 for $\Delta HAQ-DI \leq 0$ and $HAQ-DI \leq 0.5$.

CRP, C reactive protein; $\Delta HAQ-DI$, change in Health Assessment Questionnaire–Disability Index; LR+, positive likelihood ratio; LR–, negative likelihood ratio; $\Delta mTSS$, change in the modified total Sharp score during the second year of follow-up; PGA, patient global assessment; SJC28, swollen 28-joint count; TJC28, tender 28-joint count.

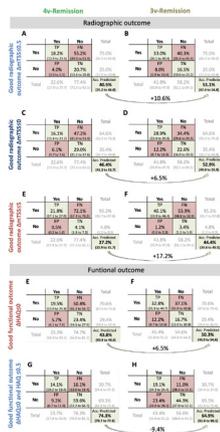


Figure 4 Pooled meta-analytic prediction accuracy of 4V-remission and 3V-remission status for the good radiographic and functional outcomes. *Footnote:* The sum of the meta-analytic percentages of TP, FN, FP and TN is slightly less than 100% due to error estimation when multi-category ($k>2$) prevalence 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 5792 analysed patients (11 randomised controlled trials (RCTs)), E and F include 3904 (8 RCTs), and G and H 5262 analysed patients (11 RCTs). *Legend:* 4V-remission=SJC28, TJC28, CRP (mg/dL) and PGA (0–10), all ≤ 1 ; 3V-remission=SJC28, TJC28 and CRP (mg/dL) ≤ 1 ; Δ HAQ, change in Health Assessment Questionnaire score; Δ mTSS, change in the modified total Sharp score from 12 months to 24 months; CRP, C reactive protein; FN, false negative; FP, false positive; PGA, patient global assessment; SJC28, swollen 28-joint count; TJC28, tender 28-joint count; TN, true negative; TP, true positive; accurately predicted=TP+TN. Between brackets is the pooled 95% CI.

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 (Δ mTSS >0.5).

4V-remission was associated with significantly higher rates of GFO (77.6%) compared with 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-remission if the definition of GFO adopted by the ACR/EULAR committee was used (4V-remission=60.5%, 4V-near-remission=22.5%, 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

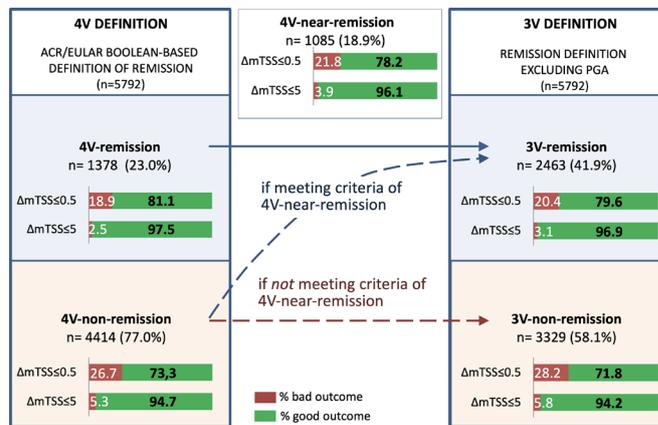


Figure 5 Reclassification of remission status and respective radiographic outcomes (n=5792). Percentages were calculated through meta-analyses. *Footnote:* Excluding PGA from the remission of remission (3V-remission) almost duplicated the percentage of patients in remission but showed only a slight increase 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 (mg/dL) and PGA (0–10), all ≤ 1 ; 4V-near-remission=SJC28, TJC28, CRP (mg/dL) ≤ 1 and PGA (0–10) >1 ; non-remission=SJC28 >1 and/or TJC28 >1 and/or CRP (mg/dL) >1 , irrespective of PGA value; 3V-remission=SJC28, TJC28, CRP (mg/dL) ≤ 1 . All definitions as observed at 6 or 12 months. *Note:* CIs and I^2 statistics of pooled radiographic outcomes can be found in online supplemental table S4. Δ mTSS, change in the modified total Sharp score during the second year of follow-up; CRP, C reactive protein; PGA, patient global assessment; SJC28/TJC28, swollen/tender 28-joint counts.

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 (1) the use of individual patient data, allowing uniform analyses procedures, (2) the availability of data collected under stringent RCT conditions, (3) the inclusion of over 5700 patients and (4) 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 recognised 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 randomisation. 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, that is, its responsiveness, should be established and compared with 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 randomisation. As patients fulfilled inclusion criteria for RCTs, generalisability 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 $\Delta\text{mTSS} \leq 0.5$ 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, that is, 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 Simplified Disease Activity Index and Clinical Disease Activity Index, 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 ($\Delta\text{mTSS} \leq 0$), 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 two 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 $\Delta\text{mTSS} \leq 0$ was the one with more favourable results for the 4V compared with the 3V-remission in terms of GRO prediction, predictive accuracy and rate of FN, but not in LR, for which the $\Delta\text{mTSS} \leq 0.5$ was more favourable. While considering these issues, one should take into account that $\Delta\text{mTSS}=1$ 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 versus 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-remission versus the 3V-remission. This position was recently reaffirmed.²² The reasons why we disagree with this approach are presented previously. Furthermore, the ACR/EULAR study analysed primarily the methotrexate-alone treatment groups of 3 trials, while we included all arms in each of 11 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 nine 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 $\Delta\text{mTSS} \geq 0.5$ (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 CRP ≤ 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 4381 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, for example, 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 two 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 recognised 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, that is, the dual target strategy, deserves consideration. The first target aims to control inflammation (biological remission) and the other one to control 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.

Author affiliations

- ¹Rheumatology, Centro Hospitalar e Universitário de Coimbra EPE, Coimbra, Portugal
²Health Sciences Research Unit: Nursing (UICISA: E), Escola Superior de Enfermagem de Coimbra, Coimbra, Portugal
³Rheumatology and Clinical Immunology, UMC Utrecht, Utrecht, The Netherlands
⁴Institut Pierre Louis d'Epidémiologie et de Santé Publique, INSERM, Sorbonne Université, Paris, France
⁵Rheumatology, Pitié Salpêtrière Hospital, AP-HP, Paris, France
⁶Faculty of Health and Applied Sciences, University of the West of England Bristol, Bristol, UK
⁷Centre for Rheumatology & Department of Neuromuscular Diseases, University College London, London, UK
⁸Rheumatology, University College London Hospitals NHS Foundation Trust, London, UK
⁹Rheumatology, Northwick Park Hospital, London North west UniversityHealthcare NHS Trust, London, UK
¹⁰Rheumatology, Leiden University Medical Center, Leiden, Zuid-Holland, The Netherlands
¹¹Clinica Universitária de Reumatologia, and i-CBR Coimbra Institute for Clinical and Biological Research, Faculty of Medicine, University of Coimbra, Coimbra, Portugal

Twitter Ricardo J O Ferreira @FerreiraRJO, Mwidimi Ndosi @ndosi and Pedro M Machado @pedromcmachado

Acknowledgements 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.

Contributors 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.

Funding This manuscript is based on research using data from data contributors AbbVie, Pfizer and UCB that have been made available through Vivli, Inc. This study was also supported by CSDR (ClinicalStudyDataRequest), which has an agreement with Roche Inc. (Project no. 1808). Data were also obtained from the Yale University Open Data Access Project (YODA Project no. 2017-1451), which has an agreement with Janssen Research & Development, LLC. PMM is supported by the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre (BRC). RJOF was supported by a grant from ARCo – Associação de Reumatologia de Coimbra, a non-profit association of health professionals.

Competing interests RJOF reports a research grant from Abvie and speaker fees from Sanofi Genzyme, Amgen, MSD and UCB Pharma. JWJ 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 and UCB Pharma. JAPS reports a research grant from Pfizer and Abvie, and speaker fees from Pfizer, AbbVie, Roche, Lilly and Novartis.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

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

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available.

ORCID iDs

Ricardo J O Ferreira <http://orcid.org/0000-0002-2517-0247>
 Johannes W G Jacobs <http://orcid.org/0000-0002-7438-3468>
 Laure Gossec <http://orcid.org/0000-0002-4528-310X>
 Mwidimi Ndosi <http://orcid.org/0000-0002-7764-3173>
 Pedro M Machado <http://orcid.org/0000-0002-8411-7972>
 Désirée van der Heijde <http://orcid.org/0000-0002-5781-158X>
 Jose A P Da Silva <http://orcid.org/0000-0002-2782-6780>

REFERENCES

- 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* 2016;75:3–15.
- Smolen JS, Landewé 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;79:685–99.
- Singh JA, Saag KG, Bridges SL, *et al.* 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Rheumatol* 2016;68:1–26.
- 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.
- van Tuyl LHD, Boers M. Patient's global assessment of disease activity: what are we measuring? *Arthritis Rheum* 2012;64:2811–3.
- 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.
- van Tuyl LHD, Boers M. Rheumatoid arthritis: remission – keeping the patient experience front and centre. *Nat Rev Rheumatol* 2017;13:573–4.
- Ferreira RJO, Duarte C, Ndosi M, *et al.* The controversy of using PGA to define remission in RA. *Nat Rev Rheumatol* 2018;14:245.
- 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* 2018;70:369–78.
- 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* 2012;51:1076–80.
- 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.
- Balogh E, Madruga Dias J, 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.
- Svensson B, Andersson MLE, Bala S-V, *et al.* Long-term sustained remission in a cohort study of patients with rheumatoid arthritis: choice of remission criteria. *BMJ Open* 2013;3:e003554.
- Ferreira RJO, Dougados M, Kirwan JR, *et al.* Drivers of patient global assessment in patients with rheumatoid arthritis who are close to remission: an analysis of 1588 patients. *Rheumatology* 2017;56:1573–8.
- 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.
- 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* 2019;78:e109.
- Ferreira RJO, Carvalho PD, Ndosi M, *et al.* Impact of patient's global assessment on achieving remission in patients with rheumatoid arthritis: a multinational study using the METEOR database. *Arthritis Care Res* 2019;71:1317–25.
- 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* 2017;69:323–9.
- Landewé RBM. Overdiagnosis and overtreatment in rheumatology: a little caution is in order. *Ann Rheum Dis* 2018;77:1394–6.
- Landewé 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;78:e110.
- Navarro-Compán 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* 2015;54:994–1007.
- 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;79:445–52.
- 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 PROSPERO, 2017. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017057099
- 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.
- 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. *Arthritis Rheum* 2010;62:2569–81.
- 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. *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* 2020;59:153–64.
- 31 Fleischmann RM, Huizinga TWJ, Kavanaugh AF, *et al.* Efficacy of tofacitinib monotherapy in methotrexate-naïve 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. *J Stat Softw* 2012;49:1–15.
- 35 Macaskill P, Gatsonis C, Deeks JJ, *et al.* Chapter 10 analysing and presenting results. In: Deeks JJ, Bossuyt JJ, Gatsonis JJ, *et al.*, eds. *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 JPT, 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, Landewé R, *et al.* Association of 17 definitions of remission with functional status in a large clinical practice cohort of patients with rheumatoid arthritis. *J Rheumatol* 2020;47:20–7.
- 40 Landewé R, Østergaard 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* 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* 2020. doi:10.1002/acr.24237. [Epub ahead of print: 27 Apr 2020].
- 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 DMFM, *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 Landewé 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. doi:10.1002/art.41461. [Epub ahead of print: 02 Aug 2020].
- 49 Santos E, 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 A-M, 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.