Validity and reliability of the EULAR instrument RAID.7 as a tool to assess individual domains of impact of disease in rheumatoid arthritis: a cross-sectional study of 671 patients

Catia Duarte, Eduardo José Ferreira Santos, Ricardo J O Ferreira, Tore K Kvien, Maxime Dougados, Maarten de Wit, Jose Antonio Pereira da Silva, Laure Gossec

INTRODUCTION

The ultimate aim of treatment of patients with rheumatoid arthritis (RA) is to consistently optimise and maintain the patient’s health-related quality of life. This aim is strongly supported by effective and persistent control of the disease process, as recommended by the current paradigm of treatment, epitomised by the treat-to-target strategy.

ABSTRACT

Objective The rheumatoid arthritis impact of disease (RAID) questionnaire comprises seven patient-important domains of disease impact (pain, function, fatigue, sleep disturbance, emotional well-being, physical well-being, coping). RAID was validated as a pooled-weighted score. Its seven individual items separately could provide a valuable tool in clinical practice to guide interventions targeting the patient’s experience of the disease. The aim was to separately assess the psychometric properties of each of the seven numeric rating scale (NRS) of the RAID (RAID.7).

Material and methods Post hoc analyses of data from the cross-sectional RAID study and from the Rainbow study, an open-label 12-week trial of etanercept in patients with RA. Construct validity of each NRS was assessed cross-sectionally in the RAID data set by Spearman’s correlation with the respective external instrument of reference. Using the rainbow data set, we assessed reliability through intraclass correlation coefficient between the screening and baseline visits and responsiveness (sensitivity to change) from 0.93 (0.73–1.13) (sleep disturbance) to 1.34 (1.01–1.64) (pain).

Results A total of 671 patients with RA features of established disease were analysed, 563 and 108 from RAID and Rainbow, respectively. The NRS correlated moderately with the respective external instrument of reference. Using 563 from 671 patients, we assessed reliability through intraclass correlation coefficient between the screening and baseline visits and responsiveness (sensitivity to change) from 0.93 (0.73–1.13) (sleep disturbance) to 1.34 (1.01–1.64) (pain).

Conclusion The separate use of the individual NRS of RAID (RAID.7) is valid, feasible, reliable and sensitive to change, representing an opportunity to improve the assessment and treatment of disease impact with minimal questionnaire burden.

Trial registration number NCT00768053.

What is already known about this subject?

- A substantial proportion of patients with rheumatoid arthritis (RA) with complete control of the inflammatory process still perceive significant impact of the disease, requiring adjunctive interventions.
- This supports the proposal for the adoption of two separate targets: one focused on the inflammatory process and the other on disease impact on patient’s lives, still to be defined.
- Rheumatoid arthritis impact of disease (RAID) is a patient-reported outcome measure that includes seven domains of highest relevance for patients with RA. However, used as a single score, it does not provide information to select and monitor interventions.

What does this study add?

- Using the individual numerical rating scales of RAID as separate scores (RAID.7) provides a tool that is valid, feasible, reliable and sensitive to change.
- RAID.7 is a useful tool to evaluate the impact of RA in different domains, with minimal interference in daily clinical practice.

How might this impact on clinical practice or future developments?

- RAID.7 allows for the regular and personalised assessment of patients’ needs, supporting the selection and the monitoring of the most appropriate interventions to mitigate the impact of RA beyond disease process remission.
However, an important proportion of patients who are otherwise in remission according to inflammatory activity still describes a substantial impact of disease, reflected by a patient global assessment (PGA) score above 1 (from 0 to 10=worst scale). A recent meta-analysis of clinical practice cohort studies estimated that 12% of patients with RA reach the American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) Boolean-based definition of remission, while 19% (95% CI 15% to 23%) fail this solely due to PGA >1. Similar rates of this near-remission are observed in randomised clinical trials (15% to 22%). Having reached a status where (inflammatory) disease activity is under control, these patients cannot be expected to improve further through reinforcement of immunosuppressive therapy. They are likely to require adjunctive interventions designed to diminish the uncontrolled domains of disease impact. To select the most appropriate interventions, it is essential that health professionals and patients have access to meaningful, feasible and reliable instruments to identify these domains and access the success of interventions.

Several validated tools have been used to assess the impact of RA in patient’s lives. However, most of them are nonspecific for RA or too complex and time-consuming to be feasible in clinical practice. Co-operative work with patient research partners, under the auspices of EULAR, led to the development of the rheumatoid arthritis impact of disease (RAID) questionnaire, which addresses the seven domains of impact of RA considered of utmost relevance by patients. These include pain, function and fatigue, considered as core domains, and also sleep, emotional and physical well-being and coping. Each domain is evaluated by patients using a single 0 (no symptoms) to 10 numeric rating scale (NRS) and has a specific weight used to calculate a single composite RAID score. The RAID was initially designed for use in clinical trials, as a composite index, and proved to be feasible, reliable and sensitive to change, in different sets of patients, with an estimation of patient acceptable symptom state ≤2 and a minimal clinical important improvement of 3. RAID domains have been shown to be associated strongly with quality of life and subjective well-being. We hypothesise that considering the seven items of RAID separately (RAID.7) would provide a valuable tool in routine clinical care to identify patients’ treatment goals and assess the efficacy of comprehensive interventions. However, the use of individual questions has only been assessed for function, sleep and coping. The aim of this study was to assess the psychometric properties of the RAID.7, that is, of each of the individual components of RAID.

MATERIAL AND METHODS
The conceptual framework for outcome measurement (OMERACT V.2.0) was used in this methodological study, to assess the psychometric properties of each of the seven NRS questions of the RAID.

Study design and patients
Data from two studies were used in this post hoc analysis: (a) the RAID study, a cross-sectional data set used to develop the RAID, which included 563 outpatients with RA with definite diagnosis from 12 European countries (Estonia, Germany, Finland, France, Greece, Italy, The Netherlands, Norway, Spain, Romania, Turkey and the UK) and (b) the Rainbow study, a 12-week, open-label, French multicentre, single-arm trial of etanercept in RA (Clinical Trials NCT 00768053). This trial included 108 patients fulfilling the ACR criteria for RA, naive to biological agents, with active disease, as defined by a Disease Activity Score 28-erythrocyte sedimentation rate (DAS28-ESR) >3.2, and at least one of the following: ≥4 swollen joints or C reactive protein (CRP) ≥10 mg/L or ESR ≥28 mm/first hour. All patients received weekly subcutaneous injections of 50 mg of etanercept and were observed at screening, baseline, weeks 4 and 12. The RAID data set was used for assessment of the validity of RAID.7, while the Rainbow was used to assess reliability and sensitivity to change.

Patient-reported outcomes of interest
The individual questions included in the RAID score were worded as in the original study and assessed through an NRS from 0 (no symptoms) to 10 (table 1). The RAID was translated by the principal investigator and the patient research partner in each country. Patients responded to the questionnaire before clinical appointments, autonomously without influence from the healthcare provider. The answers were recorded and analysed separately for each individual item.

Comparator outcome measures
The instruments used in the RAID study as reference for each NRS of RAID.7 are shown in table 1. Feelings of helplessness, despite the use of coping strategies, were also addressed by the following question ‘How helpless did you feel because of your rheumatoid arthritis although you used these strategies during the past week? which was answered on an NRS of 0 (‘not at all helpless’) to 10 (‘very helpless’).

Other variables
Sociodemographic variables (age and gender) and disease duration were collected at the first visit (RAID baseline or rainbow screening). Physician Global Assessment of disease activity, swollen and tender 28 joint counts, ESR and CRP were assessed at baseline visit in the RAID study and at screening, baseline and 12 weeks of follow-up in the Rainbow study. Disease activity scores with 28 joint counts and ESR, without PGA (DAS28(3v)-ESR) were determined and categorised as: remission <2.6, low <3.2, moderate ≤5.1 and high disease activity >5.1. In the rainbow trial, patients with an improvement of DAS28(3v)-ESR ≥1.2 at week 12...
were classified as responders. We used DAS28(3v)-ESR throughout to avoid the effects of collinearity between PGA and RAID domains.

**Statistical analyses**

**RAID study:** the following properties and measures were used in this data set:

**Construct validity**

Construct validity was calculated by Spearman’s correlation between each single RAID domain and the respective instrument of reference (reversed, if necessary). Correlations from 0.3 to 0.5 were considered weak, moderate, from 0.5 to 0.7, and strong if $>0.7$. 22

**Feasibility**

Feasibility was calculated by the percentage of missing data for each of the seven questions. The floor and the ceiling effect of each NRS were evaluated through the proportion of patients who scored 0 and 10, respectively.

**Known-group validity**

Known-group validity was evaluated by comparing the scores for each NRS across different categories of DAS28(3v)-ESR, using the Kruskal-Wallis test.

Rainbow study: the following properties were assessed in this data set.

**Reliability**

Comparing stability between screening and baseline visits (maximum of 2 weeks apart), through the intraclass correlation coefficient (ICC, two-way mixed model absolute agreement) with 95% CI. An ICC $>0.8$ was considered indicative of excellent reliability. 23 Agreement was assessed using the Bland-Altman method. 24

**Responsiveness**

Responsiveness was calculated by the standardised response means (SRM between baseline and 12-week visits, and the respective intervals of confidence were obtained through bootstrapping procedures. 25 26 SRMs $>0.8$ were considered large.

Responsiveness was also assessed by the correlation between change in DAS28(3v)-ESR and each domain through the Pearson’s coefficient of correlation. Comparison between responders and nonresponders was performed through independent samples t-test. Additionally, individual domains were compared between patients in remission/low disease activity and patients with moderate/high disease activity at the end of the study through independent sample t-test.

Quantitative data were expressed as means (SD) and categorical data as frequencies and percentages. No
imputation of missing data was performed for any statistical analysis.

Statistical analyses were performed through SPSS IBM for Windows V.24 and Medcalc software V.18.11.3.

RESULTS

A total of 671 patients were included in these post-hoc analyses (563 from the RAID study and 108 from the Rainbow study). Baseline demographic and clinical characteristics of patients are presented in table 2.

RAID dataset: validity and feasibility assessment

The mean of the NRSs of each domain of RAID varied from 3.8 to 4.8, at baseline, with pain-NRS presenting the highest score (table 3). A moderate floor effect was observed in the majority of the NRS: 4.5%–16.5% reported a score of 0, which was more pronounced for the sleep NRS (16.5% scores <1). A ceiling effect was present but less pronounced in all, with only 1.8% (coping NRS) to 4.7% (sleep NRS) of the patients scoring 10. Feasibility was excellent for all individual items, with a much lower percentage of missing data (0.7%–1.2%) than observed in the responses to the reference instruments (5.4%–21.9%) (online supplemental table S1).

All NRS were moderately to strongly correlated with the respective reference Patient Reported Outcome Measures (PROMs) (table 3): the highest correlations were observed, as expected, for pain vs MOS-SF36 bodily pain (r=0.81, 95% CI: 0.77 to 0.85), and for function vs mHAQ (r=0.75, 95% CI: 0.71 to 0.79). The coping-NRS was moderately correlated with the Helplessness question (r=0.65, 95% CI: 0.58 to 0.72).

Higher disease activity states (DAS28(3v)-ESR) were associated with significantly higher mean scores in all domains of disease impact. RAID.7 scores were similar in remission and low disease activity states (figure 1).

Rainbow dataset: reliability and responsiveness to change assessment

Reliability

The ICC from screening to baseline was excellent for fatigue (0.80, 95% CI: 0.72 to 0.86) and sleep (0.83,
95% CI:0.76 to 0.88) and moderate for the remaining domains (table 4). The Bland-Altman agreements in test–retest scores are shown in online supplemental figure 1A–G with mean differences close to zero for all NRS.

### Responsiveness

The highest SRM from baseline to week 12 was found for pain (1.34, 95% CI:1.01 to 1.64), physical well-being (1.25, 95% CI:0.97 to 1.50), function (1.21, 95% CI:0.89 to 1.49) and fatigue (1.16, 95% CI:0.89 to 1.36). Sleep and coping were the least responsive domains (SRM=0.95 (95%CI: 0.73 to 1.13) and 0.97 (95%CI:0.63 to 1.22), respectively). Responsiveness of individual RAID domains was similar to that observed with PGA (1.35, 95% CI:1.01 to 1.63) but lower than demonstrated for swollen joint counts (1.82, 95% CI:1.59 to 2.05) and DAS28(3v)-ESR (1.80, 95% CI:1.5 to 2.1) (table 4).

A correlation, although weak, was observed between change in DAS28(3v)-ESR and change in individual domains, ranging between 0.3 (sleep disturbance, coping, physical and emotional well-being) and 0.4 (pain, fatigue, and function). Patients who achieved low disease activity/remission at week 12 reported a lower impact in all domains compared with patients who remained in moderate to high disease activity (figure 2). Additionally, responders with an improvement ≥1.2 in DAS28(3v)-ESR between baseline and week 12, showed a higher change in all domains, except for sleep disturbance, compared with non-responders (figure 3).

The psychometric properties of each NRS are summarised in table 5.

### DISCUSSION

This study brings important information for the research and clinical assessment of RA patients. Our results demonstrate that each individual NRS used to assess specific domains of impact in the RAID is valid, feasible, reliable and sensitive to change in patients with RA, performing similarly to other more extensive instruments.

Individual NRS correlated moderately to strongly with the corresponding instrument of reference, with the strongest correlations being observed for pain and function which is probably related to the utmost importance of these two domains for patients with RA.27–30 The

---

**Table 4** Reliability and responsiveness of the NRS questions of RAID from the rainbow study

<table>
<thead>
<tr>
<th>RAID NRS</th>
<th>Screening</th>
<th>Baseline (2 weeks after screening)</th>
<th>12 weeks (after baseline)</th>
<th>Reliability screening baseline (ICC, 95% CI) (n=108)</th>
<th>Sensitive to change baseline 12 weeks (SRM, 95% CI) (n=97)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tender joint count</td>
<td>10.3 (5.9)</td>
<td>10.7 (5.8)</td>
<td>4.2 (4.8)</td>
<td>0.90 (0.85 to 0.93)</td>
<td>1.18 (0.90 to 1.38)</td>
</tr>
<tr>
<td>Swollen Joint Count</td>
<td>8.2 (9.8)</td>
<td>8.6 (4.1)</td>
<td>2.3 (2.6)</td>
<td>0.89 (0.83 to 0.92)</td>
<td>1.82 (1.59 to 2.05)</td>
</tr>
<tr>
<td>PGA</td>
<td>6.6 (1.7)</td>
<td>6.5 (1.9)</td>
<td>3.3 (2.3)</td>
<td>0.63 (0.50 to 0.73)</td>
<td>1.35 (1.01 to 1.63)</td>
</tr>
<tr>
<td>DAS28(3v)-ESR</td>
<td>5.1 (0.8)</td>
<td>5.1 (0.8)</td>
<td>3.3 (1.1)</td>
<td>0.93 (0.90 to 0.95)</td>
<td>1.80 (1.50 to 2.10)</td>
</tr>
<tr>
<td>Pain</td>
<td>6.6 (1.8)</td>
<td>6.5 (1.8)</td>
<td>3.1 (2.3)</td>
<td>0.64 (0.51 to 0.74)</td>
<td>1.34 (1.01 to 1.64)</td>
</tr>
<tr>
<td>Function</td>
<td>6.4 (1.9)</td>
<td>6.3 (2.0)</td>
<td>3.2 (2.4)</td>
<td>0.69 (0.58 to 0.78)</td>
<td>1.21 (0.89 to 1.49)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>6.4 (2.3)</td>
<td>6.2 (2.3)</td>
<td>3.4 (2.7)</td>
<td>0.80 (0.72 to 0.86)</td>
<td>1.16 (0.89 to 1.36)</td>
</tr>
<tr>
<td>Emotional Well-being</td>
<td>5.5 (2.5)</td>
<td>5.3 (2.4)</td>
<td>2.8 (2.4)</td>
<td>0.75 (0.65 to 0.82)</td>
<td>1.04 (0.79 to 1.26)</td>
</tr>
<tr>
<td>Physical Well-being</td>
<td>6.3 (2.0)</td>
<td>6.2 (1.9)</td>
<td>3.1 (2.5)</td>
<td>0.78 (0.70 to 0.85)</td>
<td>1.25 (0.97 to 1.50)</td>
</tr>
<tr>
<td>Sleep Disturbance</td>
<td>5.3 (2.9)</td>
<td>5.5 (2.8)</td>
<td>2.9 (2.7)</td>
<td>0.83 (0.76 to 0.88)</td>
<td>0.93 (0.73 to 1.13)</td>
</tr>
<tr>
<td>Coping</td>
<td>4.9 (2.2)</td>
<td>4.8 (2.0)</td>
<td>2.6 (2.2)</td>
<td>0.64 (0.52 to 0.74)</td>
<td>0.97 (0.63 to 1.22)</td>
</tr>
</tbody>
</table>

Results are presented as Mean (SD) at each time point. CI obtained through bootstrapping method. ICC >0.8 was considered excellent. SRM >0.8 was considered as responsive.

DAS28(3v)-ESR, Disease activity Score, 3 variables with erythrocyte sedimentation rate; ICC, intra-class correlation; NRS, numerical rating scale; PGA, patient global assessment ; SRM, standardised response mean.

---

**Figure 2** Mean NRS for each RAID domain, according to disease activity status (Das28(3v)-ESR) at week 12 in rainbow study (n=97) * correspond to p<0.05, independent samples t-test, bars correspond to mean, lines to 95% CI, obtained through bootstrapping method. DAS28(3v)-ESR: Disease activity Score, 3 variables with erythrocyte sedimentation rate; LDA: Low Disease Activity; NRS: Numerical Rating Scale; RAID, Rheumatoid Arthritis Impact of Disease.
coping NRS showed a moderate correlation with Helplessness question but not with 18-Item Coping Questionnaire. This is probably explained by the formulation of the questions, which is much closer to the NRS. The appropriateness of this coping questionnaire, as well as the best instrument, to assess this domain in RA have been previously questioned,12,31 remaining as unsewered question. Furthermore, it is recognised that this whole concept is difficult to understand by patients and capture in a scale.32 The observations above support, from our perspective, the use of the coping NRS, though it may be useful to further assess this specific NRS in other studies. It is interesting that coping also was selected as a preferred domain in a similar exercise with patients with psoriatic arthritis.33

All individual NRS were shown to be sensitive to changes induced by treatment with etanercept, with larger improvements seen on physical domains such as pain, function, and physical well-being. These findings are consistent with other studies indicating that 'physical domains' are more modifiable/responsive to immunosuppressive treatment than social and psychological measures.34-37 Patients with active disease reported higher levels of impact in all domains. However, patients with low disease activity or in remission still perceived a significant average impact in all domains, suggesting that despite the large-scale improvement in disease activity with immunosuppressive drugs did not translate into equally large improvements in patient reported outcomes. Such findings reinforce that a T2T strategy is not enough and from a patient perspective more attention should be paid to the importance of residual symptoms beyond biologic or inflammatory remission, underlining that many patients need additional support to reduce disease impact even after the inflammatory process has been abrogated.38

A very slight floor effect was observed in all domains, rendering the measure unable to discriminate between subjects in the lowest extreme of the scale, which may limit the use of the individual NRS. Physical domains, i.e., pain, function, fatigue and physical well-being had the weakest floor effect, which is consistent with the prevalence and relevance of these issues among patients with RA.27 Similarly to another study, sleep and coping had the largest floor effect,14 reflecting, perhaps, their lower specificity or relevance in RA.

Individual NRS of RAID showed to be feasible with response rates higher than 98%, making them suitable for use in research and clinical practice.

When compared with the literature, the present results suggest that individual NRS are as good or even better measures than extensive reference questionnaires.39 In fact, besides having a similar psychometric performance, individual NRS used in RAID are based on disease-specific wording, while other PROMs are not. This makes RAID NRS less prone to confounding influence by comorbidities or life events not related with RA.

Although our results suggest that individual domains capture change of disease activity over time, the correlation between change in disease activity and RAID NRS is generally weak. This highlights that what we measure as impact is only partially dependent on disease activity.

The findings of this study must be considered in the light of some limitations and strengths. This main study strength is addressing domains of impact that have been chosen as important by patients with RA.27 Similarly to another study, sleep and coping had the largest floor effect,14 reflecting, perhaps, their lower specificity or relevance in RA.

Table 5 Summary of psychometric properties of the seven NRS questions of the RAID

<table>
<thead>
<tr>
<th>NRS</th>
<th>Construct validity</th>
<th>Feasibility</th>
<th>Reliability</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain-NRS</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>Function-NRS</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>Fatigue-NRS</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>Sleep-NRS</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
</tr>
<tr>
<td>Physical well-being-NRS</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>Emotional well-being-NRS</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>Coping-NRS</td>
<td>++</td>
<td>+++</td>
<td>++</td>
<td>++</td>
</tr>
</tbody>
</table>

The + signs represent a semi-quantitative assessment of the psychometric properties with +++: good; ++: moderate; --: poor.

NRS, numerical rating scale; RAID, rheumatoid arthritis impact of disease.
the evaluation of patients with RA in randomised clinical trials and clinical practice.\textsuperscript{9} 40 41 The sample size and the cultural diversity of the patient population are also some important strengths. As limitations, we used two different sets of patients with different sociodemographic and clinical characteristics. Rainbow as a clinical trial was better suited for responsiveness assessment but could not be used for validity analysis due to the lack of the reference instruments. The population included in these studies had, on average, moderate to high disease activity, which may question the applicability of the results in less severe patients.

Our findings may have relevant implications for the management of patients with RA in clinical practice. The impact of the disease on patients’ lives must be at the core of treatment objectives and guiding targets. The abrogation of inflammatory process (biological remission) is a crucial contribution but not a guarantee that the impact of disease will also be abolished. In fact, a large proportion of patients otherwise in remission still report significant impact, as reflected by a PGA score >1. This status, designated as PGA-near-remission, affects over 60% of all patients with RA who are otherwise in remission and it has been shown to be associated with a structural damage progression similar to that of full remission.\textsuperscript{15} 42 43 These patients can only be improved by adjuvant interventions, given that the disease process is already under control; however, PGA is useless to guide the selection of the adjuvant measures needed to improve unabated domains of disease impact once the target of therapeutic drug remission has been achieved.\textsuperscript{38} The same is true for the RAID. If a single value, the composite score, is used to represent the patient’s conditions, the clinician needs to know why the patients are unsatisfied, and this requires that the diverse domains of RAID would be considered separately. So, RAID.7 may provide a useful tool to select and monitor adjuvant interventions in RA. This will, expectantly, be especially important in patients with persistent significant impact despite being in inflammatory remission or low disease activity, an object for future studies.

In conclusion, this study also highlights the very good psychometric performance of the individual domains of RAID. These results, therefore, support that RAID, considered separately in its seven domains (RAID.7), may play a useful role in the discriminant assessment of disease impact in the clinical care of individual patients with RA and, thus, in the selection of the most appropriate adjuvant interventions to ameliorate the impact of disease according to the patients’ judgement and needs.\textsuperscript{44–47} This holds great promise in promoting the ethically imperative of patient-centred care.


