RHEUMATOLOGY

Rheumatology 2023;62:98–107
https://doi.org/10.1093/rheumatology/keac257
Advance access publication 28 April 2022

Original article

Patient Experienced Symptom State in rheumatoid arthritis: sensitivity to change in disease activity and impact

Catiá Duarte [1,2], Tore K. Kvien[3,4], Joe Sexton[3], Eduardo Santos [5,6], Maarten de Wit[7], Laure Gossec[8,9] and Jose A. P. da Silva[1,2]

Abstract

Objectives. The Patient Experienced Symptom State (PESS) is a single-question, patient-reported outcome that is validated to assess global disease impact in RA. This study addresses its sensitivity to change, and reliability.

Methods. Disease activity, disease impact in the seven domains of RA Impact of Disease (RAID) and PESS were assessed in patients with RA from the NOR-DMARD registry, at two visits, 6 months apart. The PESS over the last week was scored at five levels, from 'very bad' to 'very good'. Disease impact and disease activity were compared between patients who improved, maintained or worsened PESS over time, through one-way analysis of variance, with *post hoc* Bonferroni correction. Correlations between changes in these parameters were assessed through Spearman's correlation coefficient. Sensitivity to change was assessed by standardized response mean (SRM) between the two visits. Reliability was analysed through intraclass correlation coefficient (ICC) between the two visits in patients with stable disease activity and impact.

Results. In 353 patients [76.8% females, mean (S.D.) 9.9 (9.6) years disease duration], improvement in PESS level was associated with substantial improvements in mean impact in all domains as well as disease activity ($P < 0.02$). PESS change was moderately to strongly correlated with RAID domains and disease activity (rho: 0.4–0.7). PESS was responsive to change (SRM: 0.65, 95% CI: 0.54, 0.76), particularly among RAID responders (SRM: 1.79, 95% CI: 1.54, 1.99). PESS was moderately reliable in patients with stable condition (ICC: 0.72, 95% CI: 0.52, 0.83).

Conclusion. PESS is valid, feasible, reliable and responsive, representing an opportunity to improve the assessment of disease impact with minimal questionnaire burden.

Key words: Rheumatoid Arthritis, Patient-Reported Outcomes, Patient Experienced Symptom State, Responsiveness, Reliability

CLINICAL SCIENCE

Rheumatology key messages

- Patient Experienced Symptom State (PESS) measures the patient's satisfaction with the current status of RA.
- PESS is simple, reliable and responsive to change.
- PESS seems ideal to screen patients for in-depth evaluation of disease impact.

Introduction

The current treatment paradigm for RA is typified by the treat-to-target strategy and its recommendation that disease activity should be monitored and drug therapy adjusted regularly so as to achieve and maintain a state of remission or low disease activity [1–3]. Such strategies, combined with novel therapeutic agents, have made remission a realistic goal for a majority of patients.

However, the progressive improvement of disease activity under treatment is not always paralleled by similar benefits in the patient's experience of the disease. While physicians and researchers focus on 'biological markers'

[1]Department of Rheumatology, Centro Hospitalar e Universitário de Coimbra, [2]Coimbra Institute for Clinical and Biomedical Research, Faculty of Medicine, University of Coimbra, Coimbra, Portugal, [3]Department of Rheumatology, Diakonhjemmet Hospital, [4]Faculty of Medicine, University of Oslo, Oslo, Norway, [5]Viseu Higher School of Health, Viseu, [6]Health Sciences Research Unit: Nursing, Nursing School of Coimbra, Coimbra, Portugal, [7]Patient Research Partner, Amsterdam, The Netherlands, [8]Institut Pierre Louis d'Epidémiologie et de Santé Publique, Sorbonne Université, INSERM and [9]Rheumatology Department, Pitié Salpêtrière Hospital, AP-HP, Paris, France

Submitted 22 January 2022; accepted 19 April 2022

Correspondence to: Cátia Duarte, Rheumatology Department, Centro Hospitalar e Universitário de Coimbra, Praceta Mota Pinto 3000-004 Coimbra, Portugal. E-mail: catiacmduarte@gmail.com

of disease, such as joint counts and inflammatory markers, the experience of the patients is dominated by pain, function and fatigue, among other factors [4]. This contributes frequently to disagreement between patients and physicians regarding disease status [5–8]. It is also reflected by the finding that as many as 45–60% of patients with RA who are otherwise in remission still endure significant disease impact, as reflected by a Patient Global Assessment of Disease Activity (PGA) >1 [9, 10]. PGA is actually a measure of disease impact, rather than of disease activity [11]. These incongruencies also hinder patient–physician communication and shared decision making, cornerstones of person-centred care [12]. These findings led our group to propose recently a dual-target strategy [13] where two separate targets, the inflammatory and the impact target, should be assessed and addressed in parallel, aiming at biological remission and the best possible result for the patient [14].

The patient's perspective must, indeed, remain central to disease management [13]. A wide spectrum of patient-reported outcomes (PRO) is available and validated in RA, both generic and specific to domain [15]. Some of these instruments are difficult to understand by patients as they address abstract concepts that are difficult to relate to real life challenges and goals [16, 17]. Furthermore, covering all the domains of RA impact with specific instruments would be too time-consuming for real-world practice [18].

In this context, feasible instruments and strategies that allow for a better understanding of the patient's perspective and for improved communication are important to foster better patient outcomes and are, thus, an ethical imperative [4]. We argue that a two-step approach combining a generic tool to screen for patients who perceive high impact followed by a more detailed evaluation offers an attractive strategy for the clinical setting [14].

PGA is the most frequently used PRO in RA and is part of validated indices to assess disease activity and define remission [19–22]. This question might be a natural choice to assess the patient's perception, if we take it as a measure of disease impact [11]. However, patients report difficulties in scoring PGA as a continuous variable, mainly due to its subjective nature [17, 23, 24]. Also, its scoring is affected by health literacy and background culture [17, 25].

Recently, our group explored the potential of a new tool to assess the patient's global perception of their status regarding RA: the Patient Experienced Symptom State (PESS). This question, derived from the Patient Acceptable Symptom State (PASS) [26], is composed of a single question on the patient's satisfaction with their current symptoms, offering a five-level Likert scale response (very bad, bad, acceptable, good and very good). In a cross-sectional study, PESS was shown to be easy to apply (feasible), valid and well correlated not only with other measures of disease impact, but also with disease activity. Patients who rate their status as good or very good in PESS have disease activities in the range of low disease/remission and low or absent impact of disease as measured by the RA Impact of Disease score (RAID) and its individual domains (RAID.7) [27]. Moreover, PESS is meaningful for patients and easy to understand, and has a single formulation available, which are important strengths over PGA. However, the longitudinal validity of PESS, i.e. its responsiveness and reliability, is yet to be assessed [28]. Such properties are crucial to decide whether PESS can play a useful part in defining the impact target in a dual target strategy [14].

With this study we aimed to evaluate PESS longitudinally, assessing its reliability and sensitivity to change and, especially, its ability to identify patients in need of a more detailed evaluation of impact.

Methods

Study design and patients

Data from the Norway DMARD (NOR-DMARD) registry [29], a five-centre prospective longitudinal register that includes consecutive consenting patients with inflammatory rheumatic joint diseases, were analysed. The NOR-DMARD registry has been approved by the Norwegian Data Inspectorate and Regional Ethics Committee of Eastern Norway. All patients provided written informed consent.

We selected all patients with a clinical diagnosis of RA (fulfilling the ACR 1987 and/or the ACR/EULAR 2010 classification criteria) [21, 30], included in the NOR-DMARD registry, who started a new biological (first or switch) or targeted synthetic therapy between 2012 and 2020 and were able to fill in the questionnaires. Only patients with available data regarding PESS, RAID and disease activity at baseline and 6 months of follow-up, were analysed.

This study was conducted according to the Declaration of Helsinki. The NOR-DMARD registry has been approved by the Norwegian Data Inspectorate and Regional Ethics Committee of Eastern Norway. All patients provided written informed consent.

Outcome of interest

PESS was assessed through the following question 'Consider how your rheumatic disease has affected you during the last week. If you remain in the coming months as you have been the last week, how would you rate your condition?' offering a five-level Likert scale response (very bad, bad, acceptable, good and very good) [27]. PESS was evaluated at baseline (start of a biological agent) and after 6 months.

Data collection and handling

Socio-demographic and clinical variables included gender, age, educational level, disease duration and registered comorbidities of interest (chronic low back pain, malign neoplasia and cardiovascular disease) were considered. Swollen and tender 28-joint counts performed

by rheumatologists or research nurses, ESR (mm/h), CRP (mg/dl) and medication were recorded. PGA ('Considering all the ways that your illness and health conditions affect you at this time, how do you feel?') and Physician Global Assessment of Disease Activity (PhGA) were scored through a visual analogic scale 0–100 mm (where 0 corresponds to the best possible status and 100 to the worst).

Disease activity was categorized according to the 3-variable Disease Activity Score 28 (DAS28-3vESR): high disease activity >5.1; moderate disease activity if >3.2 and ≤5.1; low disease activity (LDA) if ≥2.6 and ≤3.2; and remission <2.6. Patients with DAS28-3v improvement from baseline >1.2 or >0.6 plus at least moderate disease activity at 6 months of follow-up were classified as responders according to the EULAR criteria [22].

The Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI) [19] were also used to assess and categorize disease activity. Established thresholds for remission, low, moderate and high disease activity were considered for each index. Patients with an improvement of at least 50% from baseline were considered responders [20].

Patient-perceived impact of RA was evaluated through the RAID score [31, 32] at every visit: patients were asked to score the intensity of the impact of their disease in each of seven domains (pain, fatigue, physical function, sleep disturbance, emotional and physical well-being, and coping). Each domain was assessed on a 0–10 numerical rating scale and a combined score was derived according to the proposed algorithm. Both the combined RAID score and the individual domain scores (RAID.7) [33] were considered. Based on Minimum Clinically Important Improvement defined for RAID, patients were classified as impact responders if the change observed in the RAID score between baseline and 6 months of follow-up was ≥3 in absolute change or >50% of relative change [34].

Statistical analyses

Descriptive characteristics were presented as means (± S.D.) for continuous variables and as proportions (%) for categorical variables. No imputation of missing data was performed. Longitudinal construct validity and reliability were analysed according to the OMERACT filter [28].

Longitudinal construct validity

To assess longitudinal construct validity, we first analysed whether changes in PESS over time reflect changes in disease impact and activity. Patients were classified as 'PESS worsened', 'PESS stable' or 'PESS improved' according to change of at least one PESS category between the two visits. The mean changes in disease activity scores, RAID score and individual items of RAID in the three groups were compared through one-way analysis of variance, with *post hoc* Bonferroni correction for pairwise comparisons. Furthermore, the correlation between the degree of improvement in each of the individual domains of impact and the degree of change in PESS (number of categories) was analysed through Spearman's correlation coefficient. Correlations from 0.3–0.5 were considered weak, moderate from 0.5–0.7, and strong if >0.7.

Responsiveness of PESS was also assessed by the standardized response mean (SRM) (mean change in scores/SD of the change scores). SRMs were categorized as large (>0.80), moderate (0.5–0.80) or small (0.2–0.5) [35]. Furthermore, we compared the change of PESS in EULAR and RAID responders. SRM was calculated and compared for both groups, with an *a priori* hypothesis that it would be higher in the responder group. For responsiveness, patients considering themselves in 'very good' level at baseline were excluded from analyses.

Reliability

Reliability of PESS over time was evaluated between the two visits in the subgroup of patients with stable disease activity and impact, defined as $-0.6 < \Delta DAS283vESR < 0.6$ and $-2 < \Delta RAID < 2$. Reliability was assessed through the intra-class correlation coefficient (ICC, two-way mixed model absolute agreement) with 95% CI. An ICC >0.8 was considered indicative of excellent reliability [36].

Performance of PESS to detect responders

The ability of an improvement of at least one PESS category to identify RAID 'responders' was evaluated through a two-by-two cross-tabulation and calculation of the Crude Agreement given by [('True positive' + 'True Negative')/Total of patients)] and through Kappa statistics (k). k values <0 were considered poor, 0–0.20 weak, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good and 0.81–1.00 excellent [37]. Sensitivity, specificity, and negative and positive predictive value were also calculated. A similar analysis was performed for responders according to the different scores of disease activity. For these purposes, patients considering themselves in 'very good' level at baseline were excluded from analyses.

Ability of PESS to detect patients with high disease impact despite remission or LDA

The levels of impact, according to RAID.7 associated with different PESS states in patients who achieved at least LDA in the last visit were analysed and compared through the non-parametric Kruskall–Wallis test (<acceptable, acceptable, >acceptable).

For all analyses, only patients with available data at baseline and follow-up were included. Statistical analysis was performed using the SPSS® software, version 24 and MedCalc. Statistically significant results were assumed when $P < 0.05$ or $P < 0.02$ in case of Bonferroni correction for pairwise comparisons use.

Results

In total, 353 patients were included in the analyses (Table 1). Three-quarters were women (76.8%), with a mean (S.D.) age of 51.9 (13.9) years, and most had long-standing disease [mean (S.D.) 9.9 (9.6) years)]. 96.6% were receiving a biological DMARD (Table 1). Included

Table 1 Socio-demographic and clinical characteristics at start of bDMARDs and follow-up

Characteristic	Baseline	6 months follow-up
Female, *n* (%)	271 (76.8)	–
Age, years	51.9 (13.9)	–
Disease duration years	9.9 (9.6)	–
Glucocorticoids intake, *n* (%)	161 (45.6)	119 (33.7)
csDMARDs, *n* (%)	275 (77.9)	275 (77.9)
bDMARD intake, *n* (%)	341 (96.6)	341 (96.6)
tsDMARD intake, *n* (%)	12 (5.4%)	12 (5.4%)
Disease activity, DAS 28-3vESR	3.1 (1.5)	2.7 (1.2)
High, *n* (%)	69 (19.5)	10 (2.8)
Moderate, *n* (%)	165 (46.7)	95 (26.9)
Low, *n* (%)	48 (13.6)	59 (16.7)
Remission, *n* (%)	71 (20.1)	189 (53.5)
SDAI ($n = 302$)	18.1 (12.7)	8.1 (5.5)
High, *n* (%)	61 (20.2)	14 (4.6)
Moderate, *n* (%)	145 (48)	62 (20.5)
Low, *n* (%)	81 (26.8)	127 (42.1)
Remission, *n* (%)	15 (5.0)	99 (32.8)
CDAI ($n = 302$)	17.0 (12.2)	7.5 (8.1)
High, *n* (%)	73 (24.2)	23 (7.6)
Moderate, *n* (%)	139 (46)	58 (19.2)
Low, *n* (%)	76 (25.2)	117 (38.7)
Remission, *n* (%)	14 (4.6)	104 (34.4)
PGA (0–100)	43.3 (26.8)	27.5 (25.7)
RAID score (0–10)	4.0 (2.2)	2.9 (2.1)
Pain (0–10)	4.0 (2.5)	3.0 (2.3)
Function (0–10)	4.0 (2.6)	2.7 (2.1)
Fatigue (0–10)	4.2 (2.6)	3.1 (2.6)
Sleep disturbance (0–10)	3.5 (2.9)	2.5 (2.6)
Emotional well-being (0–10)	3.4 (2.5)	2.6 (2.3)
Physical well-being (0–10)	4.2 (2.5)	3.4 (2.4)
Coping (0–10)	3.1 (2.4)	2.2 (2.3)

Values are mean (S.D.), unless stated otherwise. csDMARDs: synthetic DMARDs; bDMARD: biological DMARD; tsDMARDs: targeted synthetic DMARDs; DAS28-ESR-3v: DAS-ESR-3 variables; SDAI: Simplified Disease Activity Index; CDAI: Clinical Disease Activity Index; PGA: Patient Global Assessment of Disease Activity; RAID: RA Impact of Disease.

and excluded patients were similar regarding socio-demographic characteristics, disease duration and disease activity at baseline.

Over the 6 months of follow-up, the mean SDAI score decreased from 18.1 (12.7) to 8.1 (5.5). The proportion of patients in high or moderate disease activity (SDAI) decreased from 68.2% to 25.1%, while the proportion of those in remission or LDA increased from 31.8% to 74.9%. The mean global RAID score decreased from 4.0 (2.2) to 2.9 (2.1), with similar changes being observed in all seven domains of impact (Table 1).

Regarding PESS, the proportion of patients in good/very good status increased during the 6 months of follow-up from 18.9% to 42.5%. At the end of follow-up, 54.1% of the patients had improved, while 33.1% maintained the same status and 12.7% of the patients had worsened (Table 2).

Longitudinal construct validity

Patients with improved, stable and worsened PESS over time differed significantly in terms of change in disease activity, and change in combined RAID and each of the RAID.7 domains. Patients who improved their PESS category over time described, on average, a marked reduction of the impact in all domains as well as in disease activity. Patients who got worse in PESS reported an increased impact in all domains in comparison with baseline, despite there being a small reduction in disease activity (Fig. 1).

Changes in PESS, in disease activity and in each of the measures of impact were significantly correlated (Fig. 2).

Moderate to strong correlations were observed (rho = 0.7) between PESS and RAID combined score, RAID pain, function and physical well-being, as well as PGA. Correlations were moderate with RAID fatigue (rho = 0.6) and RAID emotional well-being and coping (rho = 0.5). The weakest correlations (rho = 0.4), but still significant, were found for RAID sleep disturbance and disease activity (supplementary Table S1, available at *Rheumatology* online).

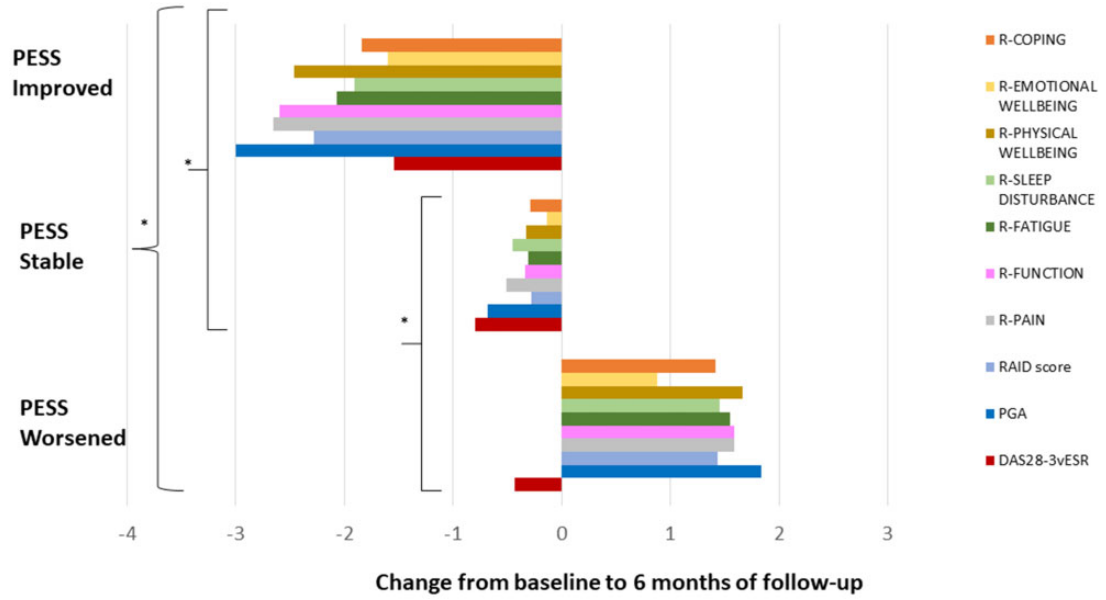
PESS was sensitive to change, with an overall SRM of 0.65 (95% CI: 0.54, 0.76). PESS was clearly more responsive among EULAR (disease activity) responders than non-responders (SRM: 0.91 *vs* 0.26, respectively). Responsiveness was higher for impact, with SRMs of

Table 2 Distribution of patients per PESS category at baseline and 6 months of follow-up

	Follow-up					
Baseline	**Very bad**	**Bad**	**Acceptable**	**Good**	**Very good**	**Total**
Very bad	**2**	10	6	2	4	**24**
Bad	7	**47**	62	30	21	**167**
Acceptable	0	16	**34**	34	11	**95**
Good	1	5	8	**21**	11	**46**
Very good	0	2	3	3	**13**	**21**
Total	**10**	**80**	**113**	**90**	**60**	**353**

Bold values in diagonal line represent patients who remain in the same PESS category.

Fig. 1 Mean change in DAS and impact in patients who improved, worsened or maintained PESS



PGA: Patients Global Assessment of Disease Activity; RAID: Rheumatoid Arthritis Impact of Disease; R: RAID. *P <0.05 (analysis of variance).

1.79 among RAID responders *vs* 0.26 in non-responders.

Overall responsiveness measured by SRM was similar for PESS and PGA (SRM: 0.63, 95% CI: 0.51, 0.74). Among RAID responders the performance was better for PESS (SRM: 1.79, 95% CI: 1.54, 1.99 *vs* 1.65, 95% CI:1.42, 1.85). Among EULAR responders, PESS and PGA showed similar responsiveness (0.91, 95% CI: 0.74, 1.03 *vs* 0.90, 95% CI: 0.71, 1.04).

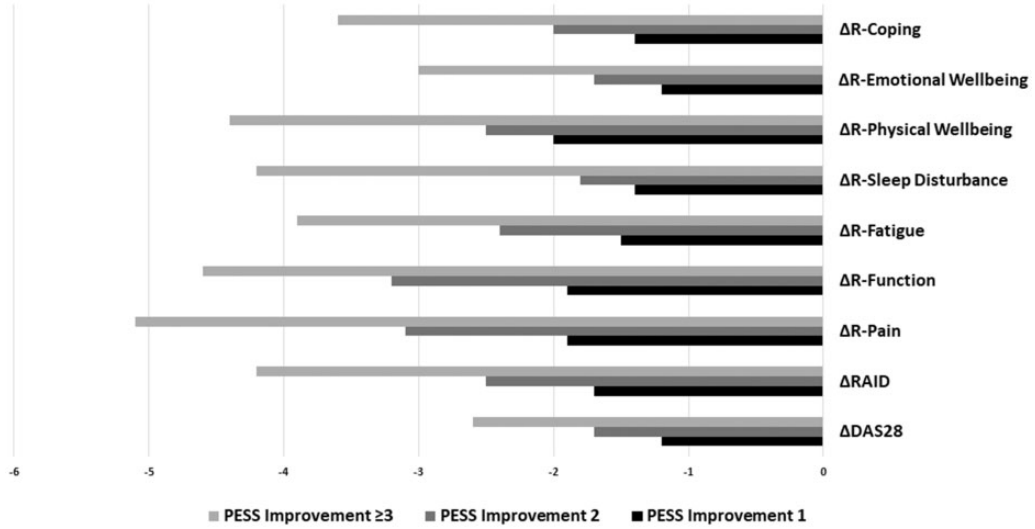
Reliability

Among patients with a stable condition ($-0.6 < \Delta DAS283vESR < 0.6$ and $-2 < \Delta RAID < 2$) ($n = 78$), around 50% remained in the same PESS level during the follow-up period and six patients had a change of >1 level (Table 3). PESS was moderately reliable in this sample according to ICC (0.72, 95% CI: 0.52, 0.83).

Performance of PESS in detecting responders and non-responders

The performance of an improvement of at least one PESS category to identify RAID (non-)responders is presented in Table 4. The diagnostic accuracy [('True positive' + 'True Negative' rates)/Total of patients] was 73%. Over 90% of patients who were RAID responders were also PESS responders (sensitivity); specificity was low: 37.8% of RAID non-responders were PESS responders. Conversely, negative-predictive value indicates that over 90% of those who did not improve at least one level in PESS were also RAID non-responders (negative-predictive value).

Fig. 2 Mean change in DAS and impact according to change in PESS during the follow-up



Numbers represent mean score improvement in global RAID score and each of its domains (0–10 NMR) as well as in DAS according to the PESS amelioration by one (black bars), two (dark grey bars) or three levels (light grey bars). PESS: Patient Experienced Symptom State; RAID: Rheumatoid Arthritis Impact of Disease.

Table 3 PESS scores at baseline and 6 months in patients with stable disease

	Follow-up					
Baseline	**Very bad**	**Bad**	**Acceptable**	**Good**	**Very good**	**Total**
Very bad	**1**	2	0	0	0	**3**
Bad	3	**12**	8	2	0	**25**
Acceptable	0	3	**10**	10	2	**25**
Good	1	1	2	**8**	4	**16**
Very good	0	0	0	0	**9**	**9**
Total	**5**	**18**	**20**	**20**	**15**	**78**

Table 4 Performance of PESS improvement ≥1 as a predictor of RAID and Disease activity response

	EULAR responder	CDAI responder	SDAI responder	RAID responder
Crude agreement	65.4	70.0	70.2	73
Kappa	0.3	0.3	0.3	0.5
Sensitivity	68.1	72.1	72	93.2
Specificity	60.7	66.4	67.2	62.2
PPV	74.9	76.8	77.9	57.6
NPV	52.5	70.8	60	94.3
+LHR	1.7	2.1	2.2	2.46

RAID responder: Δ combined RAID score ≥3 and/or ≥50%; EULAR responder according EULAR response criteria; CDAI responder: ΔCDAI ≥50%; SDAI responder: ΔSDAI ≥50%; PPV: positive predictive value, NPV: negative predictive value; +LHR: positive likelihood ratio; CDAI: Clinical Disease Activity Index; SDAI: Simplified Disease Activity Index; RAID: Rheumatoid Arthritis Impact of Disease.

A similar analysis was performed considering disease activity responders. As expected, the diagnostic accuracy (65.4–70.2%) was lower than for impact, with around 70% of responders also being PESS responders. Among different responder criteria considered, the performancee was lower for EULAR response criteria.

Ability of PESS to detect patients with high disease impact despite remission or LDA

Among patients who achieved low disease activity or remission at 6 months ($n = 248$), 40 maintained a PESS state of bad or very bad. This reflects moderate to high disease impact in the seven domains of RAID (supplementary Fig. S1, available at *Rheumatology* online), with majority of patients (55–72.5%) scoring ≥ 5 in most domains, with the exception of coping (35%).

PESS good or very good is generally associated with low impact, with levels of RAID.7 scores ≤ 2 in at least 75% of patients for most domains, with exception of fatigue and emotional well-being, which have slightly higher impact.

Discussion

This study provides important information for research and especially for clinical management of patients with RA. Our results demonstrate that PESS is very simple to use but also reliable and sensitive to change in patients with RA.

PESS was responsive to treatment and improvement in PESS status over time is associated with a significant improvement in mean impact in all seven domains of RAID and, to a lesser degree, also of disease activity. This confirms that PESS is primarily a measure of disease impact. Correlations with PESS are stronger for the domains of RAID typically associated with disease activity and of highest priority for patients: pain, function and physical well-being [38]. Our study also indicates that PESS is reliable, as it remains stable over time in steady conditions of disease activity and impact. PESS also showed a very good performance in screening for patients who do not respond to treatment in terms of impact. The ability to detect responders and non-responders regarding disease activity is smaller, but this is not the intended use of PESS.

Taken together with our prior observations [27], these results demonstrate that PESS, is feasible, valid, reliable and responsive to change, thus fulfilling all requirements of the OMERACT filter [28].

Our results are supported by the methodological robustness of the study with emphasis on its longitudinal design, large sample size, and the observation of significant change of both disease activity and disease impact during follow-up. This is reinforced by the pragmatic clinical practice setting, without constraining inclusion criteria, stiff protocols or dedicated explanations to patients or physicians. No specific guidance was given to clinicians regarding their prior knowledge of PROs when examining the patient and scoring the PhGA. This probably varied among clinicians and centres. It is unlikely that this would have affected the PhGA and other scores of disease activity, since PhGA is mainly influenced by objective aspects of the disease and only slightly affected by impact [39]. We believe, in any case, that this variability provides welcome representation of reality in regular clinics, as opposed to protocolized clinical trials. The stronger sensitivity to changes in impact rather than disease activity reinforces face validity while incorporating the strong connections between these two dimensions when departing from high disease activity. The stronger associations with the physical domains of impact are in line with evidence that these are more modifiable/responsive to immunosuppressive treatment than social and psychological measures [33, 40–42].

Some limitations should be considered. Patients were recruited in academic centres from a single country (Norway), mostly with longstanding disease, thus calling into question the generalizability of the results. In fact, sociodemographic, clinical and cultural factors, and disease duration have been shown to influence patients' perception of impact [43–46] and of disease activity [43, 47], suggesting that our results need external validation in other cohorts, including early arthritis cohorts. The relatively long interval between the two assessments (6 months as opposed to the usual 2 weeks) may have had a negative effect on the assessment of reliability. A significative proportion of patients were in remission, or at least low disease activity at baseline, which may not reflect clinical practice elsewhere. Only one-third of patients with RA in NOR-DMARD were analysed, essentially due to missing data. Excluded patients were not different regarding socio-demographic aspects, disease duration and disease activity at baseline (data not shown). However, health literacy or healthcare adherence, which may have affected the filling of PROs, were not evaluated, representing an opportunity for selection bias that deserves consideration.

We believe that our findings are relevant for clinical practice as they show that PESS provides a dependable summary view of the disease impact and the patient's perspective by means of a single, simple and easily understood question. The use of the PESS will facilitate communication between patients and physicians, potentially increasing their mutual agreement and providing a better support for the shared decision-making process [48, 49]. We believe that PESS is preferable to PGA as a measure of disease impact for several reasons. Different formulations of the question used to assess PGA are available which are not interchangeable, as they are interpreted and scored differently by patients [50, 51]. The designation of PGA of disease activity is misleading to both patients and healthcare providers, affecting their decisions. Distinguishing between two levels of PESS is certainly more meaningful to patients than separating the abstract concepts conveyed by numbers in PGA, making PGA more difficult to understand and, thus, less reliable [17]. Furthermore, a given score of PGA does not indicate whether the patient is or not personally satisfied with their current condition and to what degree. Additionally, PESS refers to remaining in a given status over time, which is certainly relevant in conveying the patient's perspective in chronic diseases.

PESS could also facilitate and foster the use of PROs in clinical practice as an integral part of disease

evaluation and management. We propose that it might be used as a screening tool to quickly detect patients who have significant burden of the disease and, thus, deserve a detailed domain-driven evaluation, especially after the disease activity has been brought under control. A patient who does not reach at least a 'good' PESS level despite disease remission or LDA should be selected for a discriminative assessment of the causes of persistent impact, so that adequate adjunctive measures can be considered and tailored. A similar approach might be considered in clinical trials. This would provide a simple and reliable evaluation of the overall satisfaction of the patients with the treatment results and allow for the needed assessment of unmet needs from the patient's perspective. A model suggesting a possible flow of care including PESS and RAID.7, in the context of a dual target strategy, has been proposed recently [14]. Naturally, the conditions of practice may advise the application of RAID, without PESS, to all patients in some clinical settings.

In conclusion, PESS showed good psychometric properties, making it suitable to be used in clinical practice and research to assess treatment success from the perspective of the patient, serving person-centred care. Further validation of PESS in other cohorts of RA is warranted.

Acknowledgements

Our most sincere thanks to all patients who contributed data to this study and to the researchers at NOR-DMARD who generously allowed the use of the database.

Funding: No specific funding was received from any funding bodies in the public, commercial or not-for-profit sectors to carry out the work described in this manuscript. NOR-DMARD receive funding from Abbvie, Amgen, Bristol Myers Squibb, Merck Sharp & Dhome, Novartis, Pfizer and Union Chimique Belge Pharma, and consulting fees from Abbvie, Biogen, Celltrion and Gilead.

Disclosure statement: The authors declare no conflict of interest for this study.

T.K.K. reports consulting fees from Abbvie, Biogen, Celltrion, Gilead, Mylan, Novartis, Pfizer, Sandoz, Sanofi; speakers bureaus Amgen, Celltrion, Egis, Evapharma, Ewopharma, Hikma, Oktal, Sandoz, Sanofi.

Data availability statement

The data underlying this article will be shared on reasonable request to the corresponding author.

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

1 Combe B, Landewe R, Daien CI *et al.* 2016 update of the EULAR recommendations for the management of early arthritis. Ann Rheum Dis 2017;76:948–59.

2 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.

3 Singh JA, Saag KG, Bridges SL Jr *et al.* 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Rheumatol 2016;68:1–26.

4 Fautrel B, Alten R, Kirkham B *et al.* Call for action: how to improve use of patient-reported outcomes to guide clinical decision making in rheumatoid arthritis. Rheumatol Int 2018;38:935–47.

5 Brites L, Dinis de Freitas J, Costa F *et al.* Patient-physician discordance in assessment of disease activity in rheumatoid arthritis patients. Acta Reumatol Port 2021;46:103–9.

6 Furu M, Hashimoto M, Ito H *et al.* Discordance and accordance between patient's and physician's assessments in rheumatoid arthritis. Scand J Rheumatol 2014;43:291–5.

7 Kaneko Y, Kuwana M, Kondo H, Takeuchi T. Discordance in global assessments between patient and estimator in patients with newly diagnosed rheumatoid arthritis: associations with progressive joint destruction and functional impairment. J Rheumatol 2014;41:1061–6.

8 Studenic P, Radner H, Smolen JS, Aletaha D. Discrepancies between patients and physicians in their perceptions of rheumatoid arthritis disease activity. Arthritis Rheum 2012;64:2814–23.

9 Ferreira RJO, Santos E, Gossec L, da Silva JAP. The patient global assessment in RA precludes the majority of patients otherwise in remission to reach this status in clinical practice. Should we continue to ignore this? Semin Arthritis Rheum 2020;50:583–5.

10 Ferreira RJO, Welsing PMJ, Jacobs JWG *et al.* Revisiting the use of remission criteria for rheumatoid arthritis by excluding patient global assessment: an individual meta-analysis of 5792 patients. Ann Rheum Dis 2021;80:293–303.

11 Ferreira RJO, Dougados M, Kirwan JR *et al.*; CoimbRA investigators, RAID investigators and COMEDRA investigators. Drivers of patient global assessment in patients with rheumatoid arthritis who are close to remission: an analysis of 1588 patients. Rheumatology (Oxford) 2017;56:1573–8.

12 Ferreira R, Duarte C, Santos EJF, da Silva JAP. Dual-target strategy: fostering person-centered care in rheumatology. Acta Reumatol Port 2021;46:99–102.

13 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.

14 Duarte C, Ferreira RJO, Santos EJF, da Silva JAP. Treating-to-target in rheumatology: theory and practice. Best Pract Res Clin Rheumatol 2022;36:101735.

15 van Tuyl LH, Michaud K. Patient-reported outcomes in rheumatoid arthritis. Rheum Dis Clin North Am 2016;42: 219–37.

16 Steed LG. A critique of coping scales. Aust Psychol 1998;33:193–202.

17 Ferreira RJO, de Wit M, Henriques M *et al.* 'It can't be zero!' Difficulties in completing patient global assessment in rheumatoid arthritis: a mixed methods study. Rheumatology (Oxford) 2020;59:1137–47.

18 Hiligsmann M, Rademacher S, Kaal KJ, Bansback N, Harrison M. The use of routinely collected patient-reported outcome measures in rheumatoid arthritis. Semin Arthritis Rheum 2018;48:357–66.

19 Aletaha D, Nell VP, Stamm T *et al.* Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. Arthritis Res Ther 2005;7:R796–806.

20 Aletaha D, Martinez-Avila J, Kvien TK, Smolen JS. Definition of treatment response in rheumatoid arthritis based on the simplified and the clinical disease activity index. Ann Rheum Dis 2012;71:1190–6.

21 Aletaha D, Neogi T, Silman AJ *et al.* 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2010;69:1580–8.

22 van Riel PLCM. The development of the disease activity score (DAS) and the disease activity score using 28 joint counts (DAS28). Clin Exp Rheumatol 2014;32: S-65-74.

23 Hirsh J, Wood P, Keniston A *et al.* Limited health literacy and patient confusion about rheumatoid arthritis patient global assessments and model disease states. Arthritis Care Res (Hoboken) 2019;71:611–9.

24 Renskers L, van Uden R, Huis AMP *et al.* Comparison of the construct validity and reproducibility of four different types of patient-reported outcome measures (PROMs) in patients with rheumatoid arthritis. Clin Rheumatol 2018; 37:3191–9.

25 Ferreira RJO, Duarte C, Ndosi M *et al.* Suppressing inflammation in rheumatoid arthritis: does patient global assessment blur the target? A practice-based call for a paradigm change. Arthritis Care Res (Hoboken) 2018;70: 369–78.

26 Tubach F, Ravaud P, Beaton D *et al.* Minimal clinically important improvement and patient acceptable symptom state for subjective outcome measures in rheumatic disorders. J Rheumatol 2007;34:1188–93.

27 Duarte C, Santos E, da Silva JAP *et al.* The Patient Experienced Symptom State (PESS): a patient-reported global outcome measure that may better reflect disease remission status. Rheumatology (Oxford) 2020;59: 3458–67.

28 Beaton DE, Maxwell LJ, Shea BJ *et al.* Instrument selection using the OMERACT filter 2.1: the OMERACT methodology. J Rheumatol 2019;46:1028–35.

29 Kvien TK, Heiberg, Lie E, Kaufmann C *et al.* A Norwegian DMARD register: prescriptions of DMARDs and biological agents to patients with inflammatory rheumatic diseases. Clin Exp Rheumatol 2005;23 (5 Suppl 39):S188–94.

30 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.

31 Gossec L, Dougados M, Rincheval N *et al.* Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. Ann Rheum Dis 2009;68:1680–5.

32 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.

33 Duarte C, Santos EJF, Ferreira RJO *et al.* 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. RMD Open 2021;7:e001539.

34 Dougados M, Brault Y, Logeart I *et al.* Defining cut-off values for disease activity states and improvement scores for patient-reported outcomes: the example of the Rheumatoid Arthritis Impact of Disease (RAID). Arthritis Res Ther 2012;14:R129.

35 Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53: 459–68.

36 Fleiss J, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Measur 1973;33: 613–9.

37 McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22:276–82.

38 Heiberg T, Kvien TK. Preferences for improved health examined in 1,024 patients with rheumatoid arthritis: pain has highest priority. Arthritis Rheum 2002;47: 391–7.

39 Desthieux C, Hermet A, Granger B, Fautrel B, Gossec L. Patient-physician discordance in global assessment in rheumatoid arthritis: a systematic literature review with meta-analysis. Arthritis Care Res (Hoboken) 2016;68: 1767–73.

40 Gossec L, Danre A, Combe B *et al.* Improvement in patient-reported outcomes after rituximab in rheumatoid arthritis patients: an open-label assessment of 175 patients. Joint Bone Spine 2015;82:451–4.

41 Smolen JS, Kremer JM, Gaich CL *et al.* Patient-reported outcomes from a randomised phase III study of baricitinib in patients with rheumatoid arthritis and an inadequate response to biological agents (RA-BEACON). Ann Rheum Dis 2017;76:694–700.

42 Boyce EG, Vyas D, Rogan EL, Valle-Oseguera CS, O'Dell KM. Impact of tofacitinib on patient outcomes in rheumatoid arthritis - review of clinical studies. Patient Relat Outcome Meas 2016;7:1–12.

43 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 (Hoboken) 2019;71:1317–25.

44 Hifinger M, Putrik P, Ramiro S *et al.* In rheumatoid arthritis, country of residence has an important influence on fatigue: results from the multinational COMORA study. Rheumatology (Oxford) 2016;55:735–44.

45 Santos EJF, Duarte C, Ferreira RJO *et al.*; 'Promoting Happiness Through Excellence of Care' Group. Determinants of happiness and quality of life in patients with rheumatoid arthritis: a structural equation modelling approach. Ann Rheum Dis 2018;77:1118–24.

46 Duarte C, Santos E, Kvien TK *et al.* Attainment of the Patient-Acceptable Symptom State in 548 patients with rheumatoid arthritis: influence of demographic factors. Joint Bone Spine 2021;88:105071.

47 Putrik P, Ramiro S, Keszei AP *et al.* Lower education and living in countries with lower wealth are associated with higher disease activity in rheumatoid arthritis: results from the multinational COMORA study. Ann Rheum Dis 2016;75:540–6.

48 Barton JL, Trupin L, Tonner C *et al.* English language proficiency, health literacy, and trust in physician are associated with shared decision making in rheumatoid arthritis. J Rheumatol 2014;41:1290–7.

49 Voshaar MJ, Nota I, van de Laar MA, van den Bemt BJ. Patient-centred care in established rheumatoid arthritis. Best Pract Res Clin Rheumatol 2015;29:643–63.

50 Ferreira RJO, Eugenio G, Ndosi M *et al.* Influence of the different "patient global assessment" formulations on disease activity score by different indices in rheumatoid arthritis. Clin Rheumatol 2018;37:1963–9.

51 Gossec L, Kirwan JR, de Wit M, RAID investigators *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.