RHEUMATOLOGY

Rheumatology 2020;0:1–10
doi:10.1093/rheumatology/keaa149

Original article

The Patient Experienced Symptom State (PESS): a patient-reported global outcome measure that may better reflect disease remission status

Cátia Duarte [1,2], Eduardo Santos [1,3,4], José A. P. da Silva[1,2], Eirik Klami Kristianslund[5], Tore K. Kvien[5], Maxime Dougados[6,7,8], Maarten de Wit[9], Laure Gossec [10,11,*] and Turid Heiberg[12,13,*]

Abstract

Objectives. In RA, Patient Acceptable Symptom State assesses disease from the patient's perspective, which does not correspond either to disease remission or to full control of disease impact. This study aims to explore the properties of a novel multilevel Patient Experienced Symptom State (PESS).

Methods. This was a cross-sectional analysis of two datasets of patients with RA. PESS was assessed through the question: 'Consider how your RA has affected you. If you remain in the coming months as you have been the last week, how would you rate your condition?', with five levels (from 'very bad' to 'very good'). Construct validity of PESS was assessed against validated disease activity [DAS28, Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI)] and impact measures [RA Impact of Disease (RAID) and modified HAQ]. Multiple pair-wise comparisons between groups and receiver-operating characteristic curves with Youden Index were performed.

Results. A total of 1407 patients [74% female, mean (S.D.) age 53.5 (13.4) years, mean disease duration 14.3 (12.0) years and mean DAS28 3.0 (1.5)] were analysed. Overall, 16.3% considered themselves as being in 'very good', 21.6% in 'good' and 31.9% in 'acceptable' state. Disease activity and impact measures differed significantly across the five levels ($P < 0.01$). Cut-off values corresponding to 'good' and 'very good' PESS states were in the range of low disease activity/remission (for 'good' and 'very good': DAS28-ESR-4v $\leq$2.6/$\leq$2.3; CDAI $\leq$5.0/$\leq$3.1; SDAI $\leq$5.1/$\leq$3.8, respectively) and very low disease impact (RAID domains all $\leq$1).

Conclusion. PESS 'very good' status corresponds to currently recommended targets for RA management and reflects full control of disease impact. PESS appears to be an easy-to-use and relevant measure in the evaluation of patients with RA.

Key words: rheumatoid arthritis, patient-reported outcomes, Patient Experienced Symptom State, disease activity, thresholds

Rheumatology key messages

- Patient Experienced Symptom State (PESS) is a single question to measure patients' current RA status.
- Levels defined by PESS had a strong association with RA disease activity and impact.
- PESS 'very good' status corresponds to current targets of disease activity and impact in RA.

[1]Department of Rheumatology, Centro Hospitalar e Universitário de Coimbra, Coimbra, [2]Coimbra Institute for Clinical and Biomedical Research - Faculty of Medicine, University of Coimbra, Coimbra, [3]Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, [4]Health Sciences Research Unit: Nursing, Nursing School of Coimbra, Coimbra, Portugal, [5]Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway, [6]Université de Paris, Paris, [7]Department of Rheumatology, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris, Paris, [8]INSERM (U1153): Clinical Epidemiology and Biostatistics, PRES Sorbonne Paris-Cité, Paris, France, [9]Patient Research Partner, Amsterdam, The Netherlands, [10]INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, Sorbonne Université, Paris, [11]Rheumatology Department, Pitié Salpêtrière Hospital, Sorbone Univeristé-AP-HP, Paris, France, [12]Faculty of Health and Welfare, Østfold University College, Halden and [13]Regional Research Support, Oslo University Hospital, Oslo, Norway

Submitted 28 November 2019; accepted 28 February 2020

Correspondence to: Cátia Duarte, Rheumatology Department, Centro Hospitalar e Universitário de Coimbra, Praceta Mota Pinto3000-004 Coimbra, Portugal. E-mail: catiacmduarte@gmail.com

*Laure Gossec and Turid Heiberg contributed equally to this paper.

CLINICAL SCIENCE

Introduction

Novel therapeutic agents and treatment strategies in RA focus on early intervention and tight control. The establishment of guiding targets for disease management like remission or low disease activity may prevent structural damage and improve long-term function [1–3].

It is essential that these goals, defined by control of inflammation and damage, correspond to desirable patient experiences, i.e. effective control of disease impact upon patients' lives. For both physicians and patients, health-related quality of life is the key overarching objective of medical practice, but this is not necessarily conveyed by measures focused on the disease process. Accordingly, the assessment of RA from the patient's perspective has warranted greater attention over recent years [4, 5].

The Patient Acceptable Symptom State (PASS) [6], defined as the highest level of regular symptom intensity deemed 'acceptable' by patients, is recognized as a clinically relevant patient-reported outcome (PRO) that facilitates the interpretation of the results of clinical trials from the patient's perspective [7]. However, evidence suggests that PASS status, assessed through a binary outcome, may not be sufficiently ambitious, as it is reached by many patients with suboptimal control of the inflammatory process, which may result in structural damage and long-term loss of function [6, 8–10]. The definition of a patient-relevant acceptable state should, ideally, be consistent with the current recommendations that establish remission as the target of treatment whenever possible [1, 2].

Additionally, patients in PASS frequently report poor levels of health in important domains such as pain, fatigue and health-related quality of life [8, 11, 12]. It seems that the expression 'acceptable' is interpreted as the maximum that the patient could tolerate or realistically expect as possible, as opposed to what they would deem 'desirable' or 'satisfying'. If the progress in RA management is to be translated into patient-relevant improvements, we should aim for a more stringent measure of patient satisfaction, an aspirational but still a realistic target for a substantial group of patients. Being in an 'acceptable' state should not be a sufficient target, in the same way that reaching moderate disease activity is not.

The considerations above regarding the limitations of binary PASS led us to hypothesize that the symptom state experienced by patients should be targeted to a higher level of ambition than a merely 'acceptable' status, and probably a 'very good' status could be more valuable than a just 'good' or 'acceptable'. To test our hypothesis, the single binary question of PASS was reformulated into a multilevel question using a 5-level Likert scale, ranging from 'very bad' to 'very good': the multilevel Patient Experienced Symptom State (PESS) [8]. This study aimed to assess the feasibility and construct validity of PESS, and also to identify levels of disease activity and disease impact corresponding to different levels of PESS.

Methods

Study design and patients

We performed a cross-sectional analysis, combining two datasets of patients with RA: the RA Impact of Disease (RAID) study [13], an international multicentre study that led to the initial development of the RAID score, and the Norwegian DMARD (NOR-DMARD) registry [14], a five-centre treatment register and longitudinal observational study that includes consecutive patients with inflammatory rheumatic joint diseases. For these cross-sectional analyses, we used the baseline visit from the RAID study and the last visit available from the longitudinal NOR-DMARD registry. Patients with missing data regarding PESS, DAS or RAID score were excluded.

This study was conducted with approval of the RAID study and the NOR-DMARD registry, the two databases used in our analysis. The RAID study was approved by the ethics committees in the participating countries. 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.

Data collection

Gender, age, disease duration, ESR (mm/h) and CRP (mg/l) were collected. Swollen joint count (SJC) (28 joints) and tender joint count (TJC) (28 joints) were performed and registered by rheumatologists or research nurses. Patient Global Assessment of Disease Activity (PGA) was assessed through a visual analogic scale 0–100 mm (where 0 corresponds to the best state and 100 to the worst) using the following phrasing 'Considering all the ways your arthritis has affected you, how do you feel your arthritis is today?' in NOR-DMARD and 'Considering all the ways that your illness and health conditions affect you at this time, how do you feel?' in the RAID study. Physician Global Assessment of Disease Activity was assessed through a 0–10 numeric rating scale in the RAID study and through a visual analogic scale 0–100 mm in NOR-DMARD. Both PGA and Physician Global Assessment of Disease Activity were converted into a 0–10 scale for Clinical Disease Activity Index (CDAI) and Simplified Disease Activity Index (SDAI) calculation.

PASS and PESS

PASS was assessed through the following question 'Consider how your RA has affected you during the last week. If you remain in the coming months as you have been in the last week, would this be acceptable or unacceptable?', with dichotomous responses of 'acceptable' or 'unacceptable'.

PESS was developed by the authors and 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 5-level Likert scale response ('very bad', 'bad', 'acceptable', 'good' and 'very good'). In the RAID study, a slightly different formulation was applied: 'Think about all the ways your RA has affected you during the last week. If you were to remain for the next few months as you were during the last week, how would you rate this state?' The PESS phrasing differed slightly in the attribution of symptoms: in NOR-DMARD the broader term ('rheumatic disease') was used to cover the variety of inflammatory rheumatic diseases included in this registry, while in the RAID study the attribution term was 'rheumatoid arthritis'. However, as only patients with RA were included in our study, we believe the two terms refer to the same disease and therefore this allows for the combination of the results. The PESS question was answered in each national language, after translation by the principal investigator and the patient research partner.

Measures of disease activity

Different composite indices were used to assess disease activity. DAS based on ESR (DAS28-ESR-4v or -3v) were calculated [15] and categorized as: high disease activity (HDA) $\geq$5.1; moderate disease activity (MDA) ($\geq$3.2 to $<$5.1); low disease activity (LDA) (2.6 to $<$3.2); and remission ($<$2.6) [16].

The SDAI [17] employs a linear sum of five untransformed, unweighted variables, including SJC28 and TJC28, patient and investigator global assessments of disease activity on a 0–10 scale, and CRP (mg/dl). Thresholds for separating remission, LDA, MDA and HDA are 3.3, 11 and 26, respectively [18].

The CDAI [19] is a modification of the SDAI without laboratory evaluation (CRP), to allow immediate clinical assessment. Thresholds for separating remission, LDA and MDA are 2.8, 10 and 22, respectively [20].

Remission was defined according to the ACR/EULAR Boolean criteria as TJC28 $\leq$1, SJC28 $\leq$1, CRP (mg/dl) $\leq$1 and PGA $\leq$1/10 [21]. Given the potential association between PGA and PASS [11], an additional analysis of a 3v-remission defined as TJC28 $\leq$1, SJC28 $\leq$1 and CRP (mg/dl) $\leq$1 [22] was also performed.

Impact measures

The impact of RA was addressed through the modified HAQ (mHAQ) [23] and the RAID score [13, 24]. mHAQ comprises eight questions, one from each category, rated on a 4-point Likert scale with a score range from 0–3. Higher scores indicate worse function and greater disability. mHAQ scores $<$0.3 are considered normal. It has been proposed that mHAQ scores can be divided into three categories of mild (mHAQ $<$1.3), moderate (1.3 $\leq$ mHAQ $\leq$ 1.8) and severe (mHAQ $>$1.8) functional losses [25].

The RAID score [13, 24] is a composite measure for RA that reflects the patient's perception of the impact of disease on seven domains of health (pain, fatigue, physical function, sleep disturbance, emotional well-being, physical well-being and coping). Each domain is assessed through a single question answered on a 0–10 numerical rating scale (0 corresponds to the best state and 10 to the worst). Scores of the individual domains were collected and the RAID score was computed according to the proposed algorithm [13, 26, 27].

Statistical analyses

Descriptive characteristics are presented as means (S.D.) for continuous variables and as proportions (%) for categorical variables. There was no imputation of missing data.

To evaluate feasibility, the response rate was calculated using the overall sample (1961 patients). To assess construct validity, mean levels of disease activity and impact were compared across the five levels of PESS using the non-parametric Kruskal–Wallis test, given the non-normal distribution. Proportions of disease activity categories were also compared across the five PESS categories using the χ^2 test. Within each of the five of levels of PESS, comparison of disease activity, impact and proportions of disease activity categories between patients with short ($\leq$2 years) and long disease duration ($>$2 years) was performed through non-parametric Mann–Whitney test and the χ^2 test, as appropriate.

The ordinal scale of PESS was dichotomized at different levels of split: 'very good *vs* $<$very good'; '$\geq$good *vs* $<$good'; '$\geq$acceptable *vs* $<$acceptable' and '$\geq$bad *vs* very bad'. This approach allowed us to use the entire cohort in the definition of cut-offs, so as to maximize precision and clinical relevance [28].

Cut-offs of disease activity scores and impact measures corresponding to the different PESS levels were calculated according to two different methods: (i) receiver operating characteristics curves were performed and used to identify, through the Youden Index (J), the cut-offs of disease activity scores and impact measures with the best trade-off between sensitivity and specificity regarding the dichotomized PESS classification [29]; and (ii) the 75th percentile method, which is defined as the cut-off of the disease activity scores or impact measures correctly classifying 75% of the patients in the targeted category of PESS. The sensitivities and specificities of the thresholds obtained by the 75th percentile method were calculated against being in the targeted category or not [9]. Sensitivity was established as the percentage of patients in a given PESS category (e.g. 'very good') whose score in the legacy variable (e.g. DAS28) was below the defined threshold. Specificity was defined as the percentage of patients outside the targeted PESS category whose score in the legacy variable was higher than the defined threshold.

To examine the robustness of the thresholds obtained by the Youden Index (J), we repeated the analysis in subgroups of patients with short and long disease duration. The thresholds obtained by Youden Index (J) for PASS/NOT PASS and for the PESS dichotomy '$\geq$acceptable *vs* $<$acceptable' were visually compared.

As confirmatory analyses, we performed a two-by-two cross-tabulation of the proportions of cases in '$\geq$good

vs <good' in PESS against equivalent dichotomizations in composite DAS (remission/LDA *vs* MDA/HDA for DAS28-ESR-4v, CDAI, SDAI and remission/non-remission for the ACR/EULAR Boolean definition). The agreement between disease activity and PESS categories was assessed by 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 [30].

Given the expected association between PGA and patients' perception of their status [11, 31, 32], a sensitivity analysis was then performed considering the DAS28-ESR without PGA (DAS28-ESR-3v) and the ACR/EULAR definition excluding PGA ('3v-remission'). Statistical analyses were performed using SPSS software version 24 and Medcalc software version 18.11.3. Statistically significant effects were assumed for $P < 0.05$.

Results

Patient characteristics

Altogether, 554 of the 1961 patients were excluded due missing data in at least one variable of interest. The response rate to the PESS question was 98.5%, revealing excellent feasibility. In total, data from 1407 patients were analysed (Table 1): 198 from the RAID study and 1209 patients from the NOR-DMARD registry. Three-quarters were women (74%), with a mean (S.D.) age of 53.5 (13.4) years, and most had longstanding disease [mean (S.D.) 14.3 (12.0 years)]. Most of the patients were in LDA or in remission, considering all composite scores, and 302 (21.5%) fulfilled the Boolean ACR/EULAR criteria for remission. Impact of disease, as reflected by mHAQ, RAID score and its individual seven items (RAID7i), was mild to moderate (RAID7i mean scores 2.0–3.7). Patients from the NOR-DMARD registry presented lower levels of disease activity and impact. (Table 1).

PESS

When asked to consider the five levels of PESS, 983 (69.8%) rated their state as acceptable or better. In total, 304 (21.6%) and 230 (16.3%) considered themselves to be in a 'good' or in a 'very good' status, respectively. Patients rating their status as a 'very good' were younger and had shorter disease duration, but gender differences were not observed (Table 2). Mean scores of different measures of disease activity and impact decreased progressively from the 'very bad' to 'very good' status, while the percentage of patients in remission increased progressively (Table 2). Patients reporting a 'very good' status were in the remission range for all composite scores of DAS and 172 (74.8%) fulfilled the Boolean ACR/EULAR criteria for remission.

Patients in 'very bad' and 'bad' states were in the range of MDA to HDA in all composite scores, had mild functional impact according to mHAQ [mean 1.3 (0.6) and 0.8 (0.4), respectively] and moderate to high impact of disease according to RAID, with almost all individual items scoring ≥5 on average. The rate of remission according to the Boolean ACR/EULAR was close to zero in these two groups. Similar mean levels of disease activity and impact in each PESS level were found when comparing patients with different years of disease duration (data not shown).

Thresholds of DAS and impact measures corresponding to the different levels of PESS

Thresholds of disease activity and global impact corresponding to PESS dichotomized cut-offs, calculated according to the Youden Index, are presented in Table 3. Thresholds of disease activity corresponding to a 'very good' status were in the remission to LDA range for all disease activity composite scores and the corresponding level of RAID score was 1.3. Thresholds of disease activity for PESS '≥good' were in the range of LDA, and for RAID score 2.6. Thresholds of disease activity for PESS '≥acceptable' (equivalent to PASS, as seen above) were in the range of LDA to MDA and corresponded to a RAID score of 4.2. For individual items of RAID sensitivities ranged from 77% (RAID Coping) to 90% (RAID Function), and specificities from 70 to 92% with the lowest value for DAS28 and the highest for PGA. The areas under the curve were ≥0.8 for all outcome measures (supplementary Table S1, available at *Rheumatology* online). Sensitivity analysis considering DAS28-ESR with and without PGA resulted in similar cut-offs for the PESS levels (Table 3). The thresholds were robust across subgroups of patients with short and long disease duration, with small fluctuations within the same range of disease activity level and impact (data not shown).

Agreement between PESS status and disease activity levels

The level of crude agreement between 'very good' PESS status and remission was excellent (>0.8, except for DAS28) (Table 4). Kappa levels were in the range of moderate agreement. The crude agreement between '≥good' PESS status and disease activity at the level of remission or LDA, was 'good' for all scores of disease activity, with the best crude agreement observed with the ACR/EULAR remission criteria and the worst with DAS28 (Table 5). The highest crude agreement was observed between the 'very good' status and remission according to the ACR/EULAR Boolean criteria (86.6%).

Considering the ACR/EULAR definition of remission without PGA (3v-ACR/EULAR remission definition), the level of agreement with '≥good' PESS status decreased as expected [from 78.7 to 69.6% (k = 0.38; 95% CI: 0.33, 0.43)]. The same was observed for 'very good' PESS status [from 86.6 to 62.7% (k = 0.22; 95% CI: 0.18, 0.26)].

Comparison between PASS and PESS

Among the 1407 included patients, 949 (67.6%) considered themselves to be in a PASS condition, which is

Table 1 Sociodemographic and clinical characteristics of RA patients

Characteristic	All (n = 1407)	RAID study (n = 198)	NOR-DMARD study (n = 1209)
Female, n (%)	1040 (74.0)	156 (79.2)	884 (73.1)
Age, years	53.5 (13.4)	54.2 (12)	53.3 (13.6)
Disease duration years	14.3 (12.0)	13.3 (11.2)	14.5 (10.4)
5-Level acceptance			
Very bad, n (%)	56 (4.0)	20 (10.1)	36 (3.0)
Bad, n (%)	308 (26.2)	59 (29.8)	309 (25.6)
Acceptable, n (%)	449 (31.9)	64 (32.3)	385 (31.8)
Good, n (%)	304 (21.6)	36 (17.2)	270 (22.3)
Very good, n (%)	230 (16.3)	21 (10.6)	209 (17.3)
DAS28-ESR-4v	3.0 (1.5)	4.3 (1.6)	2.8 (1.4)
High, n (%)	161 (11.4)	65 (32.8	96 (7.9)
Moderate, n (%)	386 (27.4)	76 (38.4)	310 (25.6)
Low, n (%)	207 (14.7)	21 (10.6)	186 (15.4)
Remission, n (%)	653 (46.7)	36 (18.2)	617 (51.0)
SDAI	10.6 (7.2)	18.8 (14.1)	9.3 (9.3)
High, n (%)	140 (10.0)	52 (26.3)	89 (7.3)
Moderate, n (%)	354 (25.2)	72 (36.4)	282 (23.3)
Low, n (%)	507 (36.1)	61 (30.8)	446 (36.9)
Remission, n (%)	405 (28.8)	13 (6.6)	392(32.5)
CDAI	9.9 (6.7)	17.9 (13.5)	8.6 (8.9)
High, n (%)	177 (12.6)	60 (30.3)	117 (9.7)
Moderate, n (%)	336 (23.9)	66 (33.3)	270 (22.3)
Low, n (%)	493 (35.0)	59 (29.8)	434(35.9)
ACR/EULAR Boolean remission, n (%)	302 (21.5)	11 (5.6)	291(24.1)
PGA	32.8 (26.3)	50.0 (23.7)	31.4 (26.5)
mHAQ (0–3)	0.5 (0.4)	0.6 (0.6)	0.4 (0.5)
RAID Score (0–10)	3.3 (3.0)	4.4 (2.1)	3.2 (2.3)
Pain (0–10)	3.7 (3.0)	4.9 (2.6)	3.5 (2.5)
Function (0–10)	3.3 (3.0)	4.7 (2.6)	3.1 (2.6)
Fatigue (0–10)	3.7 (3.0)	4.6 (2.7)	3.6 (2.7)
Sleep Disturbance (0–10)	3.1 (2.9)	4.0 (3.0)	2.9 (2.9)
Emotional Well-being (0–10)	3.0 (2.5)	4.5 (2.4)	3.4 (2.5)
Physical Well-being (0–10)	3.0 (2.5)	3.8 (2.4)	3.0 (2.5)
Coping (0–10)	2.0 (2.4)	3.9 (2.4)	2.5 (2.3)

Mean (S.D.) for continuous variables, n (%) for categorical variables. DAS28-ESR-4v: DAS-ESR-4 variables; SDAI: Simplified Disease Activity Index; CDAI: Clinical Disease Activity Index; mHAQ: modified HAQ; RAID: Rheumatoid Arthritis Impact of Disease.

very similar to the proportion of patients in 'at least acceptable' PESS (69.8%), with an excellent agreement between the two instruments [crude agreement = 94.8%; k = 0.81 (95% CI: 0.78, 0.85)]. Patients scoring in the intermediate PESS categories vary, as expected, in their PASS categorization (supplementary Fig. S1, available at *Rheumatology* online).

The cut-offs for 'acceptable' status were in the range of LDA to MDA (for DAS28) and medium impact (RAID scores of 3–5), thus very similar to those corresponding to 'at least acceptable' PESS (supplementary Table S2, available at *Rheumatology* online) and higher than the thresholds corresponding to 'good' and 'very good' PESS (remission and impact in all domains $\leq$1).

Discussion

This study introduces a new outcome measure, PESS, which includes five different levels of patient satisfaction with their current experience with RA, as opposed to the dichotomous concept of PASS. The data presented demonstrate that PESS levels were strongly related both to disease activity and impact measures, thus confirming its construct validity. Being in a 'very good' PESS status corresponds to a level of disease activity and impact that is more consistent with achieving desirable quality of life and prevention of structural damage in the long-term, than conveyed by PASS. In fact, cut-offs corresponding to a 'very good' status were in the range of LDA/remission for all composite scores of disease activity. The crude agreement between 'very good' PESS and remission according the ACR/EULAR Boolean criteria, the most stringent criteria [33], was excellent

Table 2 Comparison of sociodemographic, disease activity and impact across the five levels of Patient Experienced Symptom State

Characteristic	Very-bad (*n* = 56)	Bad (*n* = 368)	Acceptable (*n* = 449)	Good (*n* = 304)	Very good (*n* = 230)	*P*-value
Female, *n* (%)	44 (78.6)	286 (77.9)	324 (72.2)	224 (73.7)	162 (70.4)	0.21[b]
Age, years	52.7 (14.4)	53.4 (13.3)	55.7 (12.8)	53.6 (13.0)	49.3 (14.1)	<0.01[a]
Disease duration years	15.4 (13.0)	13.5 (10.8)	15.5 (11.0)	15.3 (10.0)	12.0 (8.8)	<0.01[a]
DAS28-ESR-3v	3.9 (1.6)	3.7 (1.4)	2.8 (1.1)	2.3 (0.9)	1.9 (0.9)	<0.01[a]
DAS28-ESR-4v	4.6 (1.6)	4.2 (1.4)	3.0 (1.2)	2.2 (1.0)	1.7 (0.9)	<0.01[a]
High, *n* (%)	24 (42.9)	107 (29.1)	24 (5.3)	5 (1.6)	1 (0.4)	<0.01[b]
Moderate, *n* (%)	21 (37.5)	164 (44.6)	147 (32.7)	41 (13.5)	13 (5.7)	
Low, *n* (%)	5 (8.0)	49 (13.3)	92 (20.5)	43 (14.1)	18 (7.8)	
Remission, *n* (%)	6 (10.7)	48 (13.0)	186 (41.4)	215 (70.7)	198 (86.1)	
SDAI	23.8 (13.3)	19.3 (11.5)	10.0 (7.7)	4.9 (5.4)	2.0 (3.1)	<0.01[a]
High, *n* (%)	19 (33.9)	96 (26.1)	21 (5.4)	3 (1.0)	1 (0.4)	<0.01[b]
Moderate, *n* (%)	30 (53.6)	169 (45.9)	129 (29.3)	24 (7.6)	3 (1.3)	
Low, *n* (%)	7 (12.5)	96 (26.1)	241 (53.7)	124 (40.9)	39 (17.0)	
Remission, *n* (%)	0	7 (1.9)	58 (12.9)	153 (50.5)	187 (81.3)	
CDAI	22.4 (12.5)	18.2 (11.0)	9.5 (7.5)	4.5 (5.1)	1.7 (2.9)	<0.01[a]
High, *n* (%)	23 (41.1)	113 (30.7)	37 (8.2)	3 (1.0)	1 (0.4)	<0.01[b]
Moderate, *n* (%)	26 (46.4)	159 (43.2)	122 (27.2)	26 (8.6)	3 (1.3)	
Low, *n* (%)	7 (12.5)	89 (24.2)	232 (51.7)	127 (41.8)	38(1.5)	
Remission, *n* (%)	0	7 (1.9)	58 (12.9)	148 (48.7)	188 (81.7)	
ACR/EULAR Remission, *n* (%)	0	2 (0.5)	27 (6.0)	101 (33.2)	172 (74.8)	<0.01[b]
PGA	75.4 (23.2)	57.6 (19.1)	34.0 (17.4)	14.6 (12.8)	4.3 (7.1)	<0.01[a]
mHAQ	1.3 (0.6)	0.8 (0.4)	0.5 (0.4)	0.2 (0.3)	0.1 (0.1)	<0.01[a]
RAID score	7.3 (1.5)	5.5 (1.7)	3.5 (1.4)	1.8 (1.1)	0.6 (0.9)	<0.01[a]
Pain	7.9 (1.8)	6.0 (1.8)	3.7 (1.6)	2.1 (1.3)	0.8 (1.0)	<0.01[a]
Function	8.0 (1.4)	5.5 (2.0)	3.5 (1.8)	1.6 (1.4)	0.5 (0.9)	<0.01[a]
Fatigue	7.3 (2.2)	5.9 (2.2)	4.0 (2.1)	2. (1.9)	0.9 (1.5)	<0.01[a]
Sleep Disturbance	6.8 (2.6)	5.3 (2.7)	3.3 (2.5)	1.4 (1.8)	0.5 (1.2)	<0.01[a]
Emotional Well-being	7.8 (1.9)	5.8 (1.8)	3.8 (1.6)	2.0 (1.4)	0.6 (1.3)	<0.01[a]
Physical Well-being	6.5 (2.6)	5.0 (2.2)	3.2 (1.8)	1.7 (1.6)	0.4 (1.1)	<0.01[a]
Coping	6.4 (2.0)	4.6 (2.2)	2.9 (1.7)	1.2 (1.4)	0.6 (0.9)	<0.01[a]

Mean (S.D.) for continuous variables, *n* (%) for categorical variables. [a]Kruskal–Wallis test; [b]Chi-squared test. PGA: Patient Global Assessment; DAS28-ESR-4v, Disease Activity Score, 28 joint count-ESR-4 variables; SDAI: Simplified Disease Activity Index; CDAI: Clinical Disease Activity Index; mHAQ: modified Health Assessment Questionnaire; RAID: RA Impact of Disease.

Table 3 Thresholds of disease activity and impact scores corresponding to PESS levels defined by Youden Index

Disease activity level	≥Bad *vs* very bad (Se/Sp)	≥Acceptable *vs* <acceptable (Se/Sp)	≥Good *vs* <good (Se/Sp)	Very good *vs* <very good (Se/Sp)
DAS 28-ESR-4v	4.1 (78/68)	3.3 (80/73)	2.6 (76/74)	2.3 (79/70)
DAS28-ESR-3v	3.3 (68/69)	2.9 (71/71)	2.6 (75/64)	2.4 (75/63)
SDAI	11.5 (68/88)	7.8 (72/90)	5.1 (79/84)	3.8 (85/76)
CDAI	10.1 (66/88)	7.7 (74/88)	5.0 (82/82)	3.1 (85/80)
PGA	67.0 (8782)	37.0 (81/86)	20.0 (88/86)	7.0 (85/90)
mHAQ	0.8 (80/88)	0.4 (73/82)	0.3 (85/73)	0.0 (76/82)
RAID Score	5.6 (83/91)	4.2 (86/81)	2.6 (87/84)	1.3 (89/89)

PESS: Patient Experienced Symptom state; DAS28-ESR-4v, DAS 28 joint count-ESR-4 variables; DAS28-ESR-3v: DAS 28 joint count-ESR-3 variables; SDAI: Simplified Disease Activity Index; CDAI: Clinical Disease Activity Index; mHAQ: modified HAQ; RAID: RA Impact of Disease; Se: sensitivity; Sp: specificity.

(>80%). These results indicate that 'very good' PESS status corresponds to a degree of disease/inflammation control, which is coherent with the current recommendations for RA management [1, 2].

Our results showed, as expected, a good correlation between PESS and other PROs, particularly with PGA [11, 32]. Therefore, we performed sensitivity analyses to compare PESS to legacy disease activity indices,

TABLE 4 Agreement between very-good PESS status and remission according different composite indices for disease activity

PESS very-good *vs* remission	a	b	c	d	Crude agreement (%)	Kappa (95% CI)
DAS28-ESR-4v*	198	32	455	722	65.4	0.27 (0.23, 0.31)
SDAI*	187	43	218	964	81.4	0.48 (0.43, 0.53)
CDAI*	188	42	213	964	81.9	0.49 (0.44, 0.54)
ACR/EULAR Boolean Criteria*	172	58	130	1047	86.6	0.57 (0.51, 0.62)

a: *N* of patients in 'very good' status and in remission; b: *N* of patients in 'very good' status and not in remission disease activity; c: *N* of patients in '<very good' status and in remission disease activity; d: *N* of patients in '<very good' status and not in remission disease activity. Crude agreement: (a + d)/total. PESS: Patient Experienced Symptom state; DAS28-ESR-4v: DAS 28 joint count-ESR-4 variables; CDAI: Clinical Disease Activity Index; SDAI: Simplified disease Activity Index. *All indices included Patient Global Assessment.

TABLE 5 Agreement between '≥good' PESS status and remission/low disease activity according different indices for disease activity

PESS very good/good *vs* remission/LDA	a	b	c	d	Crude agreement (%)	Kappa (95% CI)
DAS28-ESR-4v*	474	60	386	487	68.3	0.40 (0.36, 0.44)
SDAI*	503	30	469	464	68.7	0.42 (0.38, 0.46)
CDAI*	501	33	393	480	75.4	0.43 (0.39, 0.47)
ACR/EULAR Boolean Criteria*	263	261	29	844	78.7	0.52 (0.48, 0.57)

a: *N* of patients in '≥good' status and in remission or low disease activity; b: *N* of patients in '≥good' status and in moderate or high disease activity; c: *N* of patients in '<good' status and in remission or low disease activity; d: *N* of patients not in '<good' status and in moderate or high disease activity. Crude agreement: (a + d)/total. PESS: Patient Experienced Symptom state; DAS28-ESR-4v: DAS 28 joint count-ESR-4 variables; CDAI: Clinical Disease Activity Index; SDAI: Simplified Disease Activity Index; LDA: low disease activity. *All indices include Patient Global Assessment.

excluding PGA, and the results were very similar, which from our perspective reinforces the validity of the PESS.

The results also indicate that a 'very good' PESS status corresponds to low impact of disease on mHAQ (0) and each of the seven domains of health addressed by the RAID score (scores ≤1). This is further supported by the high areas under the curve (>80%) for all outcome measures, suggesting that PESS adequately captures the overall impact of RA. These findings also suggest that PESS could be more suitable to capture the overall impact of RA than PGA, which is more influenced by pain, fatigue and function than by disease activity [31].

We recognize that the PESS and PGA are both single questions that address the same concept of general status. However, we believe that the PESS may be more patient relevant: firstly because the wording of the PESS refers to remaining in a given state over time, which in a chronic disease is certainly relevant, and secondly because the Likert categories of the PESS make it easier to fill in, thus minimizing variability, than a continuous score, which patients often have difficulties with [34–36].

On the other hand, being assessed through a single question, PESS can be more suitable and feasible to evaluate the overall impact of RA in patients perspective than a complex, multi-item and scored PRO measure, which should be reserved for patients who fail to achieve the 'very good' PESS status in order to identify the underlying reasons for that.

In this study, results were similar when considering patients with different years of disease duration, suggesting that PESS is robust and reinforcing its validity. The percentage of patients who considered themselves in an 'at least acceptable' status (~70%) in our study population was similar to that reported in most previous studies reflecting clinical practice [8, 9, 11, 12]. The cut-offs of DAS28, CDAI and PGA corresponding to PESS '≥acceptable' status are, generally, in good agreement with previously published reports [8, 37] using PASS. Regarding the RAID score, our cut-offs were similar to those described in a previous publication [12], and lower than reported in two other studies [10, 38]. These similarities with studies that, like our own, convey real world data, support the adequacy of our sample for clinical practice. The strong representation of patients in MDA and LDA in our sample makes it especially adequate for our purposes, because these patients are closer to the targets we are aiming to refine. Among the different methods recommended to define cut-offs, we selected the receiver-operating characteristic curve analysis and the Youden Index (J), which allows the selection of the threshold that provides the best possible compromise between sensitivity and specificity for each investigated outcome, while other methods favour particularly sensitivity (i.e. 75th percentile) or specificity (i.e. fixed specificity) [38]. Even so, the results obtained through the Youden Index (J) were confirmed by the 75th percentile

technique, which can be considered a strength of our study.

Some limitations must be considered while interpreting our results. Despite the large sample size, most patients in both cohorts were recruited in academic centres, mostly from Norway, thus calling into question the generalizability of the results. Previous studies in RA have reported consistent cut-off values of disease activity corresponding to PASS in different countries [9], age or disease duration [8, 10] but with controversial results regarding gender [8, 10]. On the other hand, the potential influence of sociodemographic and cultural factors upon patients' perception of impact [39–41] and even upon the assessment of disease activity [39, 42] is recognized, and this suggests that our results need external validation. Similarly, comorbidities known to have a significant impact in PGA and therefore, on the different disease activity scores, were not assessed in this study [35]. Patient partners were not involved in the development of PESS. However, when PESS was developed in 2009, the involvement of patients in research was neither recommended nor frequent [43, 44]. To overcome this limitation, a patient research partner was involved in the current evaluation of PESS. Another limitation resides in the cross-sectional design, which does not allow the evaluation of the stability of these levels over time and their response to change in disease activity and collateral life-events.

The relevance of our findings for clinical practice deserves consideration. A 'very good' PESS status provides a more stringent definition of target status than PASS and can be considered a more appropriate treatment target both from the physician's and from the patient's perspective, given the advances and current recommendations in RA management [1]. The average levels and the cut-offs of disease activity and impact corresponding to the different levels of PESS seem discriminative enough to be used in the clinical setting: progressive improvement along the PESS levels seems to offer a reliable indication of improvement of both disease activity and impact. A patient who does not reach a 'very good' PESS level despite disease remission will probably require a more detailed evaluation of the reasons behind the persisting unfavourable symptom state so as to guide adjunctive interventions. The use of RAID would seem a promising next step in such circumstances. Conversely, a patient who scores PESS as at least acceptable, despite active arthritis, will probably benefit from explanations regarding the need for intensified immunosuppressive medication, in order to enhance compliance to treatment. Additionally, PESS, considering more levels of categorization, seems to convey a concept that is more easily manageable by the patient than PASS and other PRO measures, such PGA [45], which can facilitate the communication between patient and physician, and consequently provide a better support for the shared decision-making process [46, 47]. PESS may be a suitable tool to systematically screen for disease impact, as suggested by the recently proposed 'Dual Target strategy' for the management of RA [48].

We consider that this validation study supports that PESS may be a valuable tool for clinical practice and research in patients with RA. However, further external validation is required. Additionally, we think PESS also deserves to be tested and validated in other inflammatory rheumatic diseases, such as SpA and PsA.

Acknowledgements

The RAID study was supported by EULAR (grant CLI.042). NOR-DMARD has received financial support from various pharmaceutical companies [partly funded through unrestricted grants from AbbVie, BMS, MSD, Pfizer (Wyeth), Roche and UCB]. We thank the physicians and nurses who contributed to the data collection in the NOR-DMARD study. We thank the RAID investigators: A. Balanescu (Romania), D.T. Boumpas (Greece), L. Carmona (Spain), B.A. Dijkmans (The Netherlands), F. Gogus (Turkey), J.R. Kirwan (UK), M. Matucci-Cerinic (Italy), E.M. Mola (Spain), K. Otsa (Estonia), G. Schett (Germany) and T. Sokka (Finland).

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.

Disclosure statement: The authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

1 Smolen JS, Landewé R, Bijlsma J *et al*. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. Ann Rheum Dis 2017; 76:960–77.

2 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.

3 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.

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 Radner H, Chatzidionysiou K, Nikiphorou E *et al*. 2017 EULAR recommendations for a core data set to support observational research and clinical care in rheumatoid arthritis. Ann Rheum Dis 2018;77:476–9.

6 Tubach F, Dougados M, Falissard B *et al*. Feeling good rather than feeling better matters more to patients. Arthritis Rheum 2006;55:526–30.

7 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.

8 Heiberg T, Kvien TK, Mowinckel P, Aletaha D *et al*. Identification of disease activity and health status cut-off points for the symptom state acceptable to patients with rheumatoid arthritis. Ann Rheum Dis 2008;67:967–71.

9 Tubach F, Ravaud P, Martin-Mola E *et al*. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. Arthritis Care Res (Hoboken) 2012;64:1699–707.

10 Salaffi F, Carotti M, Gutierrez M, Di Carlo M, De Angelis R. Patient acceptable symptom state in self-report questionnaires and composite clinical disease index for assessing rheumatoid arthritis activity: identification of cut-off points for routine care. Biomed Res Int 2015; 2015:1–8.

11 Kvamme MK, Kristiansen IS, Lie E, Kvien TK. Identification of cutpoints for acceptable health status and important improvement in patient-reported outcomes, in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. J Rheumatol 2010;37:26–31.

12 Puyraimond-Zemmour D, Etcheto A, Fautrel B *et al*. Associations between five important domains of health and the patient acceptable symptom state in rheumatoid arthritis and psoriatic arthritis: a cross-sectional study of 977 patients. Arthritis Care Res (Hoboken) 2017;69: 1504–9.

13 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.

14 Michelsen B, Kristianslund EK, Sexton J *et al*. Do depression and anxiety reduce the likelihood of remission in rheumatoid arthritis and psoriatic arthritis? Data from the prospective multicentre NOR-DMARD study. Ann Rheum Dis 2017;76:1906–10.

15 Prevoo ML, van't Hof MA, Kuper HH *et al*. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38:44–8.

16 DAS-Score Website 2018. https://www.das-score.nl/das28/en/difference-between-the-das-and-das28/how-to-measure-the-das28/how-to-calculate-the-das28/alternative-validated-formulae.html (28 December 2018, date last accessed).

17 Smolen JS, Breedveld FC, Schiff MH *et al*. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. Rheumatology (Oxford) 2003;42: 244–57.

18 Aletaha D, Ward MM, Machold KP, Nell VP *et al*. Remission and active disease in rheumatoid arthritis: defining criteria for disease activity states. Arthritis Rheum 2005;52:2625–36.

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, Smolen JS. The definition and measurement of disease modification in inflammatory rheumatic diseases. Rheum Dis Clin North Am 2006;32:9–44, vii.

21 Felson DT, Smolen JS, Wells G *et al*. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum 2011;63:573–86.

22 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.

23 Pincus T, Swearingen C, Wolfe F. Toward a multidimensional Health Assessment Questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. Arthritis Rheum 1999;42:2220–30.

24 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.

25 Maska L, Anderson J, Michaud K. Measures of functional status and quality of life in rheumatoid arthritis: health Assessment Questionnaire Disability Index (HAQ), Modified Health Assessment Questionnaire (MHAQ), Multidimensional Health Assessment Questionnaire (MDHAQ), Health Assessment Questionnaire II (HAQ-II), Improved Health Assessment Questionnaire (Improved HAQ), and Rheumatoid Arthritis Quality of Life (RAQoL). Arthritis Care Res (Hoboken) 2011;63(Suppl 11):S4–13.

26 Salaffi F, Di Carlo M, Vojinovic J *et al*. Validity of the rheumatoid arthritis impact of disease (RAID) score and definition of cut-off points for disease activity states in a population-based European cohort of patients with rheumatoid arthritis. Joint Bone Spine 2017; 85(3): 317–22.

27 Hewlett S, Kirwan J, Bode C *et al*. The revised Bristol Rheumatoid Arthritis Fatigue measures and the Rheumatoid Arthritis Impact of Disease scale: validation in six countries. Rheumatology (Oxford) 2018;57:300–8.

28 Turner D, Schünemann HJ, Griffith LE *et al*. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. J Clin Epidemiol 2009;62:374–9.

29 Hajian-Tilaki K. The choice of methods in determining the optimal cut-off value for quantitative diagnostic test evaluation. Stat Methods Med Res 2018;27: 2374–83.

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

31 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.

32 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 (Oxford) 2017; 56:1573–8.

33 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.

34 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.

35 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.

36 Kalyoncu U, Dougados M, Daures JP, Gossec L. Reporting of patient-reported outcomes in recent trials in rheumatoid arthritis: a systematic literature review. Ann Rheum Dis 2009;68:183–90.

37 Gwinnutt JM, Hyrich KL, RAMS Co-Investigators, Lunt M, Barton A, Verstappen SM. Long-term outcomes of patients who rate symptoms of rheumatoid arthritis as 'satisfactory'. Rheumatology (Oxford) 2019; kez497, 10.1093/rheumatology/kez497.

38 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.

39 Ferreira RJO, Carvalho PD, Ndosi M *et al*. Impact of patient global assessment on achieving remission in patients with rheumatoid arthritis: a multinational study using the METEOR database. Arthritis Care Res (Hoboken) 2019;71:1317–25.

40 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.

41 Putrik P, Ramiro S, Hifinger M *et al*. In wealthier countries, patients perceive worse impact of the disease although they have lower objectively assessed disease activity: results from the cross-sectional COMORA study. Ann Rheum Dis 2016;75:715–20.

42 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.

43 Orbai AM, de Wit M, Mease P *et al*. International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. Ann Rheum Dis 2017;76: 673–80.

44 Gossec L, Dougados M, Dixon W. Patient-reported outcomes as end points in clinical trials in rheumatoid arthritis. RMD Open 2015;1:e000019.

45 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) 2019;kez467, 10.1093/rheumatology/kez467

46 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.

47 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.

48 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.