Measuring Disease Exacerbation and Flares in Rheumatoid Arthritis: Comparison of Commonly Used Disease Activity Indices and Individual Measures

Martijn A.H. Oude Voshaar, Marjan Ghiti Moghadam, Harald E. Vonkeman, Peter M. ten Klooster, Dirkjan van Schaardenburg, Janneke Tekstra, Henk Visser, Mart A.F.J. van de Laar, and Tim L. Jansen

ABSTRACT. ***Objective.*** To evaluate and compare the utility of commonly used outcome measures for assessing disease exacerbation or flare in patients with rheumatoid arthritis (RA).

Methods. Data from the Dutch Potential Optimalisation of (Expediency) and Effectiveness of Tumor necrosis factor-α blockers (POET) study, in which 462 patients discontinued their tumor necrosis factor-α inhibitor, were used. The ability of different measures to discriminate between those with and without physician-reported flare or medication escalation at the 3-month visit (T2) was evaluated by calculating effect size (ES) statistics. Responsiveness to increased disease activity was compared between measures by standardizing change scores (SCS) from baseline to the 3-month visit. Finally, the incremental validity of individual outcome measures beyond the Simplified Disease Activity Score was evaluated using logistic regression analysis.

Results. The SCS were greater for disease activity indices than for any of the individual measures. The 28-joint Disease Activity Score, Clinical Disease Activity Index, and Simplified Disease Activity Index performed similarly. Pain and physician's (PGA) and patient's global assessment (PtGA) of disease activity were the most responsive individual measures. Similar results were obtained for discriminative ability, with greatest ES for disease activity indices followed by pain, PGA, and PtGA. Pain was the only measure to demonstrate incremental validity beyond SDAI in predicting 3-month flare status.

Conclusion. These results support the use of composite disease activity indices, patient-reported pain and disease activity, and physician-reported disease activity for measuring disease exacerbation or identifying flares of RA. Physical function, acute-phase response, and the auxiliary measures fatigue, participation, and emotional well-being performed poorly. (J Rheumatol First Release May 15 2017; doi:10.3899/jrheum.160915)

Key Indexing Terms:
RHEUMATOID ARTHRITIS
DISEASE ACTIVITY
TUMOR NECROSIS FACTOR INHIBITORS
REMISSION

Disease activity cannot be directly measured in rheumatoid arthritis (RA) because of the absence of a gold standard measure of the inflammatory process of the disease. Consequently, numerous standardized measures of various symptoms and signs of the disease and measures of global disease effect have been proposed and validated over time to assess beneficial effects of treatment. Early clinical trials in this field were characterized by many different outcome

From the Arthritis Center Twente, Department of Psychology, Health and Technology, University of Twente; Arthritis Center Twente, Department of Rheumatology and Clinical Immunology, Medical Spectrum Twente, Enschede; Department of Rheumatology, VU University Medical Center and Reade Medical Center, Amsterdam; Department of Rheumatology, University Medical Center Utrecht, Utrecht; Department of Rheumatology, Rijnstate Medical Center, Arnhem; Department of Rheumatology, Viecuri Medical Center, Venlo, the Netherlands; Department of Rheumatology, University Hospital Leuven, Leuven, Belgium.

M.A. Oude Voshaar, PhD, Arthritis Center Twente, Department of Psychology, Health and Technology, University of Twente; M. Ghiti Moghadam, MD, Arthritis Center Twente, Department of Psychology, Health and Technology, University of Twente, and Department of Rheumatology, University Hospital Leuven, and Arthritis Center Twente, Department of Rheumatology and Clinical Immunology, Medical Spectrum Twente; H.E. Vonkeman, MD, PhD, Arthritis Center Twente, Department of Psychology, Health and Technology, University of Twente, and Arthritis Center Twente, Department of Rheumatology and Clinical Immunology, Medical Spectrum Twente; P.M. ten Klooster, PhD, Arthritis Center Twente, Department of Psychology, Health and Technology, University of Twente; D. van Schaardenburg, MD, PhD, Department of Rheumatology, VU University Medical Center and Reade Medical Center; J. Tekstra, MD, PhD, Department of Rheumatology, University Medical Center Utrecht; H. Visser, MD, PhD, Department of Rheumatology, Rijnstate Medical Center; M.A. van de Laar, MD, PhD, Arthritis Center Twente, Department of Psychology, Health and Technology, University of Twente, and Arthritis Center Twente, Department of Rheumatology and Clinical Immunology, Medical Spectrum Twente; T.L. Jansen, MD, PhD, Department of Rheumatology, Viecuri Medical Center.

Address correspondence to Dr. M. Ghiti Moghadam, University Hospital Leuven, Department of Rheumatology, Herestraat 49, 3000 Leuven, Belgium. E-mail: marjan.ghitimoghadam@uzleuven.be

Accepted for publication April 7, 2017.

measures and results were frequently difficult to compare[1]. To address this, a set of 7 outcome domains was recommended independently by the American College of Rheumatology (ACR) and the World Health Organization/International League against Rheumatism[2,3]. Other efforts to standardize outcomes across studies have led to the development and validation of several composite disease activity indices. Each of these combines a number of the endorsed outcomes to produce an overall disease activity score[4,5,6]. Responsiveness and discriminative ability are key properties of such measures and the performance of commonly used disease activity indices, as well as individual core set measures in detecting treatment effects and discriminating between different levels of achieved therapeutic response, have been extensively described in previous studies[7,8,9,10,11,12,13,14].

While these previous studies have provided a great deal of insight into which of the commonly applied instruments are most suitable for measuring treatment benefits in RA, not much is known about their performance when assessing exacerbation of disease activity or flares. This is of importance, however, because the occurrence of flares or disease worsening are increasingly adopted as endpoints of clinical trials[15]. Moreover, there is a need for instruments that can be used to monitor disease exacerbation in clinical practice[16].

It has been proposed by the Outcome Measures in Rheumatology flare working group that RA flare represents worsening of disease activity that would, if persistent, in most cases lead to initiation or change of therapy, and a flare represents a cluster of symptoms of sufficient duration and intensity as to require initiation, change, or increase in therapy[17]. Moreover, an extended set of outcomes for the assessment of flare was endorsed, which includes several besides the 7 core outcomes, and is intended to comprehensively cover the experience from the patient's perspective[18,19]. However, a factor to consider in addition to the relevance of outcome domains, according to experts and patients, is that the assessment of flares or disease worsening should proceed using measurement instruments that are optimally valid and reliable for this purpose to minimize flare status misclassifications (i.e., false-positives and false-negatives) and nonoptimal sensitivity to change because of random and systematic measurement errors.

In our present study, we compared the relative efficiency of a series of clinical and patient-reported measures using data from the Dutch Potential Optimalisation of (Expediency) and Effectiveness of Tumor necrosis factor-α blockers (POET) study to identify measurement instruments and outcome domains that most reliably assess disease exacerbation. The primary objective was to compare the performance of commonly used disease activity measures and indices, as well as recently endorsed patient-reported outcome (PRO) domains in assessing disease activity worsening. A secondary objective was to evaluate the incremental validity of individual outcomes over the traditional disease activity indices.

MATERIALS AND METHODS

POET study. POET was an open-label pragmatic randomized controlled trial conducted at 47 rheumatology departments in the Netherlands. To be eligible for the study, patients had to be 18 years of age or older, meet the 1987-revised ACR criteria for the classification of RA[20], have been treated for at least 1 year with concomitant tumor necrosis factor-α inhibitor (TNFi) and conventional synthetic disease-modifying antirheumatic drugs (DMARD), and be in remission or stable low disease activity for at least 6 months. Remission or stable low disease activity were defined as either a 28-joint Disease Activity Score (DAS28)[21] < 3.2 measured at least twice or the rheumatologist's clinical impression of remission or stable low disease activity in combination with at least 1 C-reactive protein (CRP) level < 10 mg/l in the 6 months prior to inclusion. There were no exclusion criteria. Study inclusion took place from March 2012 to March 2014. The study was approved by the Institutional Ethical Review Boards of all participating hospitals (grant/award number 40-00506-98-12001). Written informed consent was obtained from each patient upon study entry. The study was conducted in adherence to the Good Clinical Practice guidelines and with regulatory requirements consistent with the Declaration of Helsinki. The study is registered in the Netherlands Trial Register (number NTR3112).

Measurements. Baseline characteristics included age, sex, disease duration, medication use, and rheumatoid factor and anticyclic citrullinated peptide antibody status. Patients were evaluated at baseline and at least once every 3 months thereafter up to 1 year by the attending rheumatologist and rheumatology nurse in accordance with current Dutch treatment guidelines for RA. Clinical measurements, which are part of standard rheumatology care, were performed at every visit and included the tender joint count (TJC) in 28 joints, the swollen joint count (SJC) in 28 joints, the erythrocyte sedimentation rate (ESR), CRP, and the physician's (PGA) and patient's global assessment (PtGA) on a 0–100 visual analog scale (VAS). Additionally, physician-reported flares and changes in medication were recorded at each scheduled or unscheduled visit.

The following PRO were also administered at each study visit. Fatigue was assessed using the Bristol RA Fatigue Multi-Dimensional Questionnaire and 0–10 numerical rating scales assessing fatigue severity and effect of fatigue on daily functioning. Patient-reported well-being, disease activity, and pain were assessed using 0–100 VAS. The Health Assessment Questionnaire-Disability Index (HAQ-DI) was administered to evaluate disability. Finally, the Medical Outcomes Study Short Form-36 (SF-36) was administered to assess health-related quality of life. We used the role physical, role emotional, and social functioning scale to assess participation. The bodily pain, physical functioning, and vitality scales were used to assess pain, disability, and fatigue, respectively. The Clinical Disease Activity Index (CDAI), Simplified Disease Activity Index (SDAI), and DAS28-ESR scores were calculated using their respective standard scoring algorithms[4].

Statistical analysis. For our posthoc analysis, data were used from 462 patients randomized to stopping TNFi treatment. Standardized effect size (ES) statistics with pooled SD were obtained to compare the ability of various measures to discriminate between patients with and without flare at the 3-month assessment (T2). The first flare anchor we examined was physician-reported flare at T2. Medication escalation was also evaluated as an anchor of flare, defined as starting or increasing any biological or nonbiological DMARD (including glucocorticoids) at T2.

Responsiveness was evaluated by comparing standardized change scores (SCS), again with pooled SD from baseline to T2. The magnitude of ES and SCS was interpreted according to the guidelines provided by Cohen (i.e., trivial < 0.20, small ≥ 0.20, moderate ≥ 0.50, and large ≥ 0.80)[22]. The utility of individual measures to predict flare status at T2 was investigated using univariable logistic regression analysis. For each domain, the outcome measure with the largest ES in the analysis of discriminant validity was

selected to represent that domain. Predictive strength was quantified using R^2 as an effect size estimator. Guidelines for interpreting R^2 were again derived from Cohen: small ≥ 0.01, moderate ≥ 0.09, and large ≥ 0.20[22]. The incremental validity of outcome domains not included in the SDAI was evaluated using multivariable hierarchical logistic regression analysis with flare status at T2 as the dependent variable. CDAI was entered as a first block and individual outcome measures were entered as a second block. Incremental validity was evaluated using ΔR^2.

Multiple imputation by chained equations was performed to replace missing values. The imputation model included the dependent variable: physician-reported flare status (yes/no), the evaluated outcome measures, as well as age, followup time, and total number of flares during the total study period. Twenty datasets with imputed plausible values were obtained, with 200 iterations between datasets. Rubin's rules were used to obtain pooled variable estimates and their associated standard errors. Pooled analyses will be reported throughout our paper.

RESULTS

Baseline characteristics are presented in Table 1. The sample was characterized by established disease, with long disease duration, and 62.6% of patients with erosive disease. In accordance with the inclusion criteria, disease activity according to DAS28 was low at baseline, as was physical disability according to HAQ-DI.

Discriminant validity. While 39.4% of patients had a medication escalation at T2, only 18.7% of patients were reported to have a clinical flare at T2 by their rheumatologist. A consistently greater contrast in scores (i.e., ES of greater magnitude) was observed for clinical flare compared with medication escalation across all measures. The clinical composite scores performed best and interchangeably, followed by pain and PGA. Bodily pain assessed using the SF-36 was consistently the best performing individual measure with ES of moderate magnitude, followed by PGA and PtGA of disease activity and well-being. Measures of physical aspects of patient-reported health and fatigue had ES of small to moderate magnitude. Measures of emotional well-being consistently had ES of trivial magnitude (Figure 1).

Responsiveness. A moderate increase in disease activity according to all 3 disease activity indices (SCS ≈ 0.50) was observed for the stop group. Individual DAS28 components showed SCS of trivial (ESR) to small magnitude (TJC, SJC, well-being). PGA of disease activity was the only individual measure with SCS of moderate magnitude. Patient assessment of disease activity and both measures of pain had SCS of small magnitude, but outperformed each of the individual DAS28 components, while the SF-36 vitality and social function scales had SCS of small magnitude and smaller than the DAS28 components. Each of the remaining measures had an SCS of trivial magnitude (Figure 2).

Incremental validity. In univariable analysis, all outcomes except emotional well-being were significantly associated with clinical flare status and all outcomes except emotional well-being and fatigue were significantly associated with medication escalation at T2. Of the measures included in the SDAI, CRP was consistently most weakly associated with flare status in univariable analysis, while physician's global performed best. Of the measures not included in the SDAI, pain was consistently most strongly associated with 3 months. Pain and PGA were performed best overall. According to the analysis in the binary logistic regression models with only SDAI, R^2 was of small magnitude for medication escalation (R^2 = 0.09) and large magnitude for clinical flare (R^2 = 0.26). In Table 2, only pain contributed significantly to the classification of flare status beyond the SDAI. However, the magnitude of its independent effect was small.

Table 1. Patient characteristics. Values are mean (SD) unless otherwise specified.

Characteristics	Stop Group, n = 462
Age, yrs	59.5 (10.5)
Disease duration, yrs	12.0 (8.9)
Female, n (%)	181 (66)
DAS28	1.90 (0.8)
HAQ-DI	0.6 (0.6)
Erosive disease, n (%)	294 (62.6)
RF-positive, n (%)	316 (67.2)

DAS28: 28-joint Disease Activity Score; HAQ-DI: Health Assessment Questionnaire–Disability Index; RF: rheumatoid factor.

DISCUSSION

In our present study, the performance for assessing disease exacerbation was compared between several validated disease activity indices and individual measures that either assessed 1 of the core set variables or 1 of the domains that have been proposed as outcomes for the assessment of flare in RA.

Overall, the composite disease activity indices outperformed individual measures across the various comparisons. This was unsurprising because standardized change scores of individual measures are more affected by measurement error compared with composite scores, because random error tends to cancel itself out when information from multiple measures is combined. Nevertheless, the disease activity indices proved to be noticeably more efficient indicators of disease exacerbation compared with most individual measures, which further supports the practice of using composite tools for assessing disease activity in RA.

Interestingly, despite their different constituent measures and scoring protocols, the CDAI, SDAI, and DAS28 performed equivalently as measures of disease exacerbation in our present study. Similar sensitivity to improvement was observed as well for these 3 tools in 2 previous studies[23,24] and no clear differences in other measurement properties were observed in a comparative systematic review[25]. When measuring disease activity exacerbations, it seems that the main advantages of DAS28 are that it has been the most frequently used measure in previous studies, providing a richer frame of reference for the interpretation of study

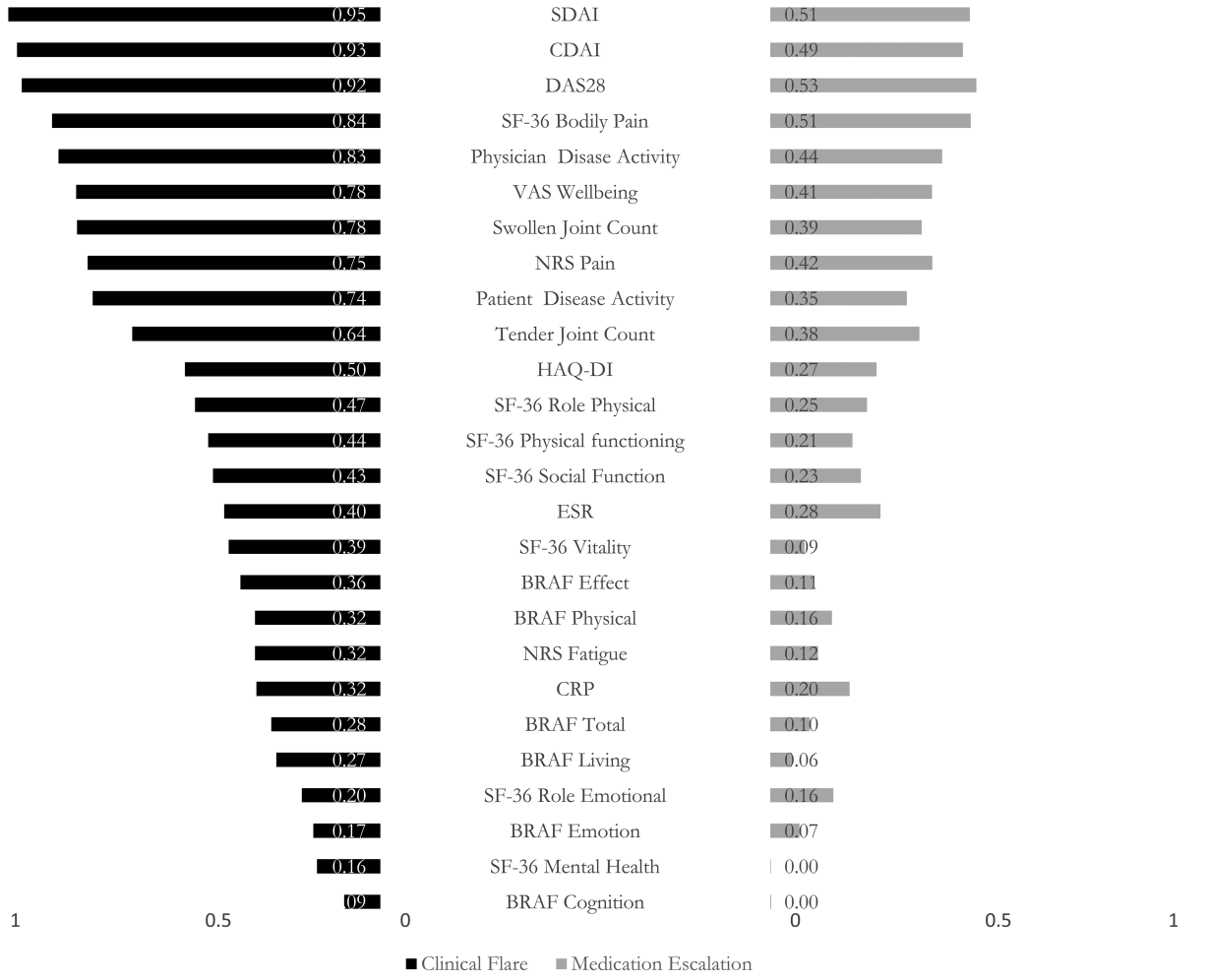

Figure 1. Discriminative ability. Bar charts represent effect size statistics. SDAI: Simplified Disease Activity Index; CDAI: Clinical Disease Activity Index; DAS28: 28-joint Disease Activity Score; SF-36: Medical Outcomes Study Short Form-36; VAS: visual analog scale; NRS: numerical rating scale; HAQ-DI: Health Assessment Questionnaire-Disability Index; ESR: erythrocyte sedimentation rate; BRAF: Bristol Rheumatoid Arthritis Fatigue Scale; CRP: C-reactive protein.

outcomes against historical results, and that it is currently the only clinical composite score for which validated flare criteria are available[26]. However, scoring for the DAS28 is more complex than for SDAI and CDAI because of the differential weighting of individual components. Further, acute-phase reactants were shown in our study to provide little additional information on disease activity exacerbation and were among the weakest univariable predictors of flare status. Because there is usually a delay in the availability of laboratory results, the inclusion of acute-phase reactants in both the SDAI and DAS28 may create a logistical barrier to obtaining disease activity scores in real time for apparently redundant information from a statistical point of view.

PGA and PtGA of disease activity and pain assessed using either VAS or the SF-36 bodily pain scale were the most sensitive individual measures of increased disease activity in the POET study. It is worth mentioning that changes in the amount of pain a patient reports were more predictive than other variables measured in our study of treatment decisions (medication escalation) and physician-reported flare. Pain intensity was also the only outcome domain that provided unique information beyond the information provided by the SDAI in the analysis of incremental validity, even though pain is indirectly represented in both the TJC and very likely PtGA, which were both controlled for in our analysis. Previous studies have repeatedly established that pain is the most important treatment priority for patients throughout the disease course[27]. These results suggest that more comprehensive evaluation of overall disease status of a patient could be obtained using a clinical composite score that includes pain.

The overall performance ranking of specific measures in our study corresponds largely to that observed in previous studies that have compared the ability of different measures to assess disease improvement. Only the poor performance of physical function contrasts with previous studies where it

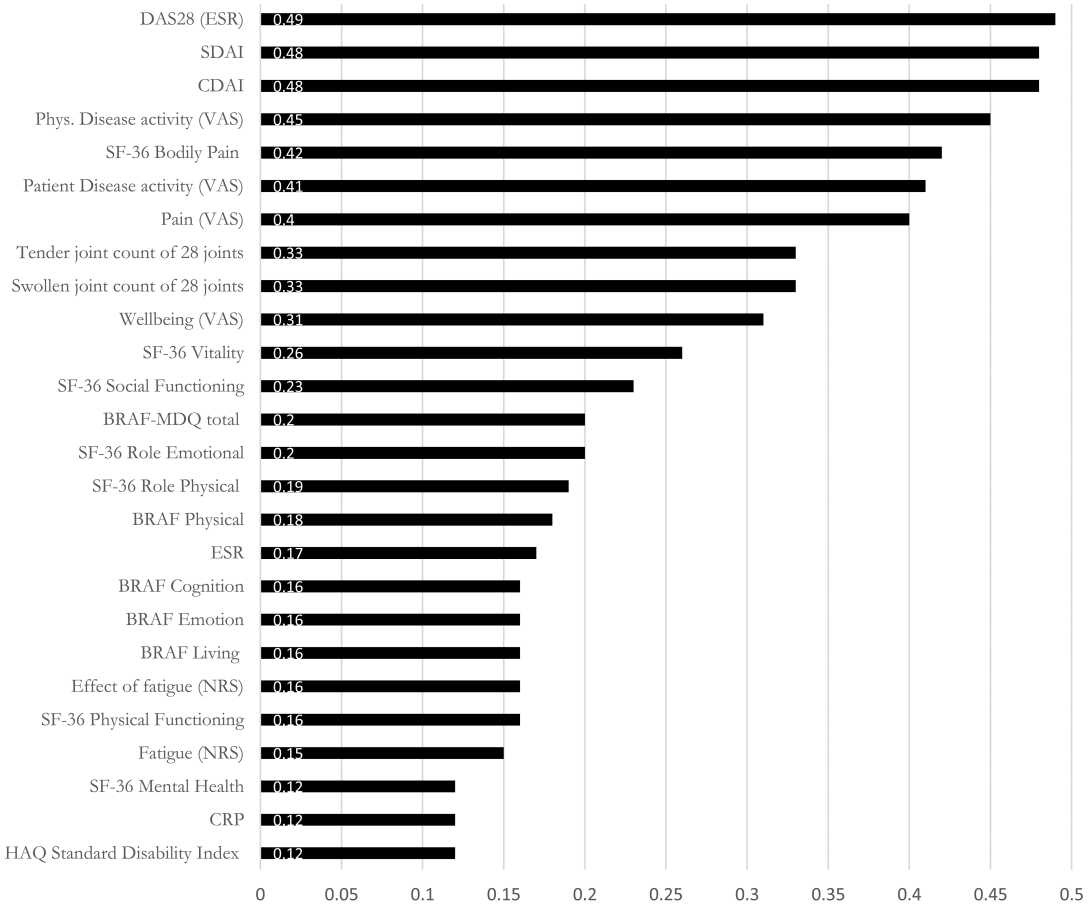

Figure 2. Standardized change scores in POET stop group. Bar charts represent effect size statistics. DAS28: 28-joint Disease Activity Score; ESR: erythrocyte sedimentation rate; SDAI: Simplified Disease Activity Index; CDAI: Clinical Disease Activity Index; VAS: visual analog scale; SF-36: Medical Outcomes Study Short Form-36; BRAF: Bristol Rheumatoid Arthritis Fatigue Scale; MDQ: multi-dimensional questionnaire; NRS: numerical rating scale; CRP: C-reactive protein; HAQ: Health Assessment Questionnaire.

has consistently been found to be one of the best performing indicators of treatment benefits[7,8,9,10,11,12] and increased disease activity in an observational study[28]. These contrasting results are likely explained by the long average disease duration and the high prevalence of patients with erosive disease at baseline in POET, which previous studies have found to be associated with decreased responsiveness when measuring physical function in patients with RA[29]. It is commonly believed that while inflammatory disease activity is the most important determinant of physical function in RA, structural damage increasingly contributes to disability later in the disease course, effectively lowering the ceiling on the amount of improvement that can be realized[30]. Our results provide further support for this notion and suggest that inferences regarding disease exacerbations from (lack of) change in physical function scores should be made with caution in populations with longstanding disease.

Little evidence was found in our present study to support an extended set of outcomes for measuring flares. PRO other than pain and PtGA were found to be weakly sensitive to disease flare and did not provide incremental information above and beyond the SDAI when predicting flare status. These results suggest that the addition of fatigue, emotional well-being, or participation to a composite score may not contribute much to its reliability or predictive validity for measuring flares or exacerbated disease. Previous studies also found limited responsiveness for measures of participation and emotional well-being, while the performance of fatigue has been mixed[6,7,8,9,10,11]. An explanation for this may be that participation restrictions and emotional well-being are complex, integrated domains that may be determined by disease activity as well as comorbidities, and personal and environmental factors.

Strengths of our study are that the POET study was designed to evaluate exacerbation of RA resulting from TNFi discontinuation in patients with longstanding stable low disease activity and that physician-rated flares were assessed per protocol, while previous studies are posthoc analyses of subsets of deteriorated patients from studies originally designed to evaluate patients expected to improve (e.g., clinical

Table 2. Univariable and multivariable predictors of flare status.

Variables	Clinical Flare, 18.7%				Medication Escalation, 39.4%			
	Univariable		SDAI Corrected		Univariable		SDAI Corrected	
	p	R^2	p	ΔR^2	p	R^2	p	ΔR^2
SDAI	< 0.01	0.26	n/a	n/a	< 0.01	0.08	n/a	n/a
Core set measures								
Swollen joint count	< 0.01	0.14	n/a	n/a	< 0.01	0.05	n/a	n/a
PtGA	< 0.01	0.15	n/a	n/a	< 0.01	0.04	n/a	n/a
PGA	< 0.01	0.19	n/a	n/a	< 0.01	0.06	n/a	n/a
CRP	< 0.01	0.05	n/a	n/a	0.01	0.03	n/a	n/a
Pain	< 0.01	0.18	< 0.01	0.03	< 0.01	0.09	0.01	0.03
Physical function	< 0.01	0.07	0.21	0.00	0.01	0.03	0.38	0.00
Auxiliary measures								
Participation	< 0.01	0.05	0.14	0.00	0.04	0.01	0.39	0.00
Fatigue	< 0.01	0.03	0.21	0.00	0.27	0.00	0.68	0.00
Emotional well-being	0.17	0.00	0.73	0.00	0.99	0.00	0.49	0.00

SDAI: Simplified Disease Activity Index; CRP: C-reactive protein; n/a: not applicable; PtGA: patient's global assessment of disease activity; PGA: physician's global assessment of disease activity.

trials) or remain stable (observational studies)[28,31,32,33,34]. A limitation of our study is that there was relatively little variability of scores for some measures (e.g., SJC) because of the homogeneous, low level of disease activity at the start of our study. Different results might have been achieved in patients with more severe disease at baseline. However, practical measurement settings in which disease worsening is to be assessed most likely involve patients starting in a state of low disease activity (e.g., tapering of medication in daily clinical practice or clinical trials).

The results of our study support the use of traditional disease activity measures and composite indices for assessing flare or disease exacerbation. Pain, PGA, and PtGA of disease activity were the best performing individual measures, while the composite DAS28, CDAI, and SDAI performed equivalently as measures of exacerbation. Patient-reported measures assessing the domains of participation, fatigue, and emotional functioning performed worst. Based on these findings, we recommend that assessment of disease flares should proceed using core set measures, preferably combined, particularly pain intensity.

REFERENCES

1. Pincus T. The American College of Rheumatology (ACR) Core Data Set and derivative "patient only" indices to assess rheumatoid arthritis. Clin Exp Rheumatol 2005;23 Suppl 39:S109-13.
2. Boers M, Tugwell P, Felson DT, van Riel PL, Kirwan JR, Edmonds JP, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol Suppl. 1994 Sep;41:86–9.
3. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum 1993;36:729–40.
4. Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol 2005;23 Suppl 39:S100-8.
5. van Gestel AM, Prevoo ML, van 't Hof MA, van Rijswijk MH, van de Putte LB, van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. Arthritis Rheum 1996;39:34–40.
6. Aletaha D, Landewe R, Karonitsch T, Bathon J, Boers M, Bombardier C, et al. Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. Ann Rheum Dis 2008;67:1360–4.
7. Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Clinical and quality-of-life measures' responsiveness to treatment in a randomized controlled trial. Arthritis Rheum 1995;38:1568–80.
8. Anderson JJ, Chernoff MC. Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. J Rheumatol 1993; 20:535–7.
9. Verhoeven AC, Boers M, van Der Linden S. Responsiveness of the core set, response criteria, and utilities in early rheumatoid arthritis. Ann Rheum Dis 2000;59:966–74.
10. Russell AS, Conner-Spady B, Mintz A, Maksymowych WP. The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. J Rheumatol 2003;30:941–7.
11. Tugwell P, Wells G, Strand V, Maetzel A, Bombardier C, Crawford B, et al. Clinical improvement as reflected in measures of function and health-related quality of life following treatment with leflunomide compared with methotrexate in patients with rheumatoid arthritis: sensitivity and relative efficiency to detect a treatment effect in a twelve-month, placebo-controlled trial. Leflunomide Rheumatoid Arthritis Investigators Group. Arthritis Rheum 2000;43:506–14.
12. Cohen SB, Strand V, Aguilar D, Ofman JJ. Patient- versus physician-reported outcomes in rheumatoid arthritis patients treated with recombinant interleukin-1 receptor antagonist (anakinra) therapy. Rheumatology 2004;43:704–11.
13. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. J Clin Epidemiol 1989;42:403–8.
14. Neogi T, Xie H, Felson DT. Relative responsiveness of

physician/assessor-derived and patient-derived core set measures in rheumatoid arthritis trials. J Rheumatol 2008;35:757–62.
15. den Broeder AA, van der Maas A, van den Bemt BJ. Dose de-escalation strategies and role of therapeutic drug monitoring of biologics in RA. Rheumatology 2010;49:1801–3.
16. Alten R, Pohl C, Choy EH, Christensen R, Furst DE, Hewlett SE, et al; OMERACT RA Flare Definition Working Group. Developing a construct to evaluate flares in rheumatoid arthritis: a conceptual report of the OMERACT RA Flare Definition Working Group. J Rheumatol 2011;38:1745–50.
17. Bingham CO 3rd, Pohl C, Woodworth TG, Hewlett SE, May JE, Rahman MU, et al. Developing a standardized definition for disease "flare" in rheumatoid arthritis (OMERACT 9 Special Interest Group). J Rheumatol 2009;36:2335–41.
18. Bartlett SJ, Bykerk VP, Cooksey R, Choy EH, Alten R, Christensen R, et al. Feasibility and domain validation of rheumatoid arthritis (RA) flare core domain set: report of the OMERACT 2014 RA Flare Group Plenary. J Rheumatol 2015;42:2185-9.
19. Berthelot JM, De Bandt M, Morel J, Benatig F, Constantin A, Gaudin P, et al; STPR group of French Society of Rheumatology. A tool to identify recent or present rheumatoid arthritis flare from both patient and physician perspectives: the 'FLARE' instrument. Ann Rheum Dis 2012;71:1110–6.
20. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.
21. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. 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.
22. Cohen J. A power primer. Psychol Bull 1992;112:155-9.
23. Aletaha D, Nell VP, Stamm T, Uffmann M, Pflugbeil S, Machold K, 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.
24. Ranganath VK, Yoon J, Khanna D, Park GS, Furst DE, Elashoff DA, et al; Western Consortium of Practicing Rheumatologists. Comparison of composite measures of disease activity in an early seropositive rheumatoid arthritis cohort. Ann Rheum Dis 2007;66:1633–40.
25. Gaujoux-Viala C, Mouterde G, Baillet A, Claudepierre P, Fautrel B, Le Loët X, et al. Evaluating disease activity in rheumatoid arthritis: which composite index is best? A systematic literature analysis of studies comparing the psychometric properties of the DAS, DAS28, SDAI and CDAI. Joint Bone Spine 2012;79:149–55.
26. van der Maas A, Lie E, Christensen R, Choy E, de Man YA, van Riel P, et al. Construct and criterion validity of several proposed DAS28-based rheumatoid arthritis flare criteria: an OMERACT cohort validation study. Ann Rheum Dis 2013;72:1800–5.
27. ten Klooster PM, Veehof MM, Taal E, van Riel PL, van de Laar MA. Changes in priorities for improvement in patients with rheumatoid arthritis during 1 year of anti-tumour necrosis factor treatment. Ann Rheum Dis 2007;66:1485–90.
28. Lie E, Woodworth TG, Christensen R, Kvien TK, Bykerk V, Furst DE, et al. Validation of OMERACT preliminary rheumatoid arthritis flare domains in the NOR-DMARD study. Ann Rheum Dis 2014;73:1781–7.
29. Ward MM. Interpreting measurements of physical function in clinical trials. Ann Rheum Dis 2007;66 Suppl 3:iii32-4.
30. Drossaers-Bakker KW, de Buck M, van Zeben D, Zwinderman AH, Breedveld FC, Hazes JM. Long-term course and outcome of functional capacity in rheumatoid arthritis: the effect of disease activity and radiologic damage over time. Arthritis Rheum 1999;42:1854–60.
31. Hays RD, Spritzer KL, Fries JF, Krishnan E. Responsiveness and minimally important difference for the patient-reported outcomes measurement information system (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. Ann Rheum Dis 2015;74:104–7.
32. Linde L, Sørensen J, Ostergaard M, Hørslev-Petersen K, Hetland ML. Health-related quality of life: validity, reliability, and responsiveness of SF-36, 15D, EQ-5D [corrected] RAQoL, and HAQ in patients with rheumatoid arthritis. J Rheumatol 2008;35:1528–37.
33. Fransen J, Häuselmann H, Michel BA, Caravatti M, Stucki G. Responsiveness of the self-assessed rheumatoid arthritis disease activity index to a flare of disease activity. Arthritis Rheum 2001;44:53–60.
34. Taal E, Rasker JJ, Riemsma RP. Sensitivity to change of AIMS2 and AIMS2-SF components in comparison to M-HAQ and VAS-pain. Ann Rheum Dis 2004;63:1655–8.