# Is Tightly Controlled Disease Activity Possible with Online Patient-reported Outcomes?

Margot J. Walter, S.H. Mohd Din, Johanna M. Hazes, E. Lesaffre, P.J. Barendregt, and Jolanda J. Luime

**ABSTRACT.** **Objective.** To evaluate the performance of patient-reported outcomes (PRO) as primary indices for identification and prediction of a 28-joint Disease Activity Score (DAS28) > 3.2 among patients with rheumatoid arthritis (RA).

**Methods.** Patients with RA completed monthly online PRO [Health Assessment Questionnaire (HAQ), Rheumatoid Arthritis Disease Activity Index (RADAI), visual analog scale (VAS) fatigue] and were clinically assessed every 3 months using the DAS28. Simple descriptive statistics, logistic regression, and the Bayesian joint modeling approach were used to analyze the data. The Bayesian joint model combines the scores and changes in the scores of 3 PRO to predict a DAS28 > 3.2 at the subsequent timepoint.

**Results.** A group of 159 patients with RA participated. Stratified summaries of the PRO by DAS28 categories at baseline provided incremental values of the PRO for more active disease. However, on an individual level, the DAS28 and the PRO fluctuated over time. The prediction of subsequent DAS score by a single instrument at single timepoints resulted in moderate sensitivity and specificity. Using the intercept and slope of the combined PRO of the first 3 measurements to predict the DAS28 state at 3 months resulted in a sensitivity of 0.81 and a specificity of 0.92. After 10-fold cross validation, the model had a sensitivity of 0.61 and specificity of 0.75 to identify patients with a DAS28 > 3.2.

**Conclusion.** PRO showed fluctuating levels of disease activity over time, while on a group level disease activity stayed the same. Using the changes in RADAI, HAQ, and VAS fatigue over time to predict future DAS28 > 3.2 resulted in moderate performance after the internal cross-validation of the model (sensitivity 0.61, specificity 0.75). (J Rheumatol First Release Feb 15 2014; doi:10.3899/ jrheum.130174)

*Key Indexing Terms:*
DISEASE ACTIVITY SCORE    PATIENT-REPORTED OUTCOME    RHEUMATOID ARTHRITIS

Tight control of disease activity is commonly monitored by clinical disease activity measures such as the Disease Activity Score (DAS)[1] or Simplified Disease Activity Index (SDAI)[2]. These measures, administered by physicians and nurses, combine tender joint count, swollen joint count (SJC), erythrocyte sedimentation rate (ESR), and a patient-reported measure of experienced global disease activity [visual analog scale (VAS) global]. The SDAI also includes the evaluator global assessment of disease activity. Because the DAS/SDAI are administered clinically and are therefore only measured during consultations, they may miss relevant fluctuations in disease activity over time, especially in patients who attend the clinic only once or twice a year.

An alternative way to assess disease activity could be the use of patient-reported outcomes (PRO). These measures can be provided to patients in a Web-based manner, enabling distant monitoring of disease activity by rheumatologists. There are many disease-related self-reported instruments, such as the Health Assessment Questionnaire (HAQ)[3], Rheumatoid Arthritis Disease Activity Index (RADAI)[4], the Medical Outcomes Study Short-Form 36[5], and the Arthritis Impact Measurement Scales 2[6], which may function as self-monitoring instruments at home. There is no single "gold standard" measure[7,8] that can serve for identification and prediction of a high disease activity state (DAS28 > 3.2) in patients with rheumatoid arthritis (RA) that is indicative for treatment escalation[9]. There is also no core set of PRO generally used in studies that fits the purpose of monitoring disease activity by patients[10,11]. Self-assessment of the DAS components demonstrated moderate to low correlations with a trained assessor, especially for the SJC[12,13]. In addition,

the unavailability of ESR limits the use of the unmodified 28-joint DAS (DAS28) score by patients themselves. While there are many studies of self-reported measures, no studies are available that investigated combined self-reported measures as an instrument that uses PRO data to predict subsequent DAS28 > 3.2.

In our study we aimed (1) to describe the change of functional status (HAQ), self-reported disease activity (RADAI), VAS fatigue, and the DAS28; (2) to identify patients with DAS28 > 3.2 by the proposed cutoff points for RADAI and the HAQ; and (3) to evaluate the performance of the HAQ, VAS fatigue, and RADAI as primary indices for identification and prediction of high disease activity (DAS28 > 3.2).

## MATERIALS AND METHODS

*Participants*. We recruited patients with clinical RA from an outpatient rheumatology clinic in Rotterdam. Participants were identified from the hospital record database and invited by their rheumatologists. Data were collected on demographics, DAS28, and PRO. Patients were eligible for inclusion if they were aged 18 years or older, were able to read and write in Dutch, and had access to a computer with Internet and e-mail capability. Because some patients wanted to participate but had no computer, we accepted that they could participate by completing the questionnaires in a paper version. We excluded patients with severe psychiatric illness or personality disorders. All patients signed an informed consent form before study enrollment. The study was approved by the independent medical ethics committee of the Erasmus MC.

*Procedures*. The duration of this study was 1 year. The patients were clinically evaluated using the DAS28 every 3 months in standard care by his/her rheumatologist or nurse practitioner. These patients were also asked to complete a Web-based patient-reported questionnaire each month over a 1-year period. The Web-based questionnaires were easy to complete, and no specific instruction was given. Patients were reminded about their next questionnaire by e-mail. Nonresponders received 2 reminders by e-mail and 1 phone call.

*Primary outcome measures*. The DAS28 was used as a primary outcome and a reference standard for moderate to high disease activity (DAS28 > 3.2)[9,14]. The DAS28 is a score ranging from 0 to 10, where a higher score indicates higher disease activity. Treatment escalation was indicated if DAS28 exceeded 3.2 points according to Dutch guidelines for treatment of RA. Data from the DAS28 were collected at baseline and at 3, 6, 9, and 12 months.

*PRO measurement*. We chose instruments that reflect different aspects of disease activity and that were available in Dutch. Over a 1-year period we measured the following PRO monthly: HAQ[3], RADAI[4], VAS global[15], and the VAS fatigue[16]. The HAQ measures functional status using 20 questions phrased in 8 subscales[3]. The score ranges from 0 to 3 (3 worst health) with a minimal clinical important difference of 0.20[17,18]. The HAQ score suggested for remission is ≤ 0.5, representing patients with almost no difficulties in daily activity. A HAQ score between 0.5 and 1.0 could be regarded as low activity, while a score above 1 would indicate moderate to high disease activity with major problems in performing daily activities[19]. The HAQ is commonly used in clinical trials. It is an effective and sensitive tool for measuring the functional status and it is correlated with the DAS score[20]. The HAQ is also a predictor of severe longterm outcomes, so it measures both disease course as well as outcome[21]. The RADAI measures self-reported disease activity[4]. It uses a scale ranging from 0 to 10, where higher scores indicate more disease activity. The cutpoints for the RADAI are < 2.2 for low disease activity, ≥ 2.2 and ≤ 4.9 for moderate disease activity, and > 4.9 for high disease activity[8]. The RADAI measures

domains similar to the DAS, but laboratory values are not required. It has proven to be reliable and sensitive to change compared to the DAS28, although low concordance for absolute values between RADAI and DAS28 were observed[22]. The VAS fatigue and VAS global are both single-item scales and measure 1 domain each. VAS fatigue asks about the severity of fatigue over the past week with anchors: no fatigue (0 mm) and extremely fatigued (100 mm). The scale is sensitive to change, valid, and reliable[16,23]. Although there are minimal important differences for the VAS fatigue for improvement and worsening, there is no guidance on the choice of cutoff points[24], while in general a 10 mm change in VAS seems to be a clinically detectable difference for patients[25]. The VAS global was also considered but because it is 1 component of the DAS28, it was not evaluated for predictive capacity of the DAS28 states. Each PRO was completed at 1-month intervals.

*Covariates*. In addition to the PRO, we also measured the effect of coping, self-efficacy, and illness perception on both the primary outcomes and the PRO. We selected these questionnaires because the underlying constructs may influence the observed relationship. Coping was measured by the Coping with Rheumatic Stressors scale, which is based on the frequency of individual coping efforts. It is divided into 5 scales, of which we used dealing with pain (decreasing activities), limitations, and optimism[26]. The self-efficacy was measured with the Dutch version of the arthritis self-efficacy scale, which constitutes 2 subscales related to self-efficacy to deal with pain and to deal with other symptoms (depression, fatigue, frustration)[27]. Illness perception was assessed by an 11-item list including aspects such as causes, experience of symptoms, consequences, timeline, and controllability of the disease[28].

*Statistical analysis*. No formal sample size calculation was done because of the exploratory nature of our study. We expected to find around 20% of patients with a change into high disease activity annually[29], leaving us with at least 30 cases of a flare of the disease activity among 150 patients.

*Analysis*. Simple descriptive statistics and diagrams were used for the pattern of change over time of the HAQ, RADAI, VAS fatigue, VAS global, medication, and the state of the DAS28. The cross-sectional relationship between the DAS28 and the PRO was evaluated by Spearman's rank correlation. Sensitivity and specificity of the proposed cutpoints of the HAQ and RADAI to identify a high disease activity state were compared to the clinical DAS28 > 3.2. Difference in PRO between patients with missing values on the DAS28 and patients without missing values were tested with the Mann-Whitney U test.

To evaluate the predictive capacity of the PRO for individual patients, we used a Bayesian joint modeling approach. This method is described in detail in Mohd Din, *et al* (submitted). With this model we tried to predict a moderate disease activity state (DAS28 > 3.2) at subsequent timepoints by the changes in the HAQ, RADAI, and VAS fatigue over time. We took the following steps: first, the skewed distributions of the PRO were transformed to values between 0 and 1, which, after a logit transformation, resulted in a normal distribution[30]. Second, the evolution of each PRO during a 3-month period was summarized into a random intercept and random slope by fitting linear mixed effects models. These 2 parts of each PRO reflect the part that is stable for the individual patient (random intercept) and the part that changes over time (random slope). Third, the intercepts and slopes of each PRO were used to estimate the DAS28 at the subsequent timepoint corrected for age, sex, self-efficacy, coping with pain, and 2 questions of the illness perception. The latter 2 steps were done at once in the Bayesian joint model, but for simplicity described here as 2 steps. This model resulted in predicted DAS28 values. These were classified into DAS28 ≤ 3.2 and DAS28 > 3.2. The predicted responses were then compared to the observed values of DAS28 and fitted by an ROC for discrimination and plotted for calibration[31]. The model described here was developed using the measurements recorded at baseline and months 1 and 2 to predict the DAS28 at Month 3. This selection of data was made because after 3 months the clinical evaluation could alter the course of disease by the adjustment of medication. This was likely to affect the scores

on the PRO as well as the subsequent DAS. The model developed with the first 3 months' data was internally validated by 10-fold cross validation.

Simple descriptive statistics were performed in STATA (version 11). The transformation of the PRO was performed in SAS. The Bayesian joint model was developed in WinBUGS and R[32].

## RESULTS

*Patients*. One hundred fifty-nine out of 174 invited patients with RA consented to the study; 76% were female; the average age was 54 years (SD 13.26); with a median disease duration of 4.5 years (min-max: 1–38 yrs). Thirty-seven percent of the patients had radiographic damage. Over time, medication did not change for 33% of the patients, 16% had a decrease in their medication by either type or dose, 39% had an increase, and 10% had a temporal increase by means of a glucocorticoid injection (triamcinolone acetonide intramuscular or intraarticular). Further details can be found in Table 1.

The majority of patients (90%) participated by Internet questionnaires, while 10% completed paper versions. Complete DAS28 data were available for 97 out of 159 patients (61%), 19% missed 1 DAS28 evaluation, and 20% missed 2 timepoints or more. Mann-Whitney U test showed no significant difference in PRO and DAS28 between patients with ≥ 1 missing DAS28 value and those with no missing values. For the PRO, complete data for all 13 timepoints were available for 64 out of 159 patients (47%), 29% missed 1–3 self-reported timepoints, and 24% missed 4 or more timepoints.

*Table 1*. Patients' characteristics at baseline (n = 159).

| Characteristics | Values |
|---|---|
| Demographic | |
| Age, yrs, median (IQR) | 55.7 (45.6–62.8) |
| Female, % | 76.1 |
| Native (Dutch) ethnicity, % | 80.3 |
| Duration of disease, yrs, median (IQR) | 4.5 (2.0–7.0) |
| Paid work, n = 154, % | 41.5 |
| Higher education, n = 154, % | 23.1 |
| Clinical measures | |
| RF-positive, % | 60.8 |
| Anti-CCP-positive, % | 52.8 |
| Radiographic damage, % | 37.0 |
| DMARD, % | |
| No medication | 6.3 |
| DMARD single | 23.9 |
| DMARD combination | 49.7 |
| Biological | 20.1 |
| DAS28, median (IQR) | 2.66 (2.01–3.44) |
| Patient-reported measures, median (IQR) | |
| HAQ score (0–3) | 0.62 (0.13–1.00) |
| RADAI (0–100) | 2.00 (0.84–2.99) |
| VAS fatigue (0–100) | 50 (29–70) |
| VAS global (0–100) | 40 (21–58) |

DMARD: disease-modifying antirheumatic drugs; RF: rheumatoid factor; DAS28: Disease Activity Score in 28 joints; HAQ: Health Assessment Questionnaire; RADAI: Rheumatoid Arthritis Disease Activity Index; VAS: visual analog scale; anti-CCP: anticyclic citrullinated peptide antibodies.

*Clinical disease activity and change over time*. Median disease activity measured by the DAS28 was 2.66 [interquartile range (IQR) 2.01–3.44], with 0.88 (SD 1.82) swollen joints and 2.1 (SD 4.34) tender joints at baseline. DAS28 remission was found in 47% of patients (DAS28 < 2.6); 19% of patients had low disease activity and 34% had moderate to high disease activity (DAS28 > 3.2). Over time there was little change in the DAS28 on a group level, while on the individual level, patients showed changes in their DAS28 score (Figure 1). Table 2 shows the percentage of patients changing. Between 8% and 14% of the patients had a change to high disease activity (DAS28 > 3.2) at any of the timepoints. Continuous low disease activity was seen in 48% (n = 76) of the patients.

*Self-reported measures*. Self-reported physical functioning on the HAQ resulted in a median of 0.50 (IQR 0.125–1.00) at baseline. Self-reported disease activity (RADAI) was low with a median of 2.0 (IQR 0.84–2.99). General health (VAS global) was scored at 39 (IQR 21–58), just below half of the scale, while fatigue had a median value of 50 (IQR 29–70). The median values of the different PRO did not evolve much over time, while individuals showed substantial variation, as reflected in Figure 1 and Table 3. Figure 1 shows the variation in self-reported measures rescaled to 0–10 scores for 6 patients (who all reflected different patterns), and Table 3 provides the minimal clinical important worsening for 3 subsequent monthly measures.

*Association between PRO and DAS28*. We used various ways to estimate the relationship between PRO and DAS28. First, we looked at group level at the correlation between DAS28 and RADAI, HAQ, and VAS fatigue. The Spearman rank correlation coefficient varied between 0.29 and 0.51, with the lowest correlation for the VAS fatigue. Second, we categorized the DAS28 in disease activity states and summarized the PRO per disease state. This provided incremental values of the PRO for more active disease (Table 4). Third, we tested the discriminatory properties of the PRO using their proposed cutpoints to identify patients with moderate to high levels of disease activity (DAS28 > 3.2). For the RADAI (cutpoint < 2.2), sensitivity at timepoint 2 for the DAS28 at timepoint 3 was 0.63 (CI 0.48–0.77) and the specificity 0.71 (CI 0.59–0.79). For the HAQ, sensitivity was 0.43 (0.29–0.59) and specificity was 0.90 (0.81–0.96).

*Change of PRO to predict DAS28*. Our aim was to identify patients with a DAS28 > 3.2 by the changes of the PRO over time. We therefore modeled the DAS28 at Month 3 using the combined scores and changes over time of the HAQ, the RADAI, and the VAS fatigue. This was done using a joint modeling technique of the Bayesian approach. The RADAI and HAQ of the random intercept had a "Bayesianly significant" positive relationship with the DAS28 at Month 3, while the variable for male sex had a significantly negative relationship (Table 5). In addition, sex and coping with pain
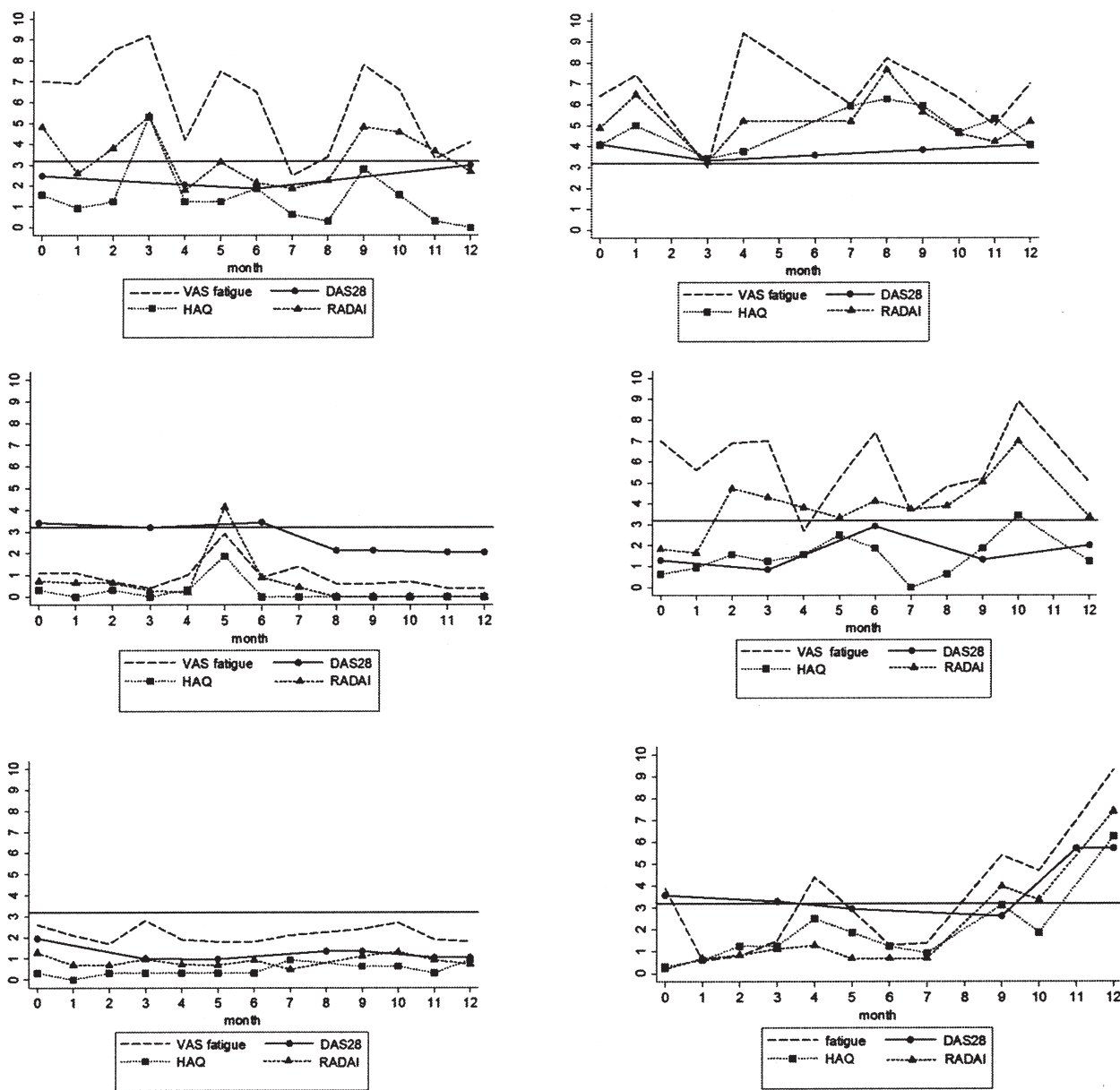
*Figure 1*. Individual patterns of 28-joint Disease Activity Score (DAS28) and patient-reported outcomes from 6 patients. HAQ: Health Assessment Questionnaire; RADAI: Rheumatoid Arthritis Disease Activity Index; VAS: visual analog scale.

*Table 2*. DAS28 changes over time presented as number of patients with a change (%). For T0–T3, values are presented for patients who had data for both timepoints.

| Time period (mos) | Change ≤ 0.6 | > 0.6 and ≤ 1.2 | > 1.2 | DAS28 ≤ 3.2 | DAS28 > 3.2 |
|---|---|---|---|---|---|
| T0–T3, n = 146 | 81 (56.3) | 43 (30.0) | 20 (13.9) | 16 (11.1) | 15 (10.4) |
| T3–T6, n = 130 | 77 (59.2) | 30 (23.1) | 23 (17.7) | 18 (13.9) | 8 (6.2) |
| T6–T9, n = 112 | 61 (54.6) | 30 (26.8) | 21 (18.8) | 9 (8.0) | 16 (14.3) |
| T9–T12, n = 102 | 58 (56.9) | 25 (24.5) | 19 (18.6) | 10 (9.8) | 8 (7.8) |

DAS28: 28-joint Disease Activity Score.

*Table 3.* No. patients (%) with a worsening on the patient-reported outcome using the minimal clinically important difference (MCID). Values are given for patients who had data for both timepoints.

| Time Period | HAQ, MCID: 0.2 | RADAI, MCID: 1.0 | VAS Fatigue, MCID: 10 mm | VAS Global, MCID: 10 mm |
|---|---|---|---|---|
| T0–T1, n = 140 | 33 (23.6) | 66 (46.1) | 47, n = 138 (34.1) | 35 (25.0) |
| T1–T2, n = 133 | 25 (18.8) | 22 (16.3) | 30, n = 130 (23.1) | 30 (22.9) |
| T2–T3, n = 135 | 18 (11.1) | 18 (13.5) | 37, n = 147 (25.2) | 25 (18.6) |
| T0–T3, n = 142 | 27 (19) | 43 (30.3) | 36, n = 145 (24.8) | 40 (28.1) |

HAQ: Health Assessment Questionnaire; RADAI: Rheumatoid Arthritis Disease Activity Index; VAS: visual analog scale.

*Table 4.* Patient-reported outcomes stratified for disease activity states according to the DAS28. Data are median (IQR).

| DAS28 | HAQ | RADAI | VAS Fatigue |
|---|---|---|---|
| < 2.6, n = 72 | 0.19 (0.00–0.50) | 1.09 (0.60–1.82) | 43 (20–66) |
| 2.6 ≤ 3.2, n = 29 | 0.50 (0.25–1.12) | 1.47 (0.92–2.20) | 53 (21–79) |
| > 3.2 ≤ 5.1, n = 39 | 0.75 (0.38–1.38) | 2.73 (1.46–3.61) | 59 (40–70) |
| > 5.1, n = 9 | 1.25 (1.00–2.13) | 4.53 (4.05–4.81) | 80 (75–82) |
| Overall, n = 159 | 0.50 (0.13–1.00) | 1.63 (0.84–2.99) | 56 (29–70) |

DAS28: 28-joint Disease Activity Scale; HAQ: Health Assessment Questionniare; RADAI: Rheumatoid Arthritis Disease Activity Index; VAS: visual analog scale; IQR: interquartile range.

*Table 5.* Bounded outcome score responses for HAQ, RADAI, and VAS fatigue.

| Variable | | DAS T3 vs 3 PRO, Months 0, 1, 2 | | |
|---|---|---|---|---|
| | | Mean | SE | 95% CI |
| Intercept | $\beta_0$ | 3.04 | 0.53 | (2.01, 4.06) |
| HAQ | | | | |
|   Intercept | $\beta_1$ | 0.19 | 0.09 | (0.02, 0.36) |
|   Slope | $\beta_2$ | –0.06 | 0.96 | (–1.99, 1.85) |
| RADAI | | | | |
|   Intercept | $\beta_3$ | 0.28 | 0.12 | (0.05, 0.53) |
|   Slope | $\beta_4$ | 1.08 | 1.02 | (–0.04, 4.12) |
| VAS fatigue | | | | |
|   Intercept | $\beta_5$ | –0.01 | 0.11 | (–0.23, 0.20) |
|   Slope | $\beta_6$ | –1.95 | 2.24 | (–7.33, 1.33) |
| Age | $\beta_7$ | 0.01 | 0.01 | (–0.01, 0.02) |
| Sex (male) | $\beta_8$ | –0.78 | 0.25 | (–1.26, –0.29) |
| Self-efficacy | $\beta_9$ | –0.004 | 0.01 | (–0.02, 0.02) |
| Coping pain | $\beta_{10}$ | –0.10 | 0.05 | (–0.20, –0.0002) |
| Coping optimism | $\beta_{11}$ | –0.02 | 0.05 | (–0.11, 0.07) |
| Control illness | $\beta_{12}$ | –0.02 | 0.04 | (–0.09, 0.05) |
| Experience symptoms | $\beta_{13}$ | 0.07 | 0.04 | (–0.01, 0.16) |
| Sensitivity | | 0.61* (0.81) | | |
| Specificity | | 0.75* (0.92) | | |

* Results of the 10-fold cross-validated model (initial values of the development model). Estimate = posterior mean of estimated regression coefficient. HAQ: Health Assessment Questionnaire; RADAI: Rheumatoid Arthritis Disease Activity Index; VAS: visual analog scale; DAS: Disease Activity Score; PRO: patient-reported outcome.

were significantly related (Table 5). Our initial development model would correctly identify 81% of the patients with a DAS28 > 3.2 and 92% of the patients with a DAS ≤ 3.2. However, after 10-fold cross-validation, a technique that corrects for overoptimism in the development model, 61% of the DAS28 > 3.2 patients were correctly identified and 75% of the patients with a DAS28 ≤ 3.2. The positive likelihood ratio was 2.7 and the negative likelihood ratio 0.51.

**DISCUSSION**
PRO are valuable tools in the clinic to guide treatment in

addition to disease activity measures performed by the physician. We aimed to assess the predictive capacity of PRO in relation to the subsequent scores on the DAS28. Moderate sensitivity and specificity were seen for the performance of 1 single PRO as primary index. Combining the monthly measurements of 3 PRO using an advanced Bayesian statistical model, taking into account the score and the change in PRO over time, the sensitivity to identify patients with a DAS28 > 3.2 at Month 3 was 0.61 and the specificity was 0.75. From discussions with rheumatologists we know that there is a strong need to reduce the pressure on their schedules. A possible way they would like to do that is by observing the patients with PRO, especially those with low levels of clinical disease activity. The patients could complete the PRO at home and with stable levels of the PRO, they only need to come in once a year for clinical assessment. With the current instruments this approach is not feasible, because a substantial portion of patients with high levels of clinical disease activity would be missed.

To our knowledge, this is the first study to investigate distant monitoring of disease activity using Web-based PRO to predict a high disease activity state. Although there are studies that investigated the relation between PRO and disease activity, none of them evaluated the predictive value of PRO. One study used the minimum clinically important difference on PRO to predict treatment response on a subsequent timepoint but did not report any risk measure that quantified the contributions of PRO in the prediction of low disease activity[33].

An important finding of our study was that individual patients showed fluctuating patterns of PRO instruments and the DAS28 (Figure 1). On a group level, the measurement did not evolve much over time. This individual fluctuation on PRO was also demonstrated by Blanchais, *et al*, using the RAPID4/3 weekly[34] (patient-reported measurement: physical function, pain, global estimate, and self-reported joint counts). One would expect that an increase of PRO values over time that indicates patient-reported disease worsening would result in a high disease activity state as measured by the DAS28 at a subsequent timepoint. This was not the case, however. PRO slopes (changes over time), which were captured in the Bayesian model, did not significantly contribute to the prediction of high disease activity. This may indicate several things. First, it could be that the time frame of the study should be weekly rather than monthly, to measure closer to the moment of clinical disease flare. Second, 3 timepoints that reflect an individual patient trajectory may be too few. The possible change of medication after each clinical visit limited us to use more than 3 consecutive monthly measures to predict the outcome of the subsequent clinical DAS28. Maybe the use of more measurements in this model would improve prediction of individual outcomes. Third, the fluctuating patterns may be influenced by other factors

unrelated to disease activity (e.g., comorbidity). This may be especially true for fatigue, which seems driven not only by disease activity.

Because we suspected that several personality aspects influence the relationship between DAS and PRO, we measured coping, self-efficacy, and illness perception at baseline. However, in the analysis most of the measured personality aspects did not contribute to prediction of the model. Only sex and coping with pain were significant.

In our study we made choices that may raise issues for discussion. First was our choice of the DAS28 to measure high disease activity. In daily clinical practice the DAS28 > 3.2 is commonly used for treatment intensification, and when our study was designed, it was probably still a valid option. Regarding the recent discussions within the OMERACT working group on flare, this may be a conservative way to identify patients needing treatment intensification. In their view, shared by us, a flare represents a change in multiple variables that requires treatment change[35]. These variables are patient global assessment, pain, swollen joints, tender joints, function, physician global assessment, and fatigue[36]. Second, there are several PRO available (RAPID3, PASS, RADAI-5, and Rheumatoid Arthritis Impact of Disease) that could be used for the evaluation of disease activity, but there is no consensus what would be best to use. Hence, we chose PRO that were familiar to us, which were validated (in Dutch) and reflecting different domains of disease activity: HAQ, VAS global, VAS fatigue, and RADAI. Reanalyzing the data with the RADAI-5[37,38] and a modified version of the RAPID3[39] did not alter the observation of moderate performance of the PRO. In an additional cross-sectional analysis (time 3 to time 3), the Spearman rank correlation coefficient between the DAS28 and the modified RAPID3 was 0.54. The correlation between the DAS28 and RADAI-5 was 0.49. In the longitudinal analysis, the predictive value (time 2 PRO predicting time 3 DAS28) resulted in a sensitivity of the modified RAPID3 of 0.29 and 0.24 for the RADAI-5. The specificity was 0.84 for the RAPID3 and 0.86 for the RADAI-5.

Limitations of our study include the choice of the time frame for clinical evaluation and the limited number of patients that had a relevant change in DAS28 score. Ideally, we would have liked to evaluate the DAS28 each month as we did with the PRO. However, that meant that patients had to come to the clinic each month, which was regarded as not feasible for them. We therefore decided to go with a time frame of every 3 months. Change in DAS28 > 1.2, which we regarded as relevant, occurred in 20 patients in the first 3 months. Which is good from a clinical viewpoint because in most patients disease was under control, but from a prediction viewpoint this change may be too little to have sufficient power to assess the effect of the PRO. One way to solve this is to study larger samples, because levels of

disease activity in patients with disease well under control are expected to change little. In addition, we had to deal with missing values on both the clinical DAS28 as well as the PRO. The missing values on the PRO were likely related to the high frequency of measurements (13 timepoints). The monthly questionnaires were too high a burden for some patients. To reflect the observable patterns in the data, we decided not to impute data if patients had one timepoint of PRO data missing. The problem of missing data on the outcome (DAS28) was solved by the Bayesian model.

A strength of our study was using a Web-based environment allowing patients to assess their disease activity at home. Collecting data in a Web-based environment has been tested and validated before on patients with a rheumatic disease[40,41,42,43], and worked well in our study. Computerized versions offer advantages over the paper version (less time-consuming) without compromising on the data validity.

PRO are very valuable and may give additional information about the patients. PRO showed fluctuating levels of disease activity over time, while on a group level, disease activity stayed the same. Using the score and changes in RADAI, HAQ, and VAS fatigue over time to predict future DAS28 disease activity (moderate to high) resulted in sensitivity of 0.81 and a specificity of 0.92 in the development set. However, the internal cross validation of the model resulted in moderate performance (sensitivity 0.61, specificity 0.75). Further research is needed to investigate the possibilities of using PRO as predictors for clinical disease activity.

## REFERENCES

1. van der Heijde DM, van 't Hof MA, van Riel PL, Theunisse LA, Lubberts EW, van Leeuwen MA, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Ann Rheum Dis 1990;49:916-20.
2. 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:S100-8.
3. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. J Rheumatol 1982;9:789-93.
4. Fransen J, Langenegger T, Michel BA, Stucki G. Feasibility and validity of the RADAI, a self-administered rheumatoid arthritis disease activity index. Rheumatology 2000;39:321-7.
5. Ware JE Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
6. Riemsma RP, Taal E, Rasker JJ, Houtman PM, Van Paassen HC, Wiegman O. Evaluation of a Dutch version of the AIMS2 for patients with rheumatoid arthritis. Br J Rheumatol 1996;35:755-60.
7. Heiberg T, Kvien TK, Mowinckel P, Aletaha D, Smolen JS, Hagen KB. 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.
8. Fransen J, van Riel PL. Outcome measures in inflammatory rheumatic diseases. Arthritis Res Ther 2009;11:244.
9. Schipper LG, Hoekstra M, Vliet Vlieland TP, Jansen TL, Lems WF, van Riel P; Nederlandse Vereniging van Reumatologie. Practice guideline 'Diagnosis and treatment of rheumatoid arthritis'. [Dutch]. Ned Tijdschr Geneeskd 2009;153:A944.
10. Kalyoncu U, Dougados M, Daurès 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.
11. Pincus T, Sokka T. Quantitative measures for assessing rheumatoid arthritis in clinical trials and clinical care. Best Pract Res Clin Rheumatol 2003;17:753-81.
12. Barton JL, Criswell LA, Kaiser R, Chen YH, Schillinger D. Systematic review and metaanalysis of patient self-report versus trained assessor joint counts in rheumatoid arthritis. J Rheumatol 2009;36:2635-41.
13. Choy EH, Khoshaba B, Cooper D, MacGregor A, Scott DL. Development and validation of a patient-based disease activity score in rheumatoid arthritis that can be used in clinical trials and routine practice. Arthritis Rheum 2008;59:192-9.
14. Fransen J, van Riel PL. DAS remission cut points. Clin Exp Rheumatol 2006;6 Suppl 43:S29-32.
15. Anderson JJ, Chernoff MC. Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. J Rheumatol 1993; 20:535-7.
16. Wolfe F. Fatigue assessments in rheumatoid arthritis: comparative performance of visual analog scales and longer fatigue questionnaires in 7760 patients. J Rheumatol 2004;31:1896-902.
17. Scott DL, Strand V. The effects of disease-modifying anti-rheumatic drugs on the Health Assessment Questionnaire score. Lessons from the leflunomide clinical trials database. Rheumatology 2002;41:899-909.
18. Greenwood MC, Doyle DV, Ensor M. Does the Stanford Health Assessment Questionnaire have potential as a monitoring tool for subjects with rheumatoid arthritis? Ann Rheum Dis 2001;60:344-8.
19. Aletaha D, Machold KP, Nell VP, Smolen JS. The perception of rheumatoid arthritis core set measures by rheumatologists. Results of a survey. Rheumatology 2006;45:1133-9.
20. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: dimensions and practical applications. Health Qual Life Outcomes 2003;1:20.
21. Wolfe F. A reappraisal of HAQ disability in rheumatoid arthritis. Arthritis Rheum 2000;43:2751-61.
22. 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.
23. Hewlett S, Dures E, Almeida C. Measures of fatigue: Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAF MDQ), Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAF NRS) for Severity, Effect, and Coping, Chalder Fatigue Questionnaire (CFQ), Checklist Individual Strength (CIS20R and CIS8R), Fatigue Severity Scale (FSS), Functional Assessment Chronic Illness Therapy (Fatigue) (FACIT-F), Multi-Dimensional Assessment of Fatigue (MAF), Multi-Dimensional Fatigue Inventory (MFI), Pediatric Quality Of Life (PedsQL) Multi-Dimensional Fatigue Scale, Profile of Fatigue (ProF), Short Form 36 Vitality Subscale (SF-36 VT), and Visual Analog Scales (VAS). Arthritis Care Res 2011;63 Suppl 11:S263-86.
24. Khanna D, Pope JE, Khanna PP, Maloney M, Samedi N, Norrie D, et al. The minimally important difference for the fatigue visual analog scale in patients with rheumatoid arthritis followed in an academic clinical practice. J Rheumatol 2008;35:2339-43.
25. Wells G, Li T, Maxwell L, MacLean R, Tugwell P. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. J Rheumatol 2007;34:280-9.

26. van Lankveld W, van't Pad Bosch P, van de Putte L, Näring G, van der Staak C. Disease-specific stressors in rheumatoid arthritis: coping and well-being. Br J Rheumatol 1994;33:1067-73.

27. Taal E, Riemsma RP, Brus HL, Seydel ER, Rasker JJ, Wiegman O. Group education for patients with rheumatoid arthritis. Patient Educ Couns 1993;20:177-87.

28. Scharloo M, Kaptein AA, Weinman J, Hazes JM, Willems LN, Bergman W, et al. Illness perceptions, coping and functioning in patients with rheumatoid arthritis, chronic obstructive pulmonary disease and psoriasis. J Psychosom Res 1998;44:573-85.

29. ten Wolde S, Hermans J, Breedveld FC, Dijkmans BA. Effect of resumption of second line drugs in patients with rheumatoid arthritis that flared up after treatment discontinuation. Ann Rheum Dis 1997;56:235-9.

30. Molas M, Lesaffre E. A comparison of three random effects approaches to analyze repeated bounded outcome scores with an application in a stroke revalidation study. Stat Med 2008;27:6612-33.

31. Steyerberg E. Clinical prediction models. A practical approach to development, validation, and updating. Rotterdam: Springer; 2009:497.

32. Lesaffre E, Rizopoulos D, Tsonaka R. The logistic transform for bounded outcome scores. Biostatistics 2007;8:72-85.

33. Strand V, Smolen JS, van Vollenhoven RF, Mease P, Burmester GR, Hiepe F, et al. Certolizumab pegol plus methotrexate provides broad relief from the burden of rheumatoid arthritis: analysis of patient-reported outcomes from the RAPID 2 trial. Ann Rheum Dis 2011;70:996-1002.

34. Blanchais A, Berthelot JM, Fontenoy AM, le Goff B, Maugars Y. Weekly home self-assessment of RAPID-4/3 scores in rheumatoid arthritis: a 6-month study in 26 patients. Joint Bone Spine 2010;77:582-7.

35. 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.

36. Bingham CO 3rd, Alten R, Bartlett SJ, Bykerk VP, Brooks PM, Choy E, et al; OMERACT RA Flare Definition Working Group. Identifying preliminary domains to detect and measure rheumatoid arthritis flares: report of the OMERACT 10 RA Flare Workshop. J Rheumatol 2011;38:1751-8.

37. Leeb BF, Haindl PM, Maktari A, Nothnagl T, Rintelen B. Patient-centered rheumatoid arthritis disease activity assessment by a modified RADAI. J Rheumatol 2008;35:1294-9.

38. Bossert M, Prati C, Vidal C, Bongain S, Toussirot E, Wendling D. Evaluation of self-report questionnaires for assessing rheumatoid arthritis activity: a cross-sectional study of RAPID3 and RADAI5 and flare detection in 200 patients. Joint Bone Spine 2012;79:57-62.

39. Pincus T, Swearingen CJ, Bergman M, Yazici Y. RAPID3 (Routine Assessment of Patient Index Data 3), a rheumatoid arthritis index without formal joint counts for routine care: proposed severity categories compared to disease activity score and clinical disease activity index categories. J Rheumatol 2008;35:2136-47.

40. Athale N, Sturley A, Skoczen S, Kavanaugh A, Lenert L. A web-compatible instrument for measuring self-reported disease activity in arthritis. J Rheumatol 2004;31:223-8.

41. Heiberg T, Kvien TK, Dale Ø, Mowinckel P, Aanerud GJ, Songe-Møller AB, et al. Daily health status registration (patient diary) in patients with rheumatoid arthritis: a comparison between personal digital assistant and paper-pencil format. Arthritis Rheum 2007;57:454-60.

42. Brophy S, Hunniford T, Taylor G, Menon A, Roussou T, Calin A. Assessment of disease severity (in terms of function) using the internet. J Rheumatol 2004;31:1819-22.

43. Richter JG, Becker A, Koch T, Nixdorf M, Willers R, Monser R, et al. Self-assessments of patients via Tablet PC in routine patient care: comparison with standardised paper questionnaires. Ann Rheum Dis 2008;67:1739-41.