The perception of rheumatoid arthritis core set measures by rheumatologists. Results of a survey

D. Aletaha1,2, K. P. Machold1, V. P. K. Nell1 and J. S. Smolen1,3

Objective. To investigate the perception of values of individual core set measures by rheumatologists, and how it differs across measures and across physicians.

Methods. We designed a survey in which 44 international expert rheumatologists explicitly marked positions on the scales of seven core-set measures that in their opinion corresponded to cut-points between remission, low, moderate and high disease activity. The measures comprised swollen and tender joint counts (SJC, TJC), CRP, ESR, patient and evaluator global assessments of activity (PGA, EGA), and the Health Assessment Questionnaire Disability Index (HAQ).

Results. The interpretation of measures across physicians was most consistent for ESR and PGA, while for CRP and joint counts there was most variation. Joint counts and CRP implied active disease at lower relative values (using normalized scales) than did PGA, EGA or ESR (P < 0.01 for most comparisons; Bonferroni-adjusted Wilcoxon signed rank test), and most physicians tended to tolerate higher numbers of tender joints than swollen joints to define similar levels of disease activity. Given these cut-points, more RA patients in a typical cross-sectional cohort would be regarded as being in remission according to joint counts (SJC, 35%; TJC, 55%) than to global scores (PGA, 18%; EGA, 9%), and fewer patients would be regarded as being in remission by physician-derived or laboratory measures than by patient-derived ones.

Conclusion. These data give insights into the integrative process of activity evaluation and will be informative for future survey designs, studies using physician opinion as the gold standard for criterion validity of disease activity, and allow ‘activity mapping’ of values on different scales based on expert opinion.

KEY WORDS: Rheumatoid arthritis, Disease activity measures, Clinical perception.

Rheumatoid arthritis (RA) is a chronic inflammatory disease which is characterized by high variation in presentation between patients; likewise, in individual patients’ signs and symptoms the manifestations of disease activity may vary over time. Several measures are generally assessed in studies of RA and have been defined to form a core set of disease activity measures [1–4], which have been shown to be valuable surrogates of disease activity [5, 6].

New therapeutic concepts suggest that disease activity should be thoroughly controlled and that a state of low disease activity or remission should be the goal in clinical practice [7–9]. Assessment of the presence of such a state is frequently assessed using combinations of markers [10], since an individual marker is not likely to reflect all aspects of RA activity [11]. However, it is still unknown what the measured values of individual disease activity surrogates mean to physicians, and how they are interpreted. This valuation of measurements by physicians may be different from the prognostic implications of measurements obtained in longitudinal studies, which has in fact been shown for several activity surrogates [5, 12–15].

Until now, assessing the impact of different levels of various disease activity measures on a physician’s clinical decision has only been possible indirectly, because information usually was derived from surveys employing complete patient profiles, or from decisions of rheumatologists in clinical practice, where they had information on many clinical and laboratory variables [3, 16–20]. Neither of these approaches allows independent assessment of the extent to which measurements on individual scales drive a clinical decision, and various statistical methods have been used to disentangle the contribution of individual measures to the overall and implicit judgement of the clinician [19, 20]. In the present study, we address this issue in a direct way by using explicit ratings of expert rheumatologists of normal and abnormal values of seven disease activity measures in a specifically designed survey.

Methods

Survey

We designed a survey in which rheumatologists were asked to indicate cut-point values on the scales of the swollen and tender joint counts (SJC, TJC, based on 28-joint counts), the erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), patient and evaluator global assessments of disease activity (PGA, EGA), and for the Health Assessment Questionnaire Disability Index (HAQ). Separately for each of these seven core-set measures, the rheumatologists were asked to assign three values which in their opinion would constitute the cut-points between remission (REM), low disease activity (LDA), moderate disease activity (MDA) and high disease activity (HDA). This was done by indicating exact numbers (for SJC, TJC and the HAQ; Fig. 1A) or by marking respective 100 mm visual analogue scales (VAS) (for ESR, CRP, PGA and EGA; Fig. 1B); for ESR and CRP, 1 mm on the VAS corresponded to 1 mm/h and 1 mg/l, respectively. The instructions for this task suggested a hypothetical situation in which only the
measure in question was available for the evaluation of the disease activity of a patient with RA, and each measure was presented on a separate page to impede direct comparison. Given the seven measures and the three cut-points, a total of 21 values were specified by the physicians. While definitions of REM, LDA, MDA, and HDA may vary between rheumatologists, they are conceptually identical for all core-set measures assessed by an individual rheumatologist. Therefore, the definitions served as anchors for each individual rheumatologist across all core-set scales and allowed the valuation of measurements on the scales and their direct comparison.

Survey respondents

The survey was performed in 2004. The addressees were experienced rheumatologists, who had >5 yr of clinical practice in rheumatology and who were involved in the conduct of RA clinical trials. Forty-seven rheumatologists were asked to complete the survey. Three gave incomplete answers and were excluded from the analysis. The remaining 44 rheumatologists were from 15 countries (14% US, 18% UK and 68% other parts of Europe); their median experience in rheumatology was 20 yr (range 6–35 yr).

Comparative analysis using normalized scales

From the survey responses, we calculated the medians, quartiles, 10th and 90th percentiles, and ranges of values for the three cut-points (REM vs LDA, LDA vs MDA, and MDA vs HDA) indicated for each of the core-set measures. Since the scales for these measures are different, we normalized their values to a scale of 0–100 (i.e. for each measure the proportion of the maximum possible value was used instead of the actual value). This allowed a more direct comparison. For ESR and CRP, which do not have a natural upper limit, we defined the maximum values as 100 mm and 100 mg/l, respectively; this was in accordance with the presentation of their scales on a 100 mm VAS in the survey (Fig. 1B). We used the Wilcoxon signed rank test, a paired non-parametric test, to statistically assess the within-respondent variability in perception of the different scales. We performed these comparisons for the levels of all pairs of the seven core-set parameters, separately for each cut-off and the two other cut-points. We adjusted P-values conservatively for multiple comparisons using the Bonferroni method (Pc).

Consideration of distributional differences

For the analysis of the scales were normalized to range from 0 to 100, distributional differences concerning some of the measures might have influenced the physicians’ responses in the survey. These include aspects such as floor effects, ceiling effects and incomplete coverage of the potential measurement scales. Therefore, we used the median cut-points as obtained from the surveys to perform further cross-sectional analyses in a routine out-patient cohort of 767 patients with RA (80% female, 55% RF-positive). We used the first documented visit of every patient with complete data available. These patients had an average duration of RA of 8.1 ± 10.6 yr and a median (quartiles) disease activity of 4.1 (3.0, 5.2) by the 28-joint Disease Activity Score (DAS28) and of 16.7 (8.1, 26.7) by the Simplified Disease Activity Index (SDAI) [21]; detailed characteristics of this cohort have been published [12]. We identified the distribution of patients to the four disease activity states individually for each core-set measure, and compared the proportions of patients in the different activity states between the measures. We quantified agreement between the classifications by different measures using kappa statistics. Kappa values of 0 indicate perfect agreement, and kappa values of 1 indicate agreement as expected by chance.

Structural implications of the responses across variables

Next we used the survey results to analyse the consistency of their implications in 65 patients of an RA inception cohort followed over a period of 3 yr (78.5% female; 64.6% RF-positive); the mean (s.d.) age of the patients was 51.9 ± 15.9 yr, and the mean (s.d.) duration of RA at baseline was 11.5 ± 12.5 weeks; all patients fulfilled the American College of Rheumatology criteria [22]. Patients were untreated at baseline, with a median (quartiles) DAS28 of 5.5 (4.8, 6.3) and an SDAI of 37.6 (22.9, 56.9), and were offered disease-modifying anti-rheumatic drugs within a few weeks of presentation. Patients were clinically assessed every 3 months for 3 years; at that end-point they showed low to moderate disease activity with a median DAS28 of 3.3 (2.5, 4.5) and a median SDAI of 10.4 (3.9, 20.1). Radiographic damage was scored by the Larsen method [23] and had a median (quartiles) of 2 (0, 8) at baseline and 14 (3, 32) at year 3; detailed characteristics of this cohort have been published [24, 25].
We calculated changes in Larsen scores between baseline and the 3-yr follow-up for all 65 patients and assessed whether disease activity states that are based on different variables are different with respect to their association with radiographic progression over time. The median cut-point value obtained for each survey item was used as the most robust measure of the physicians’ opinion in classifying patients as in REM, LDA, MDA and HDA. We then identified patients who spent more than 50% of the observation period (i.e. >18 months) in REM (activity status 1), LDA (activity status 2) or MDA or HDA (activity status 3). Finally, we assessed statistically whether there was an overall difference in progression of joint damage between patients in the various activity status (1, 2 or 3) after adjusting for differences in the core-set measures, and whether there was a difference between measures after adjusting for the different statuses. We implemented these analyses using a two-way analysis of variance model, which allowed assessment of the independent effects of the predictors ‘activity status’ and ‘core-set measure’ on the outcome. Using the interaction term activity status × core-set measure, we tested whether there was a difference in the association with progression across different measures.

**Results**

**Survey results and variation of responses**

Table 1 summarizes the responses given by the rheumatologists. For each cut-point, the responses on different measures mapped to the same clinical situation and therefore allowed between-variable comparisons. Although concepts of remission and other disease activity states vary between rheumatologists, it was expected that they would be constant in individual physicians during their completion of the various survey items. Swollen and tender joint counts mapped to similar values on the 28-joint scales, the median TJC being one unit higher at the lower and upper ends of the scale. Looking at serological surrogates of disease activity, median ESR values of 14, 32 and 51 mm were seen to correspond to CRP values of 5, 20 and 40 mg/l. The PGA ratings were numerically slightly higher than EGA ratings for all three cut-points. If only HAQ was available for the evaluation of an individual patient with RA, the rheumatologists rated the median cut-point for REM vs LDA as 0.5, that for LDA vs MDA as 1.0, and that for MDA vs HAD as 1.6.

To assess the variation of responses across rheumatologists, we calculated the relative standard deviations [coefficient of variation (CV)] for each measure and separately for each cut-point. The overall CV was 49.5%. We found that, at each cut-point level, the interpretation of ESR and PGA showed the smallest variance across rheumatologists, while CRP and joint counts showed the greatest variation (detailed data not shown). Comparing the variation in perception between physicians from North America, the UK and the rest of Europe, there was not much difference (CV = 48.2, 55.8 and 44.5%, respectively).

### Comparative analysis of normalized scales

To enable comparison of survey responses on the different measures, we normalized their values to a scale of 0–100, obtaining the percentage of maximal possible score. Figure 2A, B and C reveal that joint counts and CRP implied active disease at lower relative values than patient and evaluator global scores and ESR (corrected P < 0.01 for most comparisons; Fig. 2). More physicians tolerated higher TJC than SJC for all three cut-points (data not shown), although these differences were statistically significant only at the LDA/MDA cut-point (Fig. 2). In contrast, more stringent clinical meanings (i.e. lower cut-points) were assigned to values of EGA compared with PGA. Based on the physicians’ responses on each pair of two measures, this numerical difference

**Table 1. Indicated cut-offs for different disease activity states**

<table>
<thead>
<tr>
<th>Measure</th>
<th>10th percentile</th>
<th>25th percentile</th>
<th>50th percentile</th>
<th>75th percentile</th>
<th>90th percentile</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Swollen joint count (0–28 joints)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remission/low disease activity</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>2.0</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Low/moderate disease activity</td>
<td>2.3</td>
<td>4.0</td>
<td>5.0</td>
<td>6.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Moderate/high disease activity</td>
<td>4.6</td>
<td>7.0</td>
<td>9.0</td>
<td>12.0</td>
<td>15.0</td>
<td>20.0</td>
</tr>
<tr>
<td><strong>Tender joint count (0–28 joints)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remission/low disease activity</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
<td>3.0</td>
<td>5.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Low/moderate disease activity</td>
<td>1.4</td>
<td>4.0</td>
<td>5.0</td>
<td>8.0</td>
<td>10.0</td>
<td>14.0</td>
</tr>
<tr>
<td>Moderate/high disease activity</td>
<td>6.0</td>
<td>8.0</td>
<td>10.0</td>
<td>12.0</td>
<td>15.0</td>
<td>20.0</td>
</tr>
<tr>
<td><strong>Erythrocyte sedimentation rate (mm/h)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remission/low disease activity</td>
<td>2.5</td>
<td>5.0</td>
<td>14.0</td>
<td>19.5</td>
<td>25.0</td>
<td>30.0</td>
</tr>
<tr>
<td>Low/moderate disease activity</td>
<td>8.0</td>
<td>20.0</td>
<td>31.5</td>
<td>38.0</td>
<td>45.0</td>
<td>59.0</td>
</tr>
<tr>
<td>Moderate/high disease activity</td>
<td>29.0</td>
<td>41.0</td>
<td>51.0</td>
<td>62.0</td>
<td>65.0</td>
<td>75.0</td>
</tr>
<tr>
<td><strong>C-reactive protein (mg/l)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remission/low disease activity</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>10.0</td>
<td>14.5</td>
<td>32.0</td>
</tr>
<tr>
<td>Low/moderate disease activity</td>
<td>8.0</td>
<td>10.5</td>
<td>20.5</td>
<td>27.0</td>
<td>33.0</td>
<td>58.0</td>
</tr>
<tr>
<td>Moderate/high disease activity</td>
<td>10.0</td>
<td>21.5</td>
<td>31.0</td>
<td>40.0</td>
<td>50.0</td>
<td>65.0</td>
</tr>
<tr>
<td><strong>Patient assessment of activity (0–100 VAS)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remission/low disease activity</td>
<td>0.0</td>
<td>4.5</td>
<td>12.5</td>
<td>19.0</td>
<td>24.5</td>
<td>39.0</td>
</tr>
<tr>
<td>Low/moderate disease activity</td>
<td>6.0</td>
<td>21.0</td>
<td>32.0</td>
<td>42.0</td>
<td>48.0</td>
<td>51.0</td>
</tr>
<tr>
<td>Moderate/high disease activity</td>
<td>23.0</td>
<td>42.5</td>
<td>55.5</td>
<td>65.0</td>
<td>70.0</td>
<td>81.0</td>
</tr>
<tr>
<td><strong>Evaluator assessment of activity (0–100 VAS)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remission/low disease activity</td>
<td>0.0</td>
<td>0.0</td>
<td>8.5</td>
<td>12.0</td>
<td>20.0</td>
<td>37.0</td>
</tr>
<tr>
<td>Low/moderate disease activity</td>
<td>0.0</td>
<td>12.0</td>
<td>29.0</td>
<td>35.0</td>
<td>40.0</td>
<td>55.0</td>
</tr>
<tr>
<td>Moderate/high disease activity</td>
<td>27.0</td>
<td>39.0</td>
<td>54.0</td>
<td>66.0</td>
<td>66.5</td>
<td>72.0</td>
</tr>
<tr>
<td><strong>Health assessment questionnaire disability index (0–3)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remission/low disease activity</td>
<td>0.0</td>
<td>0.1</td>
<td>0.5</td>
<td>0.5</td>
<td>1.1</td>
<td>1.5</td>
</tr>
<tr>
<td>Low/moderate disease activity</td>
<td>0.5</td>
<td>0.6</td>
<td>1.0</td>
<td>1.5</td>
<td>1.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Moderate/high disease activity</td>
<td>0.9</td>
<td>1.0</td>
<td>1.4</td>
<td>1.6</td>
<td>2.0</td>
<td>2.8</td>
</tr>
</tbody>
</table>

*aIf the indicated ESR was <2 mm, ESR was set to 2 mm (suggested lowest possible ESR value). bIf the indicated CRP was <5 mg/l, CRP was set to 5 mg/l (lower limit of detection for most routine laboratories).
was most pronounced in the higher disease activity area (Fig. 2A: P < 0.1; Fig. 2B, P < 0.05; Fig. 2C, P < 0.01). Values on the HAQ scale had clinical implications comparable to values of PGA, EGA or ESR on their respective scales. The survey responses on all seven measures showed consistent patterns when different levels of disease activity were considered by the physicians (Fig. 2A–C).

Consideration of distributional differences

We next plotted the cumulative distributions of the measurements on each of the core-set scales (Fig. 3A–G) as observed cross-sectionally in a cohort of 767 RA patients, and applied the median cut-points from the survey to obtain proportions of patients that would fall into the four categories. This allowed consideration of the physicians' potential assumption of distributional differences of measures in this comparative analysis, e.g. that joint counts might only rarely occur in the very high range. In fact, from the panels in Fig. 3 it can be seen that the measurements of SJC and TJC, as well as CRP, show floor effects, with a considerable proportion of individuals (>60%) at the bottom end (lowest 20%) of the respective scales and only a few (<5%) at the high end (highest 10%) (Fig. 3A, B and D). On the other hand, measurements of the EGA (Fig. 3F) were more proportionally distributed in the lower two-thirds of the scale but did not cover the whole possible range, only a few observations falling into the upper third. This also contrasts with the PGA (Fig. 3E), for which about 20% of measurements fall into that part of the scale. The scales of the HAQ and PGA (Fig. 3E and G) seem to best cover the full range from the low to the high end in clinical practice.

In line with these distributional characteristics of the various measures in the observational cohort, we found that applying the median cut-points from the surveys led to considerably smaller proportions of patients classified in remission by global assessment scales (Fig. 3E and F; PGA, 17.6%; EGA, 9.3%) than by joint counts (Fig. 3A and B; SJC, 35.2%; TJC, 54.9%); similarly, this was true for the LDA cut-points. This suggests that the physicians applied relatively more stringency when judging criteria for the global assessment scales than for other core-set measures. Also, measures based on patient reporting have been regarded less stringently than physician-derived measures (TJC vs SJC; PGA vs EGA; Fig. 3). The comparison of CRP and ESR indicates a floor effect for CRP. However, although the cut-point for remission was set at the lower level of detection for CRP (5 mg/l) by the physicians, it identified a similar proportion of patients as the respective ESR cut-point (27.1 vs 28.3%). This indicates that the use of more sensitive methods of CRP measurement might not have important clinical implications in RA. However, although the proportions of patients in remission were similar for CRP and ESR, these were not necessarily the same patients: only 45% of maximum possible agreement on classifications beyond chance was achieved (Fig. 3H). For most other pairs of measures, this agreement was even lower (Fig. 3H).

Structural implications of the responses across variables

We calculated mean changes in Larsen scores over the 3-yr period. We used the median responses from the survey to define REM, LDA, MDA and HDA. According to the degree of activity that was present for most of the observation period (i.e., >18 months), we categorized patients into activity status 1 (REM for >18 months), 2 (LDA for >18 months) or 3 (MDA or HDA >18 months), separately for each core-set measure. Higher activity status was significantly associated with poorer outcome across all core-set measures (two-way ANOVA; within-measure statistics, df = 2; F = 27.4; P < 0.001; Fig. 4). Importantly, the associations of core-set measures with outcome were similar (across-measure statistics, df = 6; F = 1.3; P = 0.26). Interestingly, however, among the individual core-set measures (using the cut-points from the survey) the HAQ did not differentiate well between radiographic progression in patients who spent >50% in LDA and those in MDA/HDA, while CRP and EGA had the steepest gradients among the disease activity statuses (Fig. 4). However, these differences in within-measure associations were not statistically different across measures (interaction of activity status × measure, df = 12; F = 0.48; P = 0.93).
The results of our study indicate that there are differences in the perception of individual disease activity measures by physicians. Since such data cannot be extracted from implicit opinions, our study is the first to allow direct comparison between measures and the comparative assessment (‘activity mapping’) of their values. The interpretation by the physicians was most consistent for ESR and PGA, while for CRP and joint counts there was most variation, and there were no indications of major intercultural differences in this respect. Joint counts implied active disease at lower relative values than other measures, but when these cut-points were applied to a typical cohort of patients, still higher proportions of remissions were found than by using cut-points obtained for other measures. This indicates that, especially, tender joint counts could be valued even more stringently than in this survey to match the distributional set-up of the other disease activity characteristics in a typical cohort of RA patients. Likewise, more patients are regarded as in remission when cut-points and distributions of patient-derived measures are analysed, indicating that these measures could be better harmonized with others by more stringent evaluation by physicians. The median cut-point for HAQ remission was 0.5, which is reasonable considering the average age, disease duration and potential comorbidities of patients with RA. However, HAQ scores integrate activity and damage [14, 26–28], and therefore measure different constructs in early and late RA, but differences in disease duration were not considered by the physicians and also were not part of this study. The HAQ cut-points between low and moderate and between moderate and high disease activity were 1.0 and 1.6, respectively.

Information on the perception of individual measures has not been available hitherto. It may, however, be valuable in the choice of parameters to be measured in clinical studies, as well as in the interpretation of results from composite scores, which differ in the

![Fig. 3. Consideration of distributional differences. Data are based on single observations of 767 patients seen during routine clinical practice. (A–G) Cumulative proportion of patients according to increasing levels of seven core-set measures. The dotted vertical lines indicate the position of the median cut-points as obtained in the survey. The corresponding dotted horizontal lines mark the proportions of patients falling into REM, LDA, MDA and HDA, given the cut-points and the distribution. (H) Agreement between these classifications for pairs of measures, quantified as weighted kappa coefficients. Kappa coefficients range from 0 (agreement simply due to chance) to 1 (maximum agreement). For example, PGA and EGA showed 53% of maximum possible agreement beyond chance. ESR, 0–100 mm; CRP, 0–100 mg/l; PGA and EGA, 0–100 mm; HAQ, 0–3 SJC and TJC 0–28.

![Fig. 4. Relation of ratings to outcome. Mean and 95% confidence intervals (CI) for changes in Larsen scores over the 3-yr observation period in 65 patients in an inception cohort are shown for patients who spent most the time (i.e. >18 months) in remission (black error bars), low disease activity (dark grey error bars), or moderate and high disease activity (light grey bars), separately for the seven core-set measures. The associations with disease activity and Larsen were highly significant (P<0.001), and the associations were similar for all core-set measures (not significant). For details see text. ESR, 0–100 mm; CRP, 0–100 mg/l; PGA and EGA, 0–100 mm; HAQ, 0–3. Discussion

The results of our study indicate that there are differences in the perception of individual disease activity measures by physicians.
selection of measures and their weights [10]. Also, these data are informative if the development of new activity or response criteria is contemplated.

The data obtained in our study allow the activity mapping of values across different measures. The results suggest that the numbers of swollen and tender joints in patients with RA have similar implications for many physicians, indicating that weighting any one of the joint counts is not required. The physicians tolerated higher values on the patient global scale than on the evaluator global scale. This suggests that physicians tend to assume that patients rate their disease activity higher than their physicians. This is in line with the observed values in various patient cohorts [12,29] and clinical trials. Our data indicate that, in the evaluation of the acute-phase response, physicians assign similar clinical meanings to CRP values in mg/l and ESR values in mm that are about 10 units higher.

From a biological perspective, one would not expect that the similar relative values on different measures have similar implications. To further investigate how the clinical implications of the various measures relate to the biological consequences of the disease, we used 3-yr radiographic progression in an RA inception cohort as a marker of biological implication. We showed that the activity categories in all measures were discriminating different levels of radiographic progression. Moreover, prognostically the cut-points of the different measures had similar biological implications. Although there was a tendency for these associations to be stronger for some measures (and their cut-points) compared with others, there was no statistical evidence to support this notion. One should, however, consider that the categorization of activity status (>18 months in a particular state) might not have been sufficiently accurate to reliably identify a difference in this signal. Numerically, in this independently assessed early RA patient cohort, the steepest gradient between the different activity statuses in relation to radiographic progression was observed for EGA.

Thus, although the measures we surveyed are valued differently by physicians, the valuation of all measures had similar construct validity when related to structural damage. The differences in valuation may reflect a physician’s perception that each measure relates to somewhat different aspects of RA. The association of EGA with radiographic progression supports the frequent use of physicians’ judgements/decisions as a gold standard [1,16,17,19,20], by indicating that physicians integrate the different aspects of disease activity in a way that reflects prognostic aspects of the disease.

The criteria obtained for the different core-set measures cannot be transposed directly into clinical settings. They are hypothetical because of the way they were derived, since usually more than one measure is available to evaluate disease activity in real patients. However, each measure was anchored by the common definition of remission and low, moderate and high disease activity, and therefore allowed comparative assessment of these measures. In rare situations in which only one parameter is available or in which investigators do not want to integrate results, e.g. from subjective and objective measures, criteria based on our study might have direct practical merits. Also, it may be worthwhile investigating which combination of measures could be most useful in defining disease activity categories. However, the present survey was not designed to address such questions.

In conclusion, this study indicates that various measures are perceived differently by physicians, and differently determine the integrative evaluation of RA disease activity. These data give insights into the integrative process of activity evaluation, will be informative for future survey designs and studies using physician opinion as a gold standard for criterion validity, and will allow the activity mapping of values on different scales based on expert opinion.

### Acknowledgements

We would like to thank all participating rheumatologists for taking the time to fill in the questionnaire, for their thorough consideration of this survey, and for their general support of this endeavour.

The authors have declared no conflicts of interest.

### References


14. van Leeuwen MA, van der Heijde DM, van Rijswijk MH et al. Interrelationship of outcome measures and process variables in early...
rheumatoid arthritis. A comparison of radiologic damage, physical
disability, joint counts, and acute phase reactants. J Rheumatol
15. Dawes PT, Fowler PD, Clarke S, Fisher J, Lawton A, Shadforth MF.
Rheumatoid arthritis: treatment which controls the C-reactive protein
and erythrocyte sedimentation rate reduces radiological progression.
16. van der Heijde DM, van’t Hof M, van Riel PL, van de Putte LB.
Development of a disease activity score based on judgment in clinical
17. Prevoo ML, ‘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.
18. Soubrier M, Zerkak D, Dougados M. Should we revisit the definition
of higher disease activity state in rheumatoid arthritis (RA)?
rheumatoid arthritis: a preliminary definition. J Rheumatol 2005;32:
2016–24.
Remission and active disease in rheumatoid arthritis: defining criteria
21. Smolen JS, Breedveld FC, Schiff MH et al. A simplified disease
activity index for rheumatoid arthritis for use in clinical practice.
22. Arnett FC, Edworthy SM, Bloch DA et al. The American
Rheumatism Association 1987 revised criteria for the classification
23. Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid
arthritis and related conditions by standard reference films. Acta
24. Machold KP, Stamm TA, Eberl GJ et al. Very recent onset arthritis—
clinical, laboratory, and radiological findings during the first year of
Benefit of very early referral and very early therapy with disease-
modifying anti-rheumatic drugs in patients with early rheumatoid
26. Smolen JS, Aletaha D. Patients with rheumatoid arthritis in clinical
27. 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.
28. Aletaha D, Ward M. Duration of rheumatoid arthritis influences the
degree of functional improvement in clinical trials. Ann Rheum Dis
2005;65:227–33.
29. Antoni C, Maini RN, Grunke M et al. Cooperative on quality of
life in rheumatic diseases: results of a survey among 6000 patients