Minimal Clinically Important Difference in Radiological Progression of Joint Damage. A Definition Based on Patient Perspective

PACO M.J. WELSING, GEORGE F. BORM, and PIET L.C.M. van RIEL

ABSTRACT. ***Objective.*** To estimate a threshold for minimal clinically important radiological progression of joint damage using its longitudinal relation with functional disability in patients with rheumatoid arthritis (RA). To validate existing estimates of minimal clinically important difference (MCID) using this relation with functional disability.

Methods. We reanalyzed published data of 185 patients with early RA followed for a maximum of 9 years. Longitudinal regression (mixed models) was used, relating radiological damage (modified Sharp score) to functional disability (HAQ-DI), correcting for age (age at diagnosis and increasing disease duration), disease activity (DAS28), and demographic variables. Several shapes of the relation were investigated. Based on the observed relationship between radiological damage, functional disability, and the minimal clinically relevant increase in functional disability found in earlier studies, MCID for progression of joint damage was discussed. Existing estimates of MCID were evaluated for their influence on functional disability over the disease course.

Results. A longitudinal relation between the modified Sharp score and the HAQ-DI was found. Significant covariates were age, gender, and disease activity. The model indicated that the relation between the Sharp score and the HAQ-DI was dependent on the amount of damage (a threshold effect) and on patients' age. With lower age, no effect of joint damage on functional disability could be demonstrated and with higher age the effect of joint damage increased. With a typical patient from our cohort (age at diagnosis 55 yrs, some baseline damage, and an expected disease duration of 30 yrs), a (constant) progression of 6 points per year led to an increase of about 0.2 on the HAQ score, solely related to damage, over the disease course. This estimate of MCID was close to estimates based on expert opinion and equal or smaller than most estimates based on the smallest detectable difference.

Conclusions. The MCID, defined using longitudinal effect on functional disability, is dependent on age and (progression of) joint damage. However, with a typical patient population this MCID was similar to thresholds based on expert opinion, adding to the validity of these estimates. (J Rheumatol 2006; 33:501–7)

Key Indexing Terms:
RHEUMATOID ARTHRITIS JOINT DAMAGE
MINIMAL CLINICALLY IMPORTANT DIFFERENCE

Rheumatoid arthritis is a chronic inflammatory condition with considerable morbidity to the individual patient. Ultimately this disease can lead to destruction of the joints.

Joint destruction or radiological progression is considered one of the most important outcomes of RA, and is used to evaluate disease modifying antirheumatic drugs (DMARD) including biologicals[1,2].

There are several instruments for scoring radiological damage, using plain radiographs of hands and feet. The most widely used methods are the (modified) Sharp and the Larsen scores[3]. Both methods assess joint damage (and progression) on a continuous scale, but there is no established definition as to what constitutes clinically important progression of joint damage. This knowledge is important for interpretation of results of studies when progression of joint damage is the endpoint (at the individual patient level)[3].

Lassere, *et al* used the concept of measurement error to determine the smallest detectable difference (SDD) for the modified Sharp score as well as for the Larsen-Scott method, as a starting point for clinically relevant progression[4]. In the setting of early RA the SDD was 11 modified Sharp score units and 8 modified Larsen score units, if there was an equal distribution over the total spectrum of baseline damage and progression in the sample and the mean score of the same trained observers was always used. The SDD was 15.5 modified Sharp score units and 11 modified Larsen score units

From the Department of Rheumatology, and Department of Epidemiology and Biostatistics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

P.M.J. Welsing, MSc, Department of Rheumatology; G.F. Borm, PhD, Department of Epidemiology and Biostatistics; P.L.C.M. van Riel, MD, PhD, Professor of Rheumatology, Department of Rheumatology.

Address reprint requests to Dr. P.L.C.M. van Riel, Department of Rheumatology, Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands.
E-mail: P.vanRiel@reuma.umcn.nl

Accepted for publication October 28, 2005.

when there was an equal distribution of baseline damage and progression in the sample and the mean score of any 2 trained observers was used. Other SDD were also determined. Using another approach, pairs of radiographs, with 1-year intervals, from patients with early RA, were judged by an international expert panel of rheumatologists, for the presence or absence of a clinically important difference (defined as the amount of progression of joint damage that would make them change the second line therapy prescribed). A threshold value was chosen for a clinically relevant increase in radiological damage according to the rheumatologists, using receiving operator curves. Bruynesteyn, *et al* found that for the modified Sharp score this "clinically relevant" difference was 5 units and for the Larsen score 2 units[5].

However, these findings might not reflect the true minimal clinically relevant increase in progression of radiological damage; this should reflect the effect progression has on the patient, for instance the development (or increase) of functional disability, as this is of direct importance. A threshold for a clinically relevant difference in functional disability as measured by the Health Assessment Questionnaire-Disability Index (HAQ-DI) has been determined by asking patients to rate themselves relative to another patient and compare HAQ-DI scores. This minimal clinically relevant difference in HAQ-DI score was estimated to be between 0.2 and 0.3[6,7].

However, radiological damage is not the only factor that influences functional disability; other factors like disease activity and sociodemographic factors also have an (important) effect[8,9].

Our aim was to investigate the (longitudinal) relation between an increase in radiological damage and an increase in functional disability. This relation would then be used not only to find a threshold for clinically relevant progression of radiological damage, but also to validate existing estimates of the minimal clinically important difference (MCID) for radiological progression by evaluating their influence on functional disability over the course of the disease.

MATERIALS AND METHODS

We analyzed data from the Nijmegen inception cohort. Since 1985, all newly diagnosed patients with RA according to the 1987 American College of Rheumatology (ACR) criteria (disease duration < 1 yr) at the department of Rheumatology of the Radboud University Nijmegen Medical Centre have been included. Patients are assessed on several measures, including functional disability, radiological joint damage, and disease activity as measured by the HAQ-DI, the modified Sharp score, and the Disease Activity Score (DAS28), respectively[10]. Radiological data were collected at baseline and every 3 years with followup varying from 3 up to 9 years. A detailed description of this cohort has been reported[9].

As our study elaborates on results of 2 studies[4,5] investigating a minimal clinically relevant change in functional disability[6,7], the methods of these studies will be briefly reviewed.

Studies investigating a clinically relevant increase in functional disability. Two studies reported a threshold for clinically relevant change in functional disability on the HAQ-DI: one included 40 patients with RA; the other included 46 RA patients, with another 57 RA patients serving as a validation population. The patients had varying disease duration.

Thresholds for a clinically meaningful difference in functional disability as measured by the HAQ-DI were determined by conducting one-on-one conversations between RA patients, and comparing HAQ-DI scores. Participants rated whether their disability was the same, somewhat better, somewhat worse, much worse, or much better than their conversational partners. The minimal clinically relevant change was calculated as the difference in mean scores between the conversations where respondents rated themselves as somewhat better and conversations where the respondents rated themselves as about the same, or alternatively where respondents rated themselves as somewhat worse and about the same[6,7].

The longitudinal relation between functional disability and radiological damage. Followup data of 185 patients with early RA included in the open prospective study were used. In another article on these data the main conclusion was that in early disease cross-sectional functional disability was mainly influenced by disease activity, and that in established disease functional disability was also influenced by radiological damage[9]. This indicates that in individual patients the relation between an increase in joint damage and the development or increase in functional disability might not be linear. Furthermore, disease activity, increasing disease duration, and age might confound or modify this relation.

To investigate this relationship, mixed model regression analysis with a random intercept was used. A log transformation was applied to remove the skewness of the HAQ-DI. Statistical testing and inspection of the residuals were used to choose between the models and to evaluate the fit.

Initially, a sigmoid model was fitted with the logarithm of the HAQ-DI as dependent variable and the modified Sharp score, DAS28, age at onset of the disease, disease duration, gender, and rheumatoid factor as independent variables. This model showed poor fit, and a logarithmic transformation of the independent variables gave no improvement. Therefore, a response surface was used[11]. A quadratic response surface with Sharp score, age at diagnosis, disease duration, DAS28, gender, and rheumatoid factor showed adequate fit. The model could even be simplified by replacing 2 factors, age at diagnosis and disease duration, by a single factor: age. Inclusion of third order terms in the model did not lead to substantial improvement of fit.

Analyses were performed with SAS version 8.0 using the MIXED procedure with the Gaussian link function.

Threshold for clinically important radiological progression. Using the derived relationship between an increase in Sharp score and the HAQ-DI, corrected for important confounding (and modifying) factors, the minimal clinically relevant increase in radiological damage was investigated. An increase of 0.2 on the HAQ-DI scale was used as threshold for clinical importance. The derived relationship was also used to validate existing estimates of the MCID (or SDD) by evaluating the influence on functional disability over the disease course.

RESULTS

As described[9], 64% of the patients were women, the mean age at diagnosis was 55 years, and 78% of the patients were rheumatoid factor positive. Their median HAQ-DI score at baseline was 0.47, the mean DAS28-score was 4.4, and the median Sharp score was 11. The mean number of followup years (with data about radiological damage present) was 6.3. Functional capacity worsened (HAQ-DI increased) with disease duration by about 0.03 per year (min-max: HAQ-DI 0-3), after an initial improvement at study start.

At 9 years after study start the mean HAQ-DI had increased to 0.64 (median 0.63). Mean disease activity remained constant with disease duration, after an initial improvement at study start. Nine years after study start the mean DAS28 was about 3.8. Joint damage increased over the course of the disease; this increase seemed to be slower later

in the disease. The mean Sharp-score at 9 years after study start was 83.6 (median 83.8) and the mean yearly progression score was 8.1 Sharp points (± 7.6).

Results of the final model suggested that joint damage was related to the HAQ-DI. Female gender, higher age, and higher disease activity were also associated with a higher HAQ-DI. Age modified the relationship between Sharp score and HAQ-DI: the relationship became stronger with increasing age. Furthermore, the quadratic term for the modified Sharp in the model indicates that the influence of an increase in the Sharp score on functional disability is higher if the amount of damage already present is higher (Table 1). Figure 1A shows the relationship between HAQ-DI Sharp and age as estimated by the model. The figure indicates that with increasing age the HAQ-DI increases, that with lower ages no effect of increase in damage on functional disability can be seen, and with higher ages this effect is larger. Surprisingly, with low values of the Sharp score the functional disability seemed to decrease somewhat with increasing damage. This may be caused by an artefact of the modeling process, but models not allowing for such a decrease (e.g. sigmoid models) showed a very poor fit. It may also be a real effect caused by the process that leads to diagnosis of RA or regression to the mean. However the decrease is small and probably not of clinical relevance. Figure 1B shows the observed Sharp scores against the HAQ-DI scores and age adjusted for relevant covariates (the intercept, gender and DAS28).

Minimal clinically relevant radiological progression. Once we determined the (shape of the) relationship between progression of radiological damage and an increase in functional disability, we attempted to define the minimal clinically important radiological progression of joint damage (MCID). With a typical patient from our cohort (age 55 years at diagnosis and a baseline damage score of about 11) a disease course of about 30 years may be anticipated. In this case modeling shows that with a progression rate of about 6 Sharp points per year a clinically important increase in the HAQ-DI of 0.2 solely related to this progression of damage (not including the direct effect of age) is reached at the end of the disease course.

Table 1. Results of the linear mixed model.

Item	Coefficient	95% CI	p
Intercept	–2.967	–3.853, –2.081	< 0.0001
Age, yrs	0.022	–0.006, 0.039	0.0080
Female	0.384	0.710, –0.059	0.0220
DAS28	0.256	0.170, 0.343	< 0.0001
mSharp	–0.023	–0.024, –0.023	< 0.0001
mSharp squared	0.000053	0.000006, 0.000100	0.0286
Age*MSharp	0.00021	0.000006, 0.000414	0.0497

mSharp: modified Sharp score (min-max: 0–448); DAS28: disease activity score including 28-joint counts (min-max: 0–10); CI: confidence interval.

Figure 2 shows the influence of a yearly progression of 5 Sharp-points on the HAQ-DI (the MCID found by Bruynesteyn, *et al*[5] using expert opinion) for typical patients: patients aged 35, 55, and 75 years of age at diagnosis, respectively, with no baseline damage, a baseline damage score of 11, and a baseline damage score of 30, respectively. In Figure 2 the eventual influence over the disease course on the HAQ-DI according to the model is shown, assuming that patients reach 85 years and have a constant progression rate. It can be seen that with this progression rate the threshold for clinical relevance on the HAQ-DI is almost reached using a typical patient from our cohort (55 years of age at diagnosis and a baseline damage score of 11). With younger patients or patients with more baseline damage the eventual influence on the HAQ-DI is larger and with older patients and patients with less baseline damage this influence is less.

With higher age an increase in radiographic damage has a greater influence on the HAQ-DI (the Sharp score starts to have an influence at lower Sharp values); however, due to the shorter anticipated disease course the eventual influence is smaller and thus the minimal clinically important progression (MCID) becomes larger. With higher baseline damage the MCID becomes smaller. Most values of the SDD are equal to or larger than the estimate from our analyses of 6 modified Sharp points.

For a patient with an age at diagnosis of 35 years with a baseline damage score of 11 the MCID is 4; for a patient age 75 years at diagnosis the MCID is 20. For a patient age 55 years with no present damage the MCID is 7, and for the same age patient with a baseline damage score of 30 the MCID is 6 Sharp points per year.

DISCUSSION

We used an alternative approach to determine the minimal clinically important change in radiological joint damage. A threshold for a clinically relevant increase in radiological damage was determined, using its longitudinal effect on functional disability within individual patients, and as such using the patients' perspective. Previous work estimating the MCID for radiological joint damage used either measurement error (the statistical approach) or the opinion of rheumatologists (the opinion based approach). Our method can be regarded a data driven approach[3]. It should be noted that all these thresholds relate to an average and not to an individual patient and thus will not play a major role in day to day care of patients. However, an average threshold for clinically important progression might facilitate the interpretation of radiological outcomes of clinical studies, and make it possible to perform a responder analysis in clinical trials using radiological data.

Our results reveal some interesting points regarding the clinically important difference in radiological damage as defined by the influence of a patient's functional disability. The influence of radiological damage on functional disability might have a threshold; i.e., only after the presence of a cer-

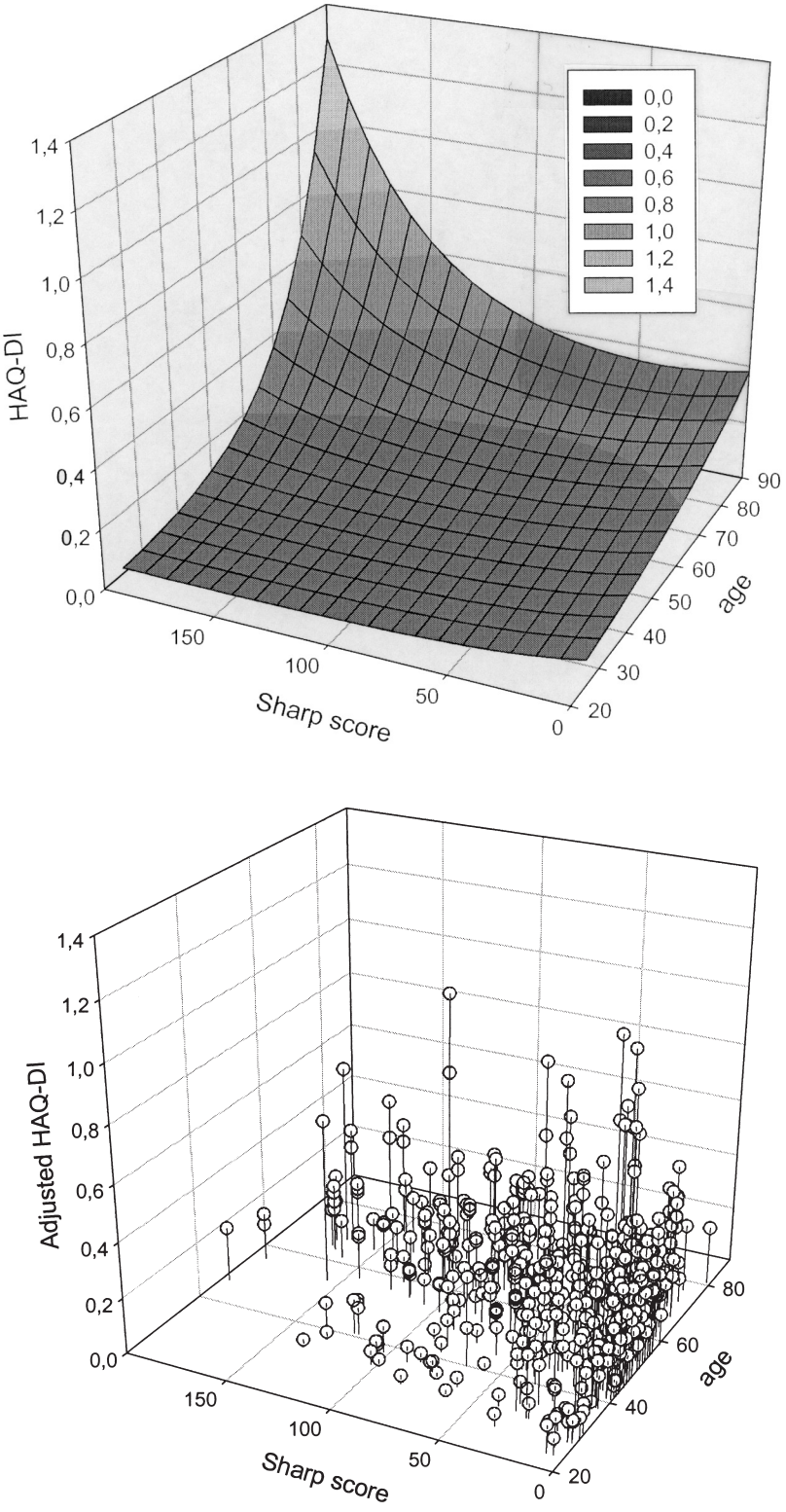

Figure 1. A. The relationship between joint damage, functional disability, and age as estimated by our model: With increasing age, the HAQ-DI increases and with lower ages, no effect of increased damage on functional disability can be seen. B. Observed Sharp scores plotted against HAQ-DI scores, after age adjustment for relevant covariates (intercept, gender, and DAS28).

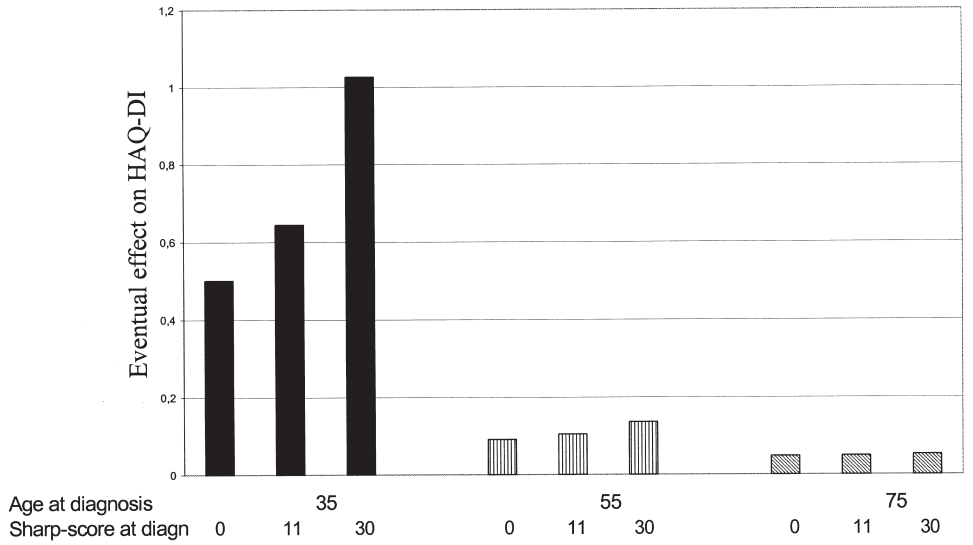

Figure 2. Eventual influence of functional disability for typical patients age 35, 55, and 75 years. The linear mixed model assumes all patients reach 85 years, and that they have a constant disease progression rate of 5 Sharp points per year. Representative patients for each age with varying baseline joint damage scores are shown.

tain amount of radiological damage, an increase in radiological damage has an influence on functional disability in a patient. Furthermore the influence of radiological damage on functional disability within individual patients seems also dependent on the age of the patient; with higher age the influence is greater (radiographic damage starts to have an effect with lower Sharp scores). Although the influence of increasing joint damage is smaller in younger patients, due to their longer anticipated disease duration the eventual influence of the progression might be greater. Therefore the MCID might be smaller, stressing the importance of slowing progression of joint damage in younger patients.

This finding might indicate that one single threshold for the clinically relevant increase in radiological damage may not be appropriate, but that this clinically relevant increase may be dependent on characteristics of both the patients and the disease. This was also the case for the MCID based on expert opinion: in the setting of early RA and with high disease activity, the MCID was smaller[12] and radiologists regarded larger values of radiological progression as clinically relevant (defined as recording "substantial progression" in their report) in contrast to rheumatologists[13]. However, when using an "average" patient from our cohort the resulting MCID was remarkably close to estimates based on the expert opinion of a panel of rheumatologists, adding to the validity of these estimates. This suggests that experts in the field of rheumatology are capable of interpreting the clinical relevance of joint damage as shown on radiographs. Usually estimates of the MCID based on the smallest detectable difference (SDD) are greater than or equal to the MCID based on expert opinion or on the relation with functional disability. The value of the SDD is dependent on the distribution of scores in the sample; the number of raters; the number of scores, i.e., the mean of 2 or more readers versus only one score; and the status or change scores used to calculate the SDD. The value is also different if one wants to generalize to any (2 or more) raters. Furthermore, it was argued by Bruynesteyn, *et al* that calculating an SDD to assess the smallest change in scores that can be deemed a "real" change is inappropriate if the films of one particular patient are scored side by side (as is usually done)[14]. In this case the change in scores is not based on 2 independently obtained scores and the measurement errors of scores of films at time points one and 2 are not independent. Therefore the smallest detectable change (SDC) should be calculated as a threshold for progression. This SDC is based on the standard error of measurement of the change score or the standard deviation of the difference between change scores of 2 reading sessions as opposed to using the standard deviation of the status scores (or of change scores) in the usual method of calculating the SDD. The SDC is smaller than the SDD and might be smaller than the MCID but this is also dependent on the raters, using means, and if one wants to generalize[14].

Although the measures used to determine functional disability and radiological damage in our study are internationally accepted, they deserve further comment. The HAQ-DI might not be sensitive enough to measure subtle changes in functional disability, especially not at the extremes of the scale[15]. This might partly explain the threshold effect for radiological damage. In our patient cohort large changes in Sharp score (and also DAS28) are necessary for a relevant change in the HAQ-DI. The HAQ-DI increases very slowly by about 0.02-0.03 per year[9]; at the cohort level, a clinically relevant increase in HAQ-DI is only reached after 7 to 10 years (partly due to an increase in age and partly due to an increase in damage). The modified Sharp method measures erosions and

joint space narrowing in joints of hands and feet. It does not take into account how many joints are involved, so a certain score may be due to much damage in a few joints or little damage in many joints. This might not have the same effect on disability and thus not have the same clinical relevance. Also this method only takes into account joints of the hands and feet; damage in other joints might have a different influence on disability. However, it has been found that progression in these joints is very similar to the progression in other joints[16]. A ceiling effect of the radiological scoring method in individual joints has also been described[17]. Therefore damage might be increasing, although the Sharp score remains constant or only increases slowly, but still influences functional disability.

As described, Sharp scores were assumed to only increase or stay the same and not to improve. Our cohort was treated prior to 1998 before the introduction of biologicals; therefore healing phenomena were probably not often present and could not influence our results.

For patients who have had joint replacement surgery, that particular joint is impossible to score and (more importantly) functional disability decreases (better function due to the surgery) even though damage has increased or is high. Furthermore radiological scoring methods can have a very large inter-observer variation[16,18]. Scoring photos more frequently (i.e., photos every 6 months versus every 3 years) increases the progression rate, since Sharp scores, when scored in the original way can only increase or be stable and not decrease. Therefore the calculation of an MCID for damage is (partly) study or observer specific.

In studies determining a clinically relevant increase in HAQ-DI scores it was shown that patients view clinically important differences in an asymmetric manner, and that a larger difference in HAQ-DI scores had to be present for patients to rate themselves as worse versus rating themselves as better[7]. Furthermore, in less disabled patients we observed a lower threshold for a clinically relevant increase in functional disability[6]. The MCID in HAQ-DI were determined by conversations between patients, rating themselves relative to other patients, after which HAQ-DI values were compared. This difference relates to inter-patient differences, and is not necessarily the same as the minimal clinical relevant change within patients, which might be smaller. In this study a conservative value for a clinically relevant difference in functional disability of 0.2 was used. These issues complicate the calculation of a clinically relevant increase in radiological damage.

It is possible that not all factors influencing functional disability were investigated in the regression models, for instance, psychological status; however, we do not believe this would confound or modify the longitudinal relationship between radiological damage and functional disability.

In conclusion we described the relationship between radiological damage and functional disability in individual patients. The magnitude of this relationship was dependent on characteristics like age, disease duration, and present damage, complicating the calculation of a threshold for clinically important radiological progression. This definition of MCID was further complicated by the properties of the radiograph scoring method (e.g., included joints, inter-observer variation, and scoring frequency) and by the definition of a clinically relevant increase in HAQ-DI score. The MCID calculated for an "average" patient from our cohort was 6 Sharp units, which was remarkably close to the MCID using expert opinion, adding to the validity of these estimates. The MCID based on the SDD is usually equal to or larger than the MCID based on either expert opinion or the influence on the HAQ-DI. It might be interesting to replicate these findings in other patient populations and investigate the role of the issues raised in our study.

REFERENCES

1. Smolen JS, Kalden JR, Scott DL, et al. Efficacy and safety of leflunomide compared with placebo and sulphasalazine in active rheumatoid arthritis: a double-blind, randomized, multicentre trial. Lancet 1999;353:259-66.
2. Lipsky PE, van der Heijde DMFM, St. Clair EW, et al. Infliximab and methotrexate in the treatment of rheumatoid arthritis. N Engl J Med 2000;343:1594-602.
3. van der Heijde D, Lassere M, Edmonds J, Kirwan J, Strand V, Boers M. Minimal clinically important difference in plain films in RA: group discussions, conclusions, and recommendations. OMERACT Imaging Task Force. J Rheumatol 2001;28:914-7.
4. Lassere M, Boers M, van der Heijde, Boonen A, Edmonds J, Saudan A, Verhoeven AC. Smallest Detectable Difference in radiological progression. J Rheumatol 1999;26:731-9.
5. Bruynesteyn K, Heijde van der D, Boers M, et al. Minimum Clinically Important difference in radiological progression of joint damage over 1 year in rheumatoid arthritis: preliminary results of a validation study with clinical experts. J Rheumatol 2001;28:904-10.
6. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. Arch Intern Med 1993;153:1337-42.
7. Wells GA, Tugwell P, Kraag GR, Baker PRA, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. J Rheumatol 1993;20:557-60.
8. Escalante A, del Rincon I. How much disability in rheumatoid arthritis is explained by rheumatoid arthritis? Arthritis Rheum 1999;42:1712-21.
9. Welsing PMJ, van Gestel AM, Swinkels HL, Kiemeney LALM, van Riel PLCM. The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. Arthritis Rheum 2001;44:2009-17.
10. van der Heijde DM, van Riel PL, Nuver-Zwart IH, Gribnau FW, van de Putte LB. Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. Lancet 1989;1:1036-8.
11. Mead R, Pike DJ. A review of response surface methodology from a biometric viewpoint. Biometrics 1975;31:803-51.
12. Bruynesteyn K, van der Heijde D, Bores M, et al. Determination of the minimal clinically important difference in rheumatoid arthritis joint damage of the Sharp/van der Heijde and Larsen/Scott Scoring methods by clinical experts and comparison with the smallest detectable difference. Arthritis Rheum 2002;46:913-20.
13. Bruynesteyn K, van der Linden S, Landewe R, Gubler F, Weijers R, van der Heijde D. Progression of rheumatoid arthritis on plain

radiographs judged differently by expert radiologists and rheumatologists. J Rheumatol 2004;31:1088-94.
14. Bruynesteyn K, Boers M, Kostense P, van der Linden SJ, van der Heijde D. Deciding on progression of joint damage in paired films individual patients: smallest detectable difference or change? Ann Rheum Dis 2005;64:179-82.
15. Gardiner PV, Sykes HR, Hassey GA, Walker DJ. An evaluation of the health assessment questionnaire in long-term longitudinal follow-up of disability in rheumatoid arthritis. Br J Rheumatol 1993;32:724-8.
16. Scott DL, Coulton BL, Popert AJ. Long term progression of joint damage in rheumatoid arthritis. Ann Rheum Dis 1986;45:373-8.
17. Kuper IH, van Leeuwen MA, van Riel PL, et al. Influence of a ceiling effect on the assessment of radiographic progression in rheumatoid arthritis during the first 6 years of disease. J Rheumatol 1999;26:268-76.
18. Swinkels HL, Laan RFJM, van 't Hof MA, van der Heijde DMFM, de Vries N, van Riel PLCM. Modified Sharp method: factors influencing reproducibility and validity. Semin Arthritis Rheum 2001;31:176-90.