Editorial

Measuring Patient Health Status in Rheumatoid Arthritis — What Is a Minimal Clinically Important Difference?

Rheumatoid arthritis (RA) is characterized by joint destruction, deformities, and disability[1]. Clinicians often use plain radiographs to determine the degree of joint destruction and to monitor the progression of the disease. In this issue of *The Journal*, Bruynesteyn and colleagues present findings about different clinicians' judgment of what is a minimal clinically important difference (MCID) in the progression of RA using plain radiographs of the hand and foot at one year intervals[2].

An expert panel of 5 international experienced rheumatologists was asked to define MCID as "the amount of progression of joint damage that would make them change the second-line therapy prescribed." When musculoskeletal consultant radiologists recorded "substantial progression" in their reports, this was defined as MCID. The Sharp/van der Heijde scores of the plain radiographs and receiver operating characteristic analyses provided a common quantification of what the 2 professions judged to be MCID. The authors concluded that the radiologists appeared to be more reserved than the rheumatologists in judging MCID for patients with RA. Despite finding a difference in judgment of the 2 professions and noting that radiological joint damage is only an intermediate outcome measure, Bruynesteyn and colleagues nevertheless emphasize the need to assess this in trials in addition to patient reported outcomes.

This editorial examines in detail the rationale for assessing MCID in relation to patient-assessed health instruments, i.e., the smallest difference that is important to patients.

WHAT IS MCID?

A minimal clinically important difference has been defined as "the smallest important difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side-effects and excessive cost, a change in the patient's management."[3] MCID facilitate the interpretation of score changes on outcome measures and are useful for informing sample size calculations in evaluative studies including clinical trials.

MCID DETERMINED BY CLINICAL JUDGMENT

As suggested by Bruynesteyn and colleagues, clinicians' repeated experience of viewing plain radiographs allows them to interpret what is substantial radiological progression in joint damage that may lead to a change in therapy. It is useful to discuss this method of estimating MCID in the context of a framework developed by Fineberg, *et al* to evaluate diagnostic technologies following debate about the adoption of computed tomography in the 1970s[4]. The proposed framework measures the effects of diagnostic technologies at 4 separate levels[5]; these were subsequently extended by the Institute of Medicine to 5[6]. Figure 1 shows the framework as applied to the assessment of RA using plain radiographs.

The "diagnostic accuracy" level is concerned with whether plain radiographs allow clinicians to accurately assess the presence and severity of RA. However, image interpretation is considered to be one of the weakest areas of clinical radiology and a source of substantial variation[7]. At this level of the framework it is therefore necessary to assess intra- and inter-observer reliability for consistency in clinician decision-making. A recent study showed good concordance at 74% between 3 experienced consultant radiologists when reporting on skeletal plain radiographs, compared with only 61% and 51% concordance when reporting radiographs for the chest and abdomen, respectively[8]. Indeed, Bruynesteyn and colleagues showed good intra- and inter-observer reliability in clinician decision-making when reporting on skeletal radiographs. This provides evidence that skeletal plain radiographs are a reliable method for monitoring changes in joint damage. Bruynesteyn and colleagues also used experienced clinicians to ensure that valid judgments were made about the progression of RA. They also appropriately addressed the

See Progression of rheumatoid arthritis on plain radiographs judged differently by expert radiologists and rheumatologists, *page 1088*

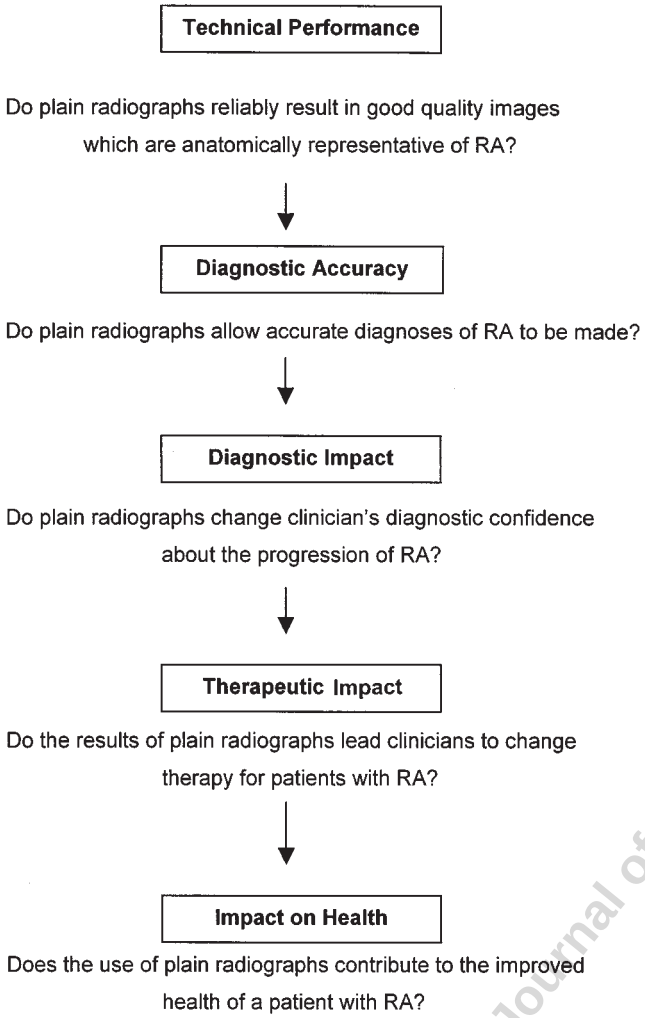

Figure 1. Evaluative hierarchy as applied to the assessment of rheumatoid arthritis using plain radiographs.

issue of independence in decision-making, which can be a source of bias, as well as the effect of availability of clinical data[9]. At the "therapeutic impact" level they defined MCID as whether a change in radiological progression of disease would lead to a change in therapy. The assumption being that a change in management is important for improving patients' health as should be addressed at the "impact on health" level of Fineberg's hierarchy.

Despite an appropriate application of this method of determining MCID there are substantial limitations to this approach. First, the reliability and validity of decision-making by clinicians when judging changes in the progression of joint damage is dependent on their experience. Moreover, variation was found in different professions' judgment of what was a MCID. The clinicians included in a study could therefore dramatically influence how MCID is determined. Second, several experienced clinicians reporting on pairs of radiographs can require substantial resources. In Bruynesteyn and colleagues' study it was only feasible to include 46 patients. Subsequently this may result in limited generalizability and lack of precision in findings. Third, the global assessment of clinician judgment about how a change in therapy should improve health outcomes does not necessarily reflect what the progression of RA means to the patient's functioning and well being.

MCID DETERMINED BY PATIENT JUDGMENT

It is now widely accepted that randomized trials and similar forms of evaluative study should incorporate the patient's views about outcome. This has led to considerable growth in the number of instruments, questionnaires, and rating scales that are designed for measuring health status and health outcomes from the perspective of the patient[10]. However, much of this work has focused on the development and testing of instruments for measurement properties of reliability and validity; relatively less attention has been given to the interpretation of instrument scores, including the MCID[11].

There are 2 broad approaches to assessing the MCID for score changes produced by patient-assessed health instruments: anchor-based and distribution-based[12,13]. Distribution-based approaches rely on statistical criteria including significance testing, sample variation, and measurement precision[14]. The most commonly applied distribution-based approach is the effect size statistic, which relates changes in instrument scores to baseline variation in the sample[11]. Several authors suggest that statistical measures are insufficient for assessing instruments; they recommend that the views of patients about the importance of the change should be included[14-16]. Anchor-based approaches assess the relationship between changes in instrument scores and an external variable[13]. This includes health transition items or global judgments of change that have been used to estimate the MCID[3,17]. Several studies have used a 15-point rating scale ranging from –7 to 7, where –7 is a very great deal worse and 7 is a very great deal better; the MCID has been judged to be in the range of 1–3[18].

A reliance on the patient in the determination of the MCID is a key advantage of health transition questions. However, few health transition ratings that have been used for assessing MCID have been evaluated for reliability and validity[19]. Retrospective judgments are also subject to recall bias, and transition ratings have been found to be unduly influenced by current health states[20]. Moreover, MCID should not be considered a fixed property and may vary across groups of patients[14,15].

FUTURE ASSESSMENT OF MCID IN PATIENTS WITH RA

Numerous outcome measures are available for assessment of patients with RA, including clinical and patient-assessed approaches. Standardization is dependent on structured

reviews of existing instruments and expert recommendations such as those made by the American College of Rheumatology and the international committee, Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT). However, the interpretation of instrument scores is a neglected area of research[21]. The assessment of important changes in health will inform sample size calculations in evaluative studies and enhance clinical understanding of the meaning of score changes produced by outcome measures. Future studies involving patient-assessed health instruments should assess the MCID through the concurrent use of health transition questions. However, further research should also address the reliability and validity of such questions.

STEPHEN BREALEY, BSc,
Research Fellow,
York Trials Unit,
Department of Health Sciences,
2nd Floor, Area 4,
Seebohm Rowntree Building,
University of York,
Heslington, York, YO10 5DD, United Kingdom

Address reprint requests to S. Brealey. E-mail: sb143@york.ac.uk

REFERENCES

1. Molenaar ETH, Voskuyl AE, Dinant HJ, Bezemer D, Boers M, Dijkmans BAC. Progression of radiologic damage in patients with rheumatoid arthritis in clinical remission. Arthritis Rheum 2004;50:36-42.
2. Bruynesteyn K, van der Linden SM, Landewe RBM, Gubler F, Weijers R, van der Heijde DMFM. Progression of rheumatoid arthritis on plain radiographs judged differently by expert radiologists and rheumatologists. J Rheumatol 2004;31:1088-94.
3. Jaeschke R, Singer J, Guyatt GH. Measurements of health status: ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-15.
4. Brealey S. Measuring the effects of image interpretation: an evaluative framework. Clin Radiol 2001;56:341-7.
5. Fineberg HV, Bauman R, Sosman M. Computerized cranial tomography: Effect on diagnostic and therapeutic plans. JAMA 1977;238:224-7.
6. Institute of Medicine. Policy statement: computed tomographic scanning. Washington, DC: National Academy of Sciences; 1977.
7. Robinson PJA. Radiology's Achilles' heel: error and variation in the interpretation of the Rontgen image. Br J Radiol 1997;70:1085-98.
8. Robinson PJA, Wilson D, Coral D, Murphy A, Verow P. Variation between experienced observers in the interpretation of accident and emergency radiographs. Br J Radiol 1999;72:323-30.
9. Brealey S, Scally AJ. Bias in plain film reading performance studies. Br J Radiol 2001;74:307-16.
10. Garratt AM, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ 2002;324:1417-9.
11. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2:i-iv:1-74.
12. Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. J Eval Clin Pract 2000;6:39-49.
13. Norman GR, Sridhar FG, Guyatt GH, Walter SD. Relation of distribution and anchor-based approaches in interpretation of changes in health-related quality of life. Med Care 2001; 39:1039-47.
14. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56:395-407.
15. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res 2003;12:349-62.
16. Liang MH, Lew RA, Stucki G, Fortin PR, Daltroy L. Measuring clinically important changes with patient-oriented questionnaires. Med Care 2002;40 Suppl:II45-51.
17. Juniper EF, Price DB, Stampone PA, Creemers JP, Mol SJ, Fireman P. Clinically important improvements in asthma-specific quality of life, but no difference in conventional clinical indexes in patients changed from conventional beclomethasone dipropionate to approximately half the dose of extra fine beclomethasone dipropionate. Chest 2002;121:1824-32.
18. Wells G, Beaton D, Shea B, et al. Minimal clinically important differences: review of methods. J Rheumatol 2001;28:406-12.
19. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. Med Care 2000;38 Suppl:II84-90.
20. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. J Clin Epidemiol 2002;55:900-8.
21. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum 2000;43:1478-87.