Minimum Clinically Important Difference: The Crock of Gold at the End of the Rainbow?

JOHN R. KIRWAN

ABSTRACT. The minimum clinically important difference (MCID), like the crock of gold at the end of the rainbow, is attractive but unattainable. Empirical data on how rheumatologists make clinical decisions show a wide variety of approaches and lack of agreement in decision making. Clinical importance needs to consider the magnitude of both the benefits and adverse events. A proposal for future attempts to define MCID could explore links between short term changes in outcomes to improvement in disability outcome many years later. Defining response to treatment could be explored using different approaches and involving patients and other professional groups. (J Rheumatol 2001;28:439–44)

Key Indexing Terms:
OUTCOME AND PROCESS ASSESSMENT RANDOMIZED CLINICAL TRIALS
MINIMAL CLINICALLY IMPORTANT DIFFERENCE

There are a number of empirical and theoretical reasons to suppose that the definition of "minimal clinically important difference" (MCID) will prove as elusive as the crock of gold that legend tells us waits at the end of a rainbow.

The empirical evidence rests on analysis of the clinical judgments of many rheumatologists in several countries, which has shown that different rheumatologists have very different decision making patterns. The theoretical arguments relate to the range of potential definitions, their mutual contradiction, and the realization that when defining MCID in relation to response to treatment, any definition is dependent on factors other than direct therapeutic effect.

An alternative approach might be to define outcomes of clinical intervention in terms that are more easily understood and interpreted by practicing clinicians. Such descriptions could usefully employ a consensus view of the extent of patient response to intervention that is based on perceptions of change seen as important by a defined proportion of knowledgeable clinicians (or patients) in the particular circumstances in which the treatment has been tested.

EMPIRICAL EVIDENCE

The empirical evidence rests on analyses of the clinical judgments of many rheumatologists in several countries[1,2]. Clinical judgment analysis requires rheumatologists to make decisions about the severity of arthritis, or about changes in severity, in many patients. (Patient data may be taken directly from the clinical encounter or be conveniently provided in the form of "paper patients"[3].) A multiple regression equation is then calculated relating the clinician's decisions to the data available about the patients. A judgment policy model is derived by expressing the relative weight of each variable as a percentage contribution to the equation, and by expressing the ability of the equation to adequately capture the decision making process as the square of the multiple correlation coefficient (R^2)[1]. This has shown[4] that different rheumatologists have very different decision making patterns, both in relation to judgment of current disease activity in rheumatoid arthritis (RA)[4,5] and in relation to changes in response to therapy in patients treated with intramuscular gold[6,7].

It must be appreciated that rheumatologists are not good at describing how they come to their decisions[8-10]. This is illustrated by the simple example in Figure 1[4]. Two rheumatologists described the weight or importance they gave to 5 aspects of "current disease activity" (top charts), while an analysis of the decisions they made in practice for actual patients showed that these decisions related to clinical status in very different ways (bottom charts). There is long-standing and extensive evidence that inaccurately described decision making processes may lead to lower levels of agreement in decision making[11], and evidence within rheumatology that such agreement can be facilitated by the provision of accurate descriptions derived from regression analysis[8,10].

Rheumatologists' judgments of "change in response to treatment" show very great variations. The assessments made by 38 Australian rheumatologists of 5 patients are illustrated in Figure 2. In mitigation, it might be considered that the visual analog scale (VAS) was used differently by different rheumatologists, so that there may be less disagreement when the threshold for a "clinically important change" is used as an alternative. Unfortunately this is not sustained in practice. Figure 3 shows the proportion of 48 British rheumatologists identifying clinically important

From the Rheumatology Unit, University of Bristol Division of Medicine, Bristol Royal Infirmary, Bristol, UK.

J.R. Kirwan, MD, Consultant and Reader in Rheumatology.

Address reprint requests to Dr. J.R. Kirwan, Rheumatology Unit, Bristol Royal Infirmary, Bristol BS2 8HW, UK.
E-mail: John.Kirwan@Bristol.ac.uk

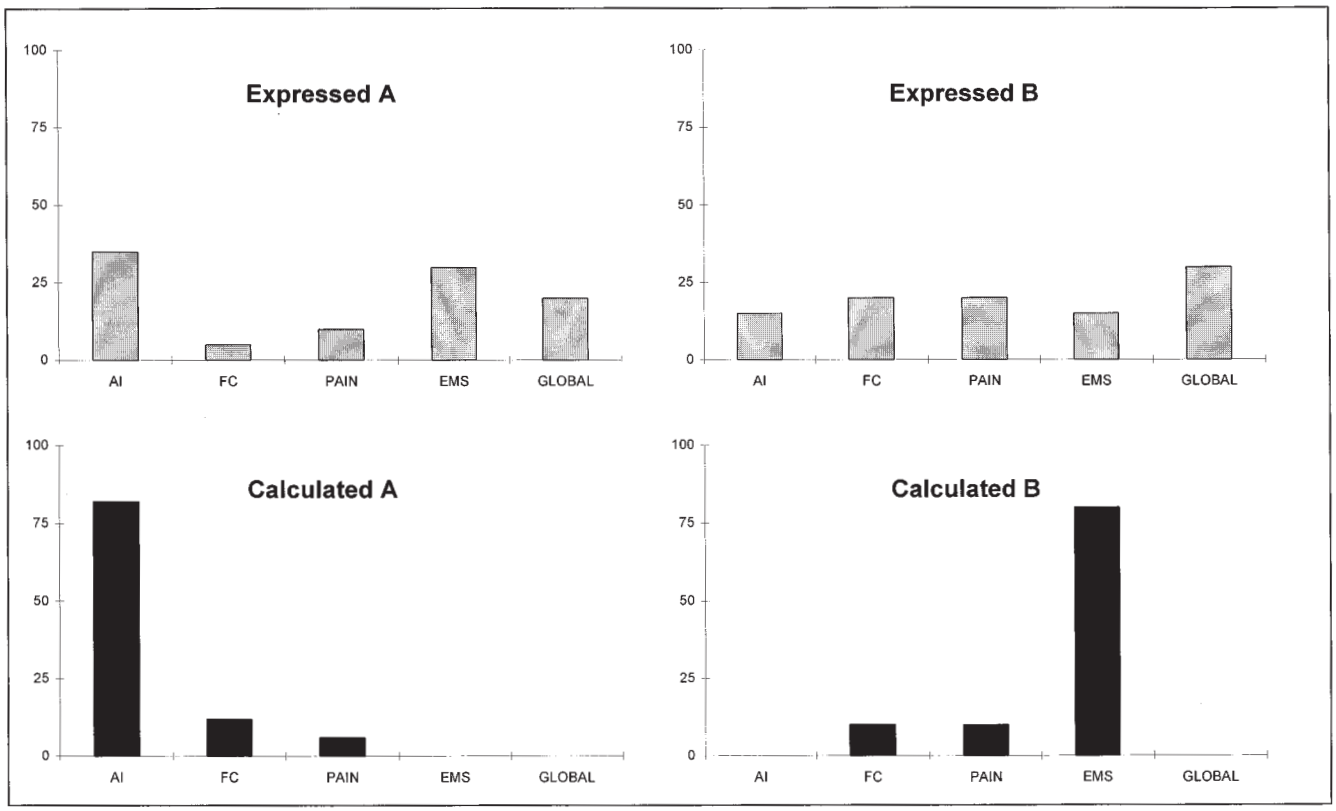

Figure 1. Weight or importance 2 rheumatologists gave to 5 aspects of "current disease activity" (top charts), and analysis of the decisions they made in practice for actual patients (bottom charts). AI: articular index, FC: functional capacity, EMS: early morning stiffness, Global: patient's opinion of overall severity.

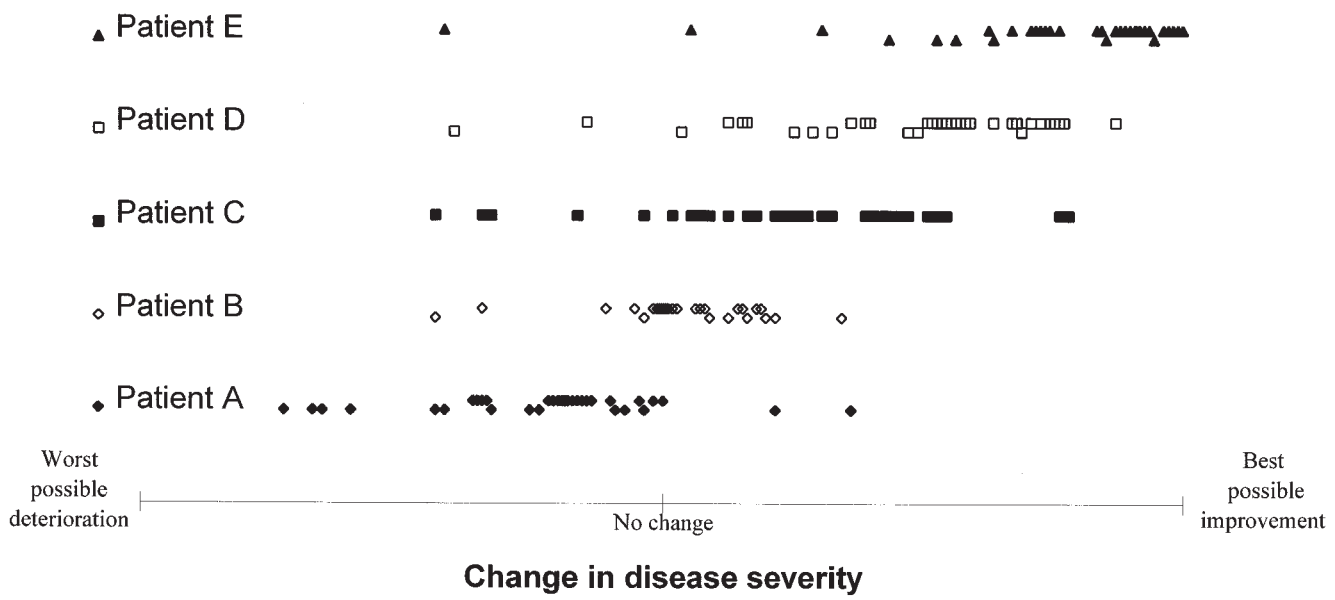

Figure 2. The opinion of 38 rheumatologists on the change in disease status of 5 patients.

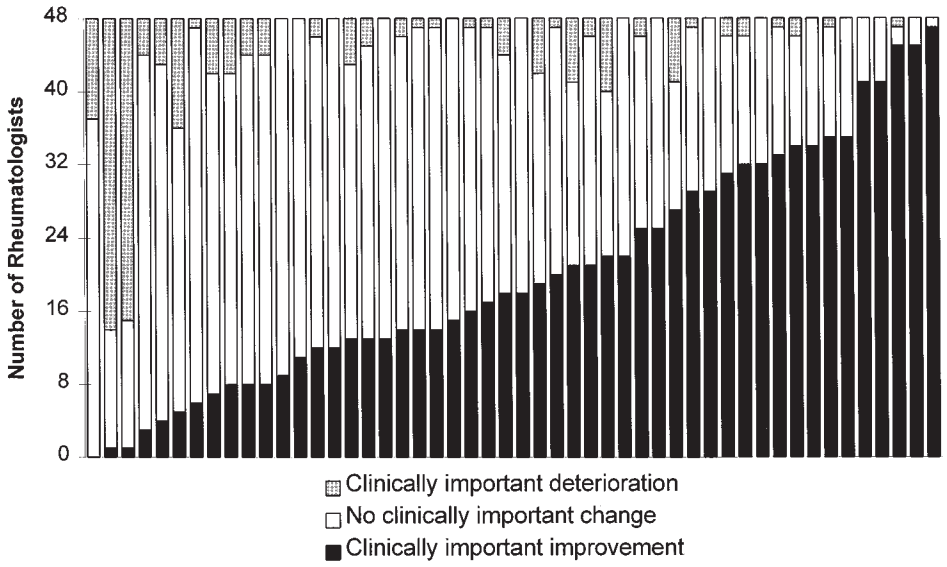

Figure 3. Clinically important change in 50 patients as judged by 48 rheumatologists.

change in the same 50 patients[6]. There are few patients for whom there is clearcut agreement that they have had a clinically important improvement or deterioration.

Another approach is to relate the number of doctors agreeing about whether a patient's condition has changed in a clinically important way to the average VAS score given for the change seen (Figure 4). There are extensive overlaps, and even at the extremes there are few patients about whom all rheumatologists agree. However, perhaps not all need agree — perhaps only a large proportion of agreement will be needed to define patients who have changed in a clinically important way. It is possible to relate the proportion of doctors required to agree about a patient's change to the proportion of patients for whom agreement can be reached, as illustrated in Figure 5. This shows that if 80% of doctors are required to agree, then only 18% of patients qualify for agreement. If 70% of doctors are required to agree, then 46% of patients qualify for agreement. If the requirement for the proportion of doctors required to agree is relaxed to 51% (a bare majority), then agreement can be reached for 84% of patients. However, the slope of the curve is steep and the decision on how many doctors must agree is arbitrary.

The question arises as to whether these variations will make any difference to clinical trial design. In fact, they have a profound effect on the number of patients required to ensure that a given study will detect improvements that are considered clinically important by a particular rheumatologist (that is, MCID). Calculations have been undertaken to assess the consequences of this consideration[11]. In order to make conservative estimates of the variation in trial design, only those rheumatologists with highly consistent judgments (repeated judgments on the same data, $r > 0.8$) which could be well modeled ($R^2 > 0.8$) were used in the calculations. There were 14 such rheumatologists, and the trial sizes required to detect 50% of patients responding to therapy are as shown in Table 1. They range from 206 to 4134 depending on which rheumatologist will be making the outcome assessments. One interesting conclusion from this study was the recognition that for many rheumatologists no single variable could change sufficiently to make a clinically important difference by itself, but rather combined modest changes in a number of variables may be considered useful even though each alone might be insignificant. This finding supports previous suggestions for defining a clinical response in terms of several variables[12,13].

A further challenge to those designing and interpreting clinical trials is the way clinicians may treat the outcome of a study containing many patients in the same way as data from a single individual. When data from the patients reported above were presented in a format that resembled clinical trial results, the judgments of 14 rheumatologists were essentially the same as when the same data had been presented as originating from individual patients[14]. There was no suggestion that a treatment effect observed in a clinical trial could be smaller to achieve the same degree of clinical importance as the same effect in one patient. This runs counter to the expectations of many statisticians.

An interesting approach to combining the decisions of a group of rheumatologists was the development of the disease activity score (DAS)[2]. The DAS was based on clinical judgment analysis applied to decisions made by rheumatologists in The Netherlands. The judgments were decisions about changing or starting second line therapy for outpatients with RA and the DAS was the multiple linear regression equation that related that "collective" decision

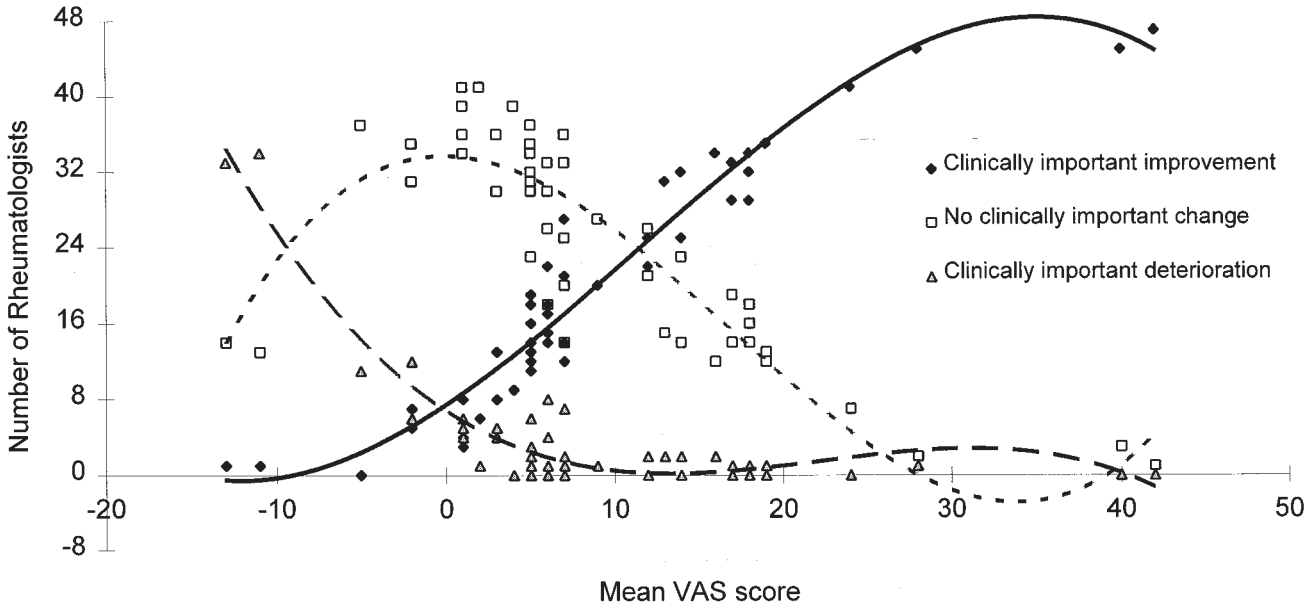

Figure 4. Number of rheumatologists identifying clinically important changes in relation to mean VAS score for change in response to therapy.

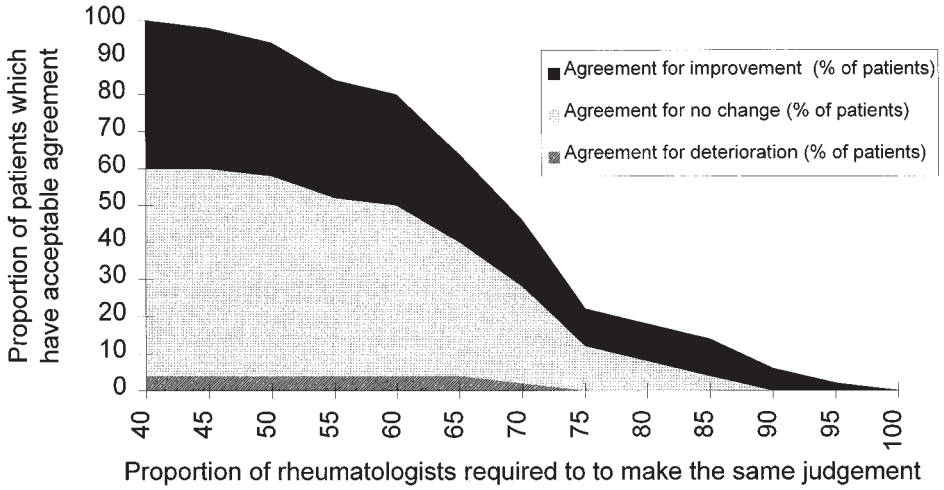

Figure 5. ROC curve. Proportion of rheumatologists required to make the same judgment of "clinically important change" against the proportion of patients on which rheumatologists can be said to agree.

making process to the clinical status of the patients. The decisions were manifestly clinically important — they did, after all, change treatment options. This process should therefore mean that the cutoff level at which the DAS relates to a definite decision to change therapy should give at least an upper limit on the definition of MCID.

The development of the DAS has been an important contribution to defining clinically important change, but there are several reasons for caution in its use. For example, the doctors were all from closely linked clinical settings, and the amalgamation procedure for the decisions masks underlying variability between the doctors. Furthermore, there is the issue of the context in which the judgments were made, which leads to a number of theoretical considerations.

THEORETICAL ARGUMENTS

The decisions on which the DAS was constructed were made against a background of understanding of the potential benefits of second line therapies and their potential toxicity. That is, each decision was about whether to change a specific treatment with a specific balance of benefits and risks. If the study had been about nonsteroidal antiinflammatory drugs or immunosuppression it is likely that the threshold for changing treatment would have been much

Table 1. Trial size required to demonstrate superiority of treatment in reaching a clinically important improvement for 14 doctors who observed different rates of improvement in the same group of 50 patients (from reference 11).

Doctor	Observed improvement (% of patients) taking gold	Trial size* to detect 50% improvement
14	50	206
6	54	234
9	56	250
4	46	252
7	42	308
3	68	390
15	36	414
8	34	458
12	34	458
13	66	460
11	76	560
5	28	626
10	88	1244
2	6	4134

*alpha = 0.05; beta = 0.05; n1 = n2.

lower and higher, respectively. Thus a MCID cannot be defined in isolation from the context of the consequences of its interpretation or use. There are always risks of side effects and always costs (or savings). If there were no side effects and no costs, then any improvement in symptoms would be worth having — no matter how small. If there were high risks and high costs, some people may not even be prepared to accept the chance of perfect health as being a sufficiently large clinical improvement to try out the treatment[15]. Thus, in the latter context, there is no clinically important difference, while if it is free of cost and adverse effects, even the minimal detectable difference will be clinically important. It is not unreasonable to suppose that MCID will be different for decisions about (say) selective cyclooxygenase-2 inhibitors and the use of anti-tumor necrosis factor therapy.

If an attempt is made to overtly incorporate both risks and benefits in considering clinically important change, then the magnitude of the risk multiplied by its likelihood must be less than the magnitude of the benefit multiplied by its likelihood. To calculate the likelihood of benefits (and risks) it is necessary to know their distribution within the treated population, rather than simply the average benefit. It is possible that for some treatments, a comparison of risks and benefits may favor benefits over a wide range of assumptions about MCID, and this could be tested using a sensitivity analysis. If the benefits of treatment are insensitive to large variations in the definition of MCID, the difficulties of defining or agreeing on the definition melt away for this particular treatment outcome in this particular study. It is possible that this approach would work for published studies of nonsteroidal antiinflammatory drugs versus placebo, or glucocorticoids versus placebo. This discussion centers on an evaluation of benefits by clinicians, but perhaps the recipients of the benefits, patients themselves, should be the people who define the benefits. They will choose different outcomes than doctors and health care professionals[16].

A further challenge is the current incomplete understanding of the pathological mechanisms in arthritis. Present clinical decision making probably combines several different aspects of pathology into a single decision rule, when in practice the different pathologies have different responses to different treatments. An important recent example in RA is the difference in response of erosion development and clinical inflammation to low dose oral glucocorticoid therapy[17-20]. In osteoarthritis, it has been noted for some time that pain, cartilage loss, and bony overgrowth are only poorly related[21,22], and evaluating analgesics by measuring radiographic change would be considered inappropriate.

A WAY FORWARD

While challenging the very notion of an independently definable minimal clinically important change, the empirical and theoretical arguments above nevertheless allow a number of ways in which some standardization of outcome can be pursued. One approach would be to relate short term changes to longterm clinical progression. Perhaps one particularly useful criterion would be the relationship between changes in a particular measure of the disease and the longterm development of disability. If a link to unequivocal improvement in disability outcome many years later can be demonstrated (well within the range that would be considered clinically important by most people), then the strength of this link could be used to define a MCID. This approach is explored elsewhere in the conference in relation to radiological progression[23]. Another possibility links clinical changes to changes in the underlying pathological process. If, for example, a treatment changes the underlying synovial pathology that is related to pain or joint destruction[24], then the extent of that effect may be sufficient to define a MCID.

The definition of "response criteria," based on current expert (or other) opinion, may nevertheless provide a useful way of describing the outcome of interventions in more understandable terms than population mean changes. In defining response criteria it is not necessary to resort to the authority of a MCID to justify them. Majority voting for such criteria has face validity, but lacks rigor. A better approach might be to use actual clinical decisions to identify patients for which a given proportion of experts agree (as in Figures 3 and 5). An analysis of these judgments made on the patients with agreement will then produce an equation similar in nature to the DAS[2], and the cutoff values for clinically important change calculated.

CONCLUSION

Expert clinicians have widely differing interpretations of minimum clinically important difference in patients with RA. The MCID, like the crock of gold at the end of the rainbow, is attractive but unattainable. Attempts to define a MCID for response will be less rewarding than agreements to define response criteria with no claim to be more than what they are. They will be the opinions of the group that defined the criteria at the time the definition was made. Use of expert panels to derive response criteria provides a very medically oriented definition of response. Opportunities exist for involving other professional groups and patients in defining response to treatment.

REFERENCES

1. Kirwan JR, Chaput de Saintonge DM, Joyce CRB. Clinical judgment analysis. Descriptive modelling of physicians' assessment of patients based on social judgment theory. Q J Med 1990;76:935-49.
2. Van der Heijde DMFM, van't Hof MA, van Reil PLCM, et al. Judging disease activity in clinical practice in rheumatoid arthritis: First steps in the development of a disease activity score. Ann Rheum Dis 1990;49:916-20.
3. Kirwan JR, Chaput de Saintonge DM, Joyce CRB, Currey HLF. Clinical judgement in rheumatoid arthritis. I. Rheumatologists' opinions and the development of paper patients. Ann Rheum Dis 1983;42:644-7.
4. Kirwan JR, Chaput de Saintonge DM, Joyce CRB, Currey HLF. Clinical judgement in rheumatoid arthritis. II. Judging 'current disease activity' in rheumatoid arthritis. Ann Rheum Dis 1983;42:648-51.
5. Kirwan JR, Bellamy N, Condon H, Buchanan WW, Barnes CG. Judgement of 'current disease activity' in rheumatoid arthritis — an international comparison. J Rheumatol 1983;10:901-5.
6. Kirwan JR, Chaput de Saintonge DM, Joyce CRB, Currey HLF. Clinical judgment in rheumatoid arthritis. III. British rheumatologists' judgements of 'change in response to therapy'. Ann Rheum Dis 1984;43:686-94.
7. Kirwan JR, Brooks PM, Currey HLF. Measuring physicians' judgement — The use of clinical data by Australian rheumatologists. Austr NZ J Med 1985;15:738-44.
8. Kirwan JR, Chaput de Saintonge DM, Joyce CRB, Currey HLF. Clinical judgement analysis — practical application in rheumatoid arthritis. Br J Rheumatol 1983;22 Suppl:18-23.
9. Kirwan JR, Chaput de Saintonge DM, Joyce CRB, Holmes J, Currey HLF. Inability of rheumatologists to describe their true policies for assessing rheumatoid arthritis. Ann Rheum Dis 1986;45:156-61.
10. Kirwan JR, Barnes CG, Davies P, Currey HLF. Analysis of clinical judgment helps to improve agreement in the assessment of rheumatoid arthritis. Ann Rheum Dis 1988;47:138-43.
11. Chaput de Saintonge DM, Kirwan JR, Evans SJW, Crane GJ. How can we design trials to detect clinically important changes in disease severity? Br J Clin Pharmacol 1988;26:355-62.
12. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
13. Paulus HE, Egger MJ, Ward JR, Williams HJ. Analysis of improvement in individual rheumatoid arthritis patients treated with disease modifying anti-rheumatic drugs, based on the findings in patients treated with placebo. The CSSRD Group. Arthritis Rheum 1990;33:477-84.
14. Chaput de Saintonge DM, Evans S, Crane G, Kirwan JR. Clinical trials and clinical practice: Do doctors use the same criteria to judge outcome? Br J Clin Pharmacol 1997;44:583-6.
15. Hewlett S, Kirwan JR, Smith A. Perception of disability using time-trade off [abstract]. Br J Rheumatol 2000;39 Suppl:S169.
16. Hewlett S, Kirwan JR. Patients with rheumatoid arthritis rate the value of functions differently to both health professionals and normal volunteers [abstract]. Arthritis Rheum 1997;40 Suppl:534.
17. Kirwan JR and the Arthritis and Rheumatism Council Low-Dose Glucocorticoid Study Group. The effect of glucocorticoids on joint destruction in rheumatoid arthritis. N Engl J Med 1995;333:142-6.
18. Hickling P, Jacoby RK, Kirwan JR, and The Arthritis and Rheumatism Council Low-Dose Glucocorticoid Study Group. Joint destruction after glucocorticoids are withdrawn in early rheumatoid arthritis. Br J Rheumatol 1998;37:930-6.
19. Kirwan JR. The relationship between synovitis and erosions in rheumatoid arthritis. Br J Rheumatol 1997;36:225-8.
20. Wassenberg S, Rau R, Ratingen HZ. Low dose prednisolone therapy retards radiographically detectable destruction in early rheumatoid arthritis [absract]. Arthritis Rheum 1999;42 Suppl:S243.
21. van Baar ME, Dekker J, Lemmens JAM, Oostendorp RAB, Bijlsma JWJ. Pain and disability in patients with osteoarthritis of the hip or knee: the relationship with articular, kinesiological, and psychological characteristics. J Rheumatol 1998;25:125-33.
22. Cicuttini FM, Baker J, Hart DJ, Spector TD. Association of pain with radiological changes in different compartments and views of the knee joint. Osteoarthritis Cartilage 1996;4:143-7.
23. Kirwan JR. Links between radiological change and other outcomes in rheumatoid arthritis. J Rheumatol 2001;28.
24. Mulherin D, FitzGerald O, Bresnihan B. Synovial tissue macrophage populations and articular damage in rheumatoid arthritis. Arthritis Rheum 1996;39:115-24.