Minimal Clinically Important Difference in Plain Films in RA: Group Discussions, Conclusions, and Recommendations

DÉSIRÉE van der HEIJDE, MARISSA LASSERE, JOHN EDMONDS, JOHN KIRWAN, VIBEKE STRAND, and MAARTEN BOERS for the OMERACT Imaging Task Force

ABSTRACT. Analysis of progression of structural damage on an individual patient level in randomized controlled trials provides extra information in addition to the analysis on a group level. A cutoff level is required to define which patients show progression and which patients do not. The objective of the mimimal clinically important difference (MCID) module for plain films was to elaborate the various concepts to determine a MCID for plain films, and if possible, to define a MCID for specific scoring methods. The module comprised preconference reading material, a plenary session, small group discussions, and a plenary report of the group sessions, combined with interactive voting. The following conclusions and recommendations were made: the smallest detectable difference (SDD) beyond measurement error is a good starting point to define MCID; SDD is study-specific; SDD should be reported for all radiographic endpoints used in a trial as a quality control; the expert panel approach is a reasonable method to define MCID, but defined in this way MCID may be smaller than current SDD; more research is needed to validate expert panel based MCID in different datasets and with different experts; a predictive, data driven MCID is the ultimate goal, but is not yet available; the SDD can be used as a proxy for MCID until a data driven MCID is available; analysis at the group level (comparison of means or medians) should remain primary in studies that include progression of joint damage as outcome measure; the proportion of patients showing more progression than the SDD is a secondary outcome measure. (J Rheumatol 2001;28:914–7)

Key Indexing Terms:
SMALLEST DETECTABLE DIFFERENCE
MINIMAL CLINICALLY IMPORTANT DIFFERENCE
SCORING METHODS
RADIOGRAPHS

INTRODUCTION

OMERACT 4 initiated discussions of the smallest detectable difference (SDD) in progression in damage assessed on plain radiographs in rheumatoid arthritis (RA). It was decided that this concept required further elucidation in view of the ultimate goal: to define minimal clinically important differences (MCID) in radiographs in RA. The SDD is based on measurement error. It is defined as the smallest amount of change that can reliably be distinguished from random measurement error. In comparison, MCID is the minimum amount of change that is considered clinically meaningful, and various definitions of clinical outcome can be used. It inherently contains an element of judgment about importance. The objective of the MCID module for plain radiographs in RA at OMERACT 5 was to define and discuss the various concepts required to determine a MCID for plain films and, if possible, to define MCID for 2 scoring methods: Sharp–van der Heijde and Larsen-Scott[1,2].

How is application of the MCID concept relevant to changes on radiographs in RA?

To analyze the effect of treatment on structural damage as measured by radiography, analyses are performed on a group level: median/mean changes are presented per treatment group and tested to determine whether they achieve a prespecified level of significance. This ascertains whether a treatment is able to retard progression of structural damage compared with another treatment (or placebo). However,

From the Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, Maastricht, The Netherlands and Limburg University Center, Diepenbeek, Belgium; Department of Rheumatology, St. George Hospital, Sydney, Australia; Rheumatology Unit, University of Bristol Division of Medicine, Bristol Royal Infirmary, Bristol, UK; Division of Immunology, Stanford University, San Francisco, USA; and Department of Clinical Epidemiology, VU University Hospital, Amsterdam, The Netherlands.

D. van der Heijde, MD, PhD, Professor of Rheumatology, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, and Limburg University Center; M. Lassere, MB, BS, FRACP, PhD, FAFPHM, Research Staff Specialist; J. Edmonds, MBBS, FRACP, Professor of Rheumatology, Department of Rheumatology, St. George Hospital; J. Kirwan, BSc, MD, FRCP, Consultant and Reader in Rheumatology, Rheumatology Unit, University of Bristol Division of Medicine, Bristol Royal Infirmary; V. Strand, MD, Clinical Associate Professor, Division of Immunology, Stanford University; M. Boers, MSc, MD, PhD, Rheumatologist, Professor of Clinical Epidemiology, Department of Clinical Epidemiology, VU University Hospital.

Address reprint requests to Dr. D. van der Heijde, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, PO Box 5800, 6202 AZ Maastricht, The Netherlands.
E-mail: dhe@sint.azm.nl

these data do not offer information on the magnitude of effect on an individual patient basis. These levels of analyses are expressed as one of the axes of the cube of classification of discrimination: change at a group level versus change on an individual patient level[3]. If we want to define how many patients in each treatment group experience a clinically meaningful benefit, a cutoff value needs to be defined so that patients with change greater than this cutoff value are considered to have clinically important deterioration. However, this definition cannot be deduced from results presented on the group level. Even if mean differences between active treatment and control groups are appreciably less than the SDD, the differences may be important because that treatment may have an important effect on many patients[4].

The advantage of presenting data on a patient level is that it makes data more interpretable to physicians as well as patients. Analyses on an individual patient level are already widely applied in interpreting clinical trial data. For example, mean changes in number of swollen joints, C-reactive protein level, or Disease Activity Index (DAS) are examined on a group level and the proportions of patients fulfilling American College of Rheumatology ≥ 20%, ≥ 50%, ≥ 70% criteria (ACR 20/50/70) or DAS response criteria between active and control treatments are examined on an individual level.

The aim of the MCID module for plain radiographs in RA was to derive a definition of the cutoff level for progression and/or lack of deterioration. Three different methods to define this cutoff level were examined and the validity of these methods in the context of radiographic methodology discussed. The methods include: (1) distribution based models, (2) judgmental/authority based models, and (3) predictive/data driven models; these are described in more detail in the introductory paper[5] and can be summarized as follows.

Distribution based models: although there may be several ways to describe features of the distribution, during OMERACT 4 the limits of agreement method was selected as a direct measurement of the error since it can be used to calculate SDD[6]. SDD for progression of structural damage in various radiographic data sets, as well as status of SDD determinations for widely utilized disease activity instruments, were presented.

Judgmental/authority based models: This type of model, based on expert opinion, may provide an external gold standard. Results from a study utilizing an expert panel to define MCID for plain radiographs were presented.

Finally, the predictive/data driven model: this model is used as an evidence based approach and will require extensive data.

MATERIALS AND METHODS

The module consisted of preconference reading material, a plenary session, small group sessions, and report of the group discussions to a plenary session that included interactive voting. The preconference material comprised an introductory paper to the module[5], a paper on the application of the limits of agreement method by Bland and Altman to plain films to determine the SDD[6], a paper on the updated research agenda for imaging[7], the SDD calculated for various clinical disease activity and radiological status measures[8], and results of a study of MCID based on an expert panel[9]. The plenary session gave the background for the MCID concept in radiological progression, and presented relative information needed for the small group discussions. Data presented during the plenary session reviewed the robustness of SDD analyses of radiological progression in various datasets[10] and the links between radiological change and other clinical outcomes[11]. Participants then broke into 8 groups of about 18 persons and discussed 2 of a total of 6 questions in varying combinations.

The following were discussed: If we want to use a dichotomous variable, is the SDD a good starting point as a measurement-error based number to define patients in a clinical trial with progression versus those without progression? Is the SDD robust in various clinical trial datasets? In other words can one SDD be applied to all trials or should it be study-specific? Can expert consensus studies be utilized to define MCID? Would it be feasible to derive MCID from large data sets if these were available? What would be the implications if an expert panel based MCID were to be smaller than the SDD in a given trial? Use MCID or SDD? Should 95% limits of agreement be applied or are less strict limits acceptable?

All discussions were restricted to the situation of a clinical trial with a 1 year duration. Results of the actual MCID were considered to be scoring method dependent and specific.

The OMERACT method to reach consensus has been described[12]. A group leader who had been briefed by the module leaders led the group discussion. Discussions were based on the nominal group technique[13]. This is a technique to generate ideas, freely discuss the items, and make decisions where appropriate. A rapporteur from each group summarized and presented conclusions of the group to the following plenary session. All groups who discussed a given question presented their views, which was immediately followed by interactive voting on this particular subject.

RESULTS

Five groups discussed the suitability of the SDD as a starting point for defining MCID or the robustness of SDD defined in different trials with varying patient characteristics, disease duration, baseline damage, readers, etc. In general, it was felt that the SDD is a conservative measure with a high specificity (avoiding the false positive result of defining patients as progressors who in fact had no deterioration), but a relatively low sensitivity, causing false negative results and missing patients with progression. It was therefore suggested that efforts should be dedicated to minimizing measurement error. It was evident to all groups that the SDD was considered study-specific, as SDD from various datasets showed differences too large to ignore[10]. Although several problems were inherent in the definition of SDD, most groups felt that it could be helpful in presenting data on an individual patient level. Moreover, it was suggested that publishing the SDD for all radiographic endpoints could be used as a measure of quality control in a clinical trial. All groups emphasized that the SDD may be a proxy for MCID, a starting point to define MCID but not equivalent to it. Following these discussions, the majority of participants voted that SDD appears to be a good starting point to define patients with radiographic progression versus

those without progression in a clinical trial. Additionally it was almost unanimously voted that SDD is study-specific.

Some participants considered it a problem that the SDD is calculated post-hoc, and thus it would be impossible to base sample sizes on its calculation. However, it was recognized that the initial analysis should be performed on a continuous measure and this should be used for the power calculation. SDD from the same observers in other but comparable datasets might also be employed as an indication of the expected SDD for a trial.

The groups who addressed whether expert consensus studies could be utilized to define MCID thought that, in principle, the method was valid. However, they emphasized that more information is needed on the reproducibility of results with other expert panels and different datasets. It was considered reassuring that the MCID defined by the expert panel was of similar magnitude to the SDD of the same dataset. An MCID identified by an expert panel was perceived as a followup step to defining SDD. In the plenary vote, the majority of participants judged the expert panel method a valid means to define MCID.

Most of the discussions urged that a link between progression of radiographic damage and longterm clinical outcomes was the ultimate goal in defining MCID. Having concluded this, group discussions diverged, raising a variety of challenges posed by application of this concept. Which clinical outcomes to select? According to whose perspective? Although the patient perspective was considered most important, many groups stressed the equal importance of the societal perspective. Longterm clinical outcomes to be considered included: physical function, work/role disability, surgery, and health resource utilization. And followup of patients should be of sufficient duration — 10 years as a suggested minimum.

Additionally, data should include multiple clinical subsets with variations in age, disease duration, rheumatoid factor, baseline radiographic damage, and country. Although groups believed that a definition of MCID based on the above types of data would be possible, it would be very difficult to accomplish. All agreed that insufficient data are available at present to allow such analyses. Moreoever, it is unclear with currently available outcome measures and statistical techniques if differentiation between small changes (e.g., an increase of 5 units in one year compared with 3 units) may yield a different effect on outcome 10 years later. One group emphasized that knowledge regarding the true linearity of radiographic progression would be essential information in deriving MCID in a longitudinal database. All groups noted that more knowledge regarding the lag time between diagnosis and development of erosions by radiography as well as the time to develop new erosions was essential in better understanding the concepts of SDD and MCID.

Given all these points elucidated in group discussions it was rather surprising (but also hopeful?) that a large majority of participants voted that it is indeed feasible to derive MCID from longitudinal data, assuming that datasets were available. Given the wide variability in choice of clinical outcomes and the various clinical subsets available, it is highly unlikely that a universal definition of MCID will be developed.

Exploratory discussions addressed what to do if the MCID value based on an expert panel was smaller than SDD, in other words that the expert panel judged a progression rate clinically relevant that was less than could be reliably detected beyond measurement error. One group suggested that this could be inherent in the scoring systems currently used. Current scoring methods sum erosions and joint space narrowing, whereas experts judge other features such as osteopenia, cysts, etc. Most participants voted that at the most conservative, the SDD value should be used. However, 28% were unable to answer the question.

In further discussions of this question, it was suggested that the difference between a smaller MCID based on an expert panel and the greater SDD could reflect the stringent use of the 95% limits of agreement method. SDD calculated on a lower limit of agreement will be smaller and the decision of what limit to use is similar to the choice of a power and/or type 2 error for a study (often arbitrarily set at 80%). This idea caused confusion and split opinions. Again, a large proportion of the participants (26%) felt that they could not answer this question. Some participants expressed very strongly that they did not want to change widely accepted conventions. The majority voted to keep the 95% limits of agreement, but there was a broad spectrum of opinion.

CONCLUSIONS AND RECOMMENDATIONS (Table 1)

It was generally accepted that analyses of clinical trial data of plain radiographs in RA on an individual patient basis can make data more interpretable. Results expressed on this level can also be applied in subsequent analyses to determine number needed to treat and to facilitate economic analyses. However, individual patient analyses of radiographic data should be applied only as secondary analyses. Initial analyses should examine continuous variables to

Table 1. Conclusions of participants in the group and plenary sessions on MCID in plain films in RA.

Points of interest
SDD good starting point for MCID
SDD is study specific
Report SDD for all endpoints as a quality control
Expert panel approach valid to define MCID
More research needed to validate expert panel based MCID
Predictive data-driven MCID ultimate goal
Use SDD as a proxy for MCID until data-driven MCID available
Proportion of progressors is a secondary outcome measure

determine whether results meet appropriate criteria for significance and whether the treatment is effective and/or equivalent to another recognized active therapy. Thereafter, the proportion of patients with improvement/deterioration in each treatment group can be elucidated as an aid in interpreting the importance of the results. It must be kept in mind that statistical power may be lost when converting a continuous variable, used for analysis on a group level, into a dichotomous variable, used to facilitate analyses on a patient level.

It was agreed that determination of the SDD is a good starting point for defining MCID. Although expert based determination of MCID is a followup step to expressing SDD, more data are needed to judge the validity, reproducibility, and performance of these definitions in other datasets and with different experts. Although a data driven, predictive MCID is the ultimate goal, this is a challenging task. It is uncertain whether sufficient data and/or satisfactory outcome measures are currently available to allow definition of MCID for plain radiographs in RA. Moreover, it can be expected that different MCID will be defined according to the different datasets available, in part determined by the variability in patient disease characteristics and radiographic progression. At present, the SDD appears to be the only available way to categorize patients as progressors and nonprogressors. Finally, it was recommended that the SDD for radiographic data in each clinical trial be reported to facilitate their use in quality control.

REFERENCES

1. van der Heijde DM. Plain X-rays in rheumatoid arthritis: overview of scoring methods, their reliability and applicability. Baillieres Clin Rheumatol 1996;10:435-53.
2. Scott D, Houssien D, Laasonen L. Proposed modification to Larsen's scoring method for hand and wrist radiographs. Br J Rheumatol 1995;34:56.
3. Beaton DE, Bombardier C, Katz JF, Wright JG, OMERACT MCID working group. Looking for important change/differences in studies of responsiveness. J Rheumatol 2001;28:400-5.
4. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomized trials. BMJ 1998;316:690-3.
5. Lassere M, van der Heijde D. Foundations of the minimal clinically important difference for imaging. J Rheumatol 2001;28:890-1.
6. Lassere M, Boers M, van der Heijde D, et al. Smallest detectable difference in radiological progression. J Rheumatol 1999;26:731-9.
7. Strand V, Lassere M, van der Heijde D, Johnson K, Boers M. Recent rheumatoid arthritis clinical trials using radiographic endpoints — updated research agenda. J Rheumatol 2001;28:887-9.
8. Lassere M, van der Heijde D, Johnson K, Edmonds J. The reliability of measures of disease activity and disease damage in rheumatoid arthritis: Implications for the smallest detectable difference, minimal clinically important difference, and analysis of treatment effects in randomized controlled trials. J Rheumatol 2001;28:892-903.
9. Bruynesteyn K, van der Heijde D, Boers M, et al. Minimal 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.
10. Lassere M, van der Heijde D, Johnson K, et al. Robustness and generalizability of smallest detectable difference in radiological progression. J Rheumatol 2001;28:911-3.
11. Kirwan J. Links between radiological change, disability, and pathology in rheumatoid arthritis. J Rheumatol 2001;28:881-6.
12. Fried BJ, Boers M, Baker PR. A method for achieving consensus on rheumatoid arthritis outcome measures: the OMERACT conference process. J Rheumatol 1993;20:548-51.
13. Delbecq AL, van de Ven AH, Gustafson DH. Group techniques for program planning. A guide to nominal group and delphi processes. Glenview: Scott, Foresman; 1975.