Foundations of the Minimal Clinically Important Difference for Imaging

MARISSA N.D. LASSERE, DÉSIRÉE van der HEIJDE, and KENT R. JOHNSON

ABSTRACT. This article develops a generic conceptual framework for defining and validating the concept of minimal clinically important difference. We propose 3 approaches. The first uses statistical descriptions of the population ("distribution based"), the second relies on experts ("opinion based"), and a third is based on sequential hypothesis formation and testing ("predictive/data driven based"). The first 2 approaches serve as proxies for the third, which is an experimentally driven approach, asking such questions as "What carries the least penalty?" or "What imparts the greatest gain?" As an experimental approach, it has the expected drawbacks, including the need for greater resources, and the need to tolerate trial and error *en route*, compared to the other 2 models. (J Rheumatol 2001;28:890-1)

Key Indexing Terms:
SMALLEST DETECTABLE DIFFERENCE   RANDOMIZED CONTROLLED TRIALS
MINIMUM CLINICALLY IMPORTANT DIFFERENCE

INTRODUCTION

At OMERACT 4, the smallest detectable difference was suggested as a useful starting point in determining what is a minimum clinically important difference (MCID) in radiographic progression[1]. The smallest detectable difference (SDD) is that difference beyond random error of the measurement. In a clinical trial it is applied and interpreted in the context of individual responder status, and is therefore similar in concept to that of the American College of Rheumatology response criteria.

Further evaluation of the validity and robustness of the SDD, and whether it should, in principle, serve as a MCID, are topics on the OMERACT 5 Imaging Module agenda. This article develops a conceptual framework for defining and validating the concept of MCID in imaging.

A STATEMENT OF THE PROBLEM

Rheumatologists have argued that judgments and recordings such as "a little more synovitis," "more pain," "radiographs are a little worse," or even "doing well" may be ineffectual at the time of the assessment, let alone 1, 2, or 5 years later. Numerous systems of formal measurement of disease activity, physical function, and structural damage have been developed. Some, particularly the measures of functional status like the Health Assessment Questionnaire (HAQ), are good predictors of future disability, morbidity, and even mortality in the limited setting of clinical research. However, there remains little agreement among rheumatologists and researchers on which measurements are the most important, what the measurements mean, what amount of improvement or deterioration in a group of patients or in an individual patient is clinically important, and how the measurements should guide clinical decisions.

Which measurements are the most important? This mainly depends on the person you ask. However, an international group with representatives from various backgrounds such as rheumatologists, epidemiologists, and workers in pharmaceutical companies and registration agencies decided to use a core set to be included in all studies[2].

What do the measurements mean, what amount of improvement or deterioration is clinically important, and how should the measurements guide clinical decisions? Rheumatologists constantly make decisions about continuing or changing therapy. What information do rheumatologists base these decisions on? How can we test empirically what is a clinically important improvement in HAQ score[3], in C-reactive protein, in the Disease Activity Index[4] in different individuals? How long should any change be maintained? Although we now have instruments expressly developed to quantify clinical measurements such as pain, physical function, joint synovitis, joint damage, and laboratory measures in rheumatoid arthritis (RA), we have a poor understanding of what these changes mean, particularly in individual men and women with different duration of rheumatoid disease. The classification of the discrimination (differences and changes) can be visualized in a cube[5]. In

From the Department of Rheumatology, St. George Hospital, Sydney, Australia; the Department of Rheumatology, University Hospital Maastricht, Maastricht, The Netherlands, and Limburg University Center, Diepenbeek, Belgium; and CDER, Food and Drug Administration, Rockville, MD, USA.

M. Lassere, MB, BS, PhD, FAFPHM, FRACP, Staff Specialist in Rheumatology, Department of Rheumatology, St. George Hospital; D. van der Heijde, MD, PhD, Professor of Rheumatology, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, and Limburg University Center; K.R. Johnson, MD, Medical Officer, Center for Drug Evaluation and Research, FDA (Note: Views are not necessarily those of the FDA).

Address reprint requests to Dr. M. Lassere, Department of Rheumatology, St. George Hospital, Kogarah, NSW 2217, Australia. E-mail: lasserem@sesahs.nsw.gov.au

this module we will limit the discussions to radiographic measures in RA, particularly how to define and validate MCID for imaging within individual patients.

HOW TO DEFINE MCID FOR IMAGING

What is supposed by the MCID concept in the setting of plain films? How should this question be approached? In reviewing the field, and thinking about various approaches, there appear to be essentially 3 models that can be invoked in strategizing approaches to defining the MCID and these can be fitted within the cube at various levels:

1. Distribution based models

These are purely formal approaches. They begin with the enumeration of the distribution of the measure in the population, and then employ various mathematical functions to describe features of that distribution. For ease of understanding and interpretation, the distribution and its descriptors are often then transposed into qualitative language. The language and the mathematical functions used in 3 such applications of this model are (a) effect sizes (incorporating the mean and the standard deviation, a measure of variability); (b) standard error of the measurement (incorporating the standard deviation and the reliability coefficient of the measure, so arguably better); and (c) limits of agreement, a direct measurement of the error (which takes into account only the reliability). The SDD as determined at OMERACT 4 used this measurement-error distribution based approach as a starting point for defining the MCID in radiological progression. It is important to realize that, in the end, all distribution based methods are arbitrary, based purely on convention.

2. Experiential/opinion based models

These are heuristic approaches, based on adequate surveys of expert opinion, formally executed using the clinician's global assessment. Their authority is predicated on the foundations of experience, knowledge, data, and anecdote. In the end, this model is based on the belief system of the clinician (and importantly for some measures — the belief system and preferences of patients). This form of evidence is intuitive and often is the only evidence that is readily available. As an example of this approach, a study to define MCID based on the judgments of a clinical expert panel will be presented in this module.

3. Predictive/data driven models

This is an experimentally driven approach, asking such questions as "What carries the least penalty?" or "What imparts the greatest gain?" As an experimental approach, it has the expected drawbacks, including the need for greater resources, and the need to tolerate trial and error *en route*,

compared to the other 2 models. But this approach, unlike the experiential approach above, does *not* assume that by sufficiently surveying and analyzing clinical judgment the "truth" will eventually come into focus. It may be that what we believe does not, in fact, contain the truth. Additionally, it does not assume that truth somehow resides in formal descriptive structures. In the final analysis, the above 2 models seem proxies for the third approach. The predictive/data driven model is an evidenced based approach, grounded in experiment, built on knowledge of the causal pathways of disease, and requiring the sequential formation and testing of hypotheses. The experimental venues will include both epidemiology and randomized controlled trials. Robust natural history data will be necessary but rarely sufficient because it cannot address secular changes or effects of interventions. This model will lead to a dynamic process of MCID research, not a static conclusion. For MCID in imaging, this could be the predictiveness of change in radiographic scores, for example, future disability, work loss, or mortality.

OVERVIEW OF THE MODULE

The data retrieved at OMERACT 4 on SDD will serve as a starting point. The robustness of the SDD concept in different datasets, and with different readers, will be presented and compared to the magnitude of SDD of clinical measures. Thereafter the availability of data to use the third strategy, the linkage to other outcomes, will be reviewed. Finally, a MCID based on an expert panel will be presented. After setting the stage with these introductions, group discussions will address questions regarding the validity of these methodologies, the conclusions that can be drawn, and what further research is needed.

REFERENCES

1. Lassere M, Boers M, van der Heijde D, et al. Smallest detectable difference in radiological progression. J Rheumatol 1998;26:731-9.
2. Boers M, Tugwell P, Felson DT, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol 1994;21 Suppl 41:86-9.
3. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the Health Assessment Questionnaire, disability and pain scales. J Rheumatol 1982;9:789-93.
4. Van der Heijde D, van't Hof M, van de Putte L, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Ann Rheum Dis 1990;49:916-20.
5. Beaton D, Bombardier C, Katz J, Wright J, OMERACT MCID Working Group. Looking for important change/differences in studies of responsiveness. J Rheumatol 2001;28:400-5.