Towards a Definition of "Difference" in Osteoarthritis

NICHOLAS BELLAMY, ALISON CARR, MAXIME DOUGADOS, BEV SHEA, and GEORGE WELLS

ABSTRACT. To assess existing information regarding detectable differences in osteoarthritis (OA), a systematic literature search was conducted up to December 1999. Thirty-three articles were considered methodologically relevant to the definition and categorization of detectable differences in OA. It was determined that the musculoskeletal literature contains a wealth of information that relates to observed changes, much of which is derived from the clinical trials literature, but there have been relatively few methodological studies that have systematically evaluated the nature, categorization, and relevance of the change. Furthermore, most of those that have been published take the perspective of an individual or groups of experts other than that of the patient. This summary of the current literature reveals that the diverse sources of information go part way towards developing an understanding of detectable differences and their importance in the area of OA research and clinical practice. Stakeholders' interests as well as factors that modulate perceptions of importance need to be taken under consideration. In particular, the patient's perspective of the importance of change at an individual level requires further evaluation. This area of clinical research is relatively underdeveloped, but there is considerable opportunity for progress. (J Rheumatol 2001;28:427–30)

Key Indexing Terms:
DISCRIMINATION OSTEOARTHRITIS
MINIMALLY CLINICALLY IMPORTANT DIFFERENCE

INTRODUCTION

Health status measurement in osteoarthritis (OA) has undergone progressive evolution in the last 60 years[1], with more rapid change in the last 20 years[2]. Core set domains of pain, physical function, patient global assessment, and for studies of one year or longer, imaging, were established by international consensus at the OMERACT 3 conference[3], and subsequently ratified by the Osteoarthritis Research Society International Task Force on clinical trials[4]. The latter were published within guidelines for the execution of future studies, and contained descriptions of relevant measurement techniques. The last 20 years have seen progress in the development of general measures of musculoskeletal status [e.g., the Health Assessment Questionnaire, Arthritis Impact Measurement Scale (AIMS), and AIMS2], generic health related quality of life measures (Medical Outcomes Survey Short Form-36, EUROQOL, NHP, HUI), and disease- specific measures for OA knee and hip disease [Indices of Clinical Severity, Western Ontario and McMaster University OA (WOMAC) Index, WOMBAT Index] and OA hand disease (Algofunctional Index, AUSCAN Index)[2]. Studies of the relative responsiveness of the WOMAC suggest that disease-specific measures may offer advantage over generic arthritis measures and that disease-specific measures are more responsive than generic Health Related Quality of Life (HRQOL) measures[5]. From a conceptual standpoint, the combination of the disease-specific OA measure and a generic HRQOL measure is advantageous in dissecting the impact of interventions on the hierarchy of health states.

There are several approaches to defining detectable and/or important differences in health state. A taxonomy for responsiveness has recently been proposed by Beaton, *et al*[6], which employs a tri-axial classification system according to who is being analyzed (individuals or groups), when the change is being measured (over time/at what point in time), and the type of change being quantified (e.g., observed change versus important change)[6]. The nature of the change being quantified may be considered from various standpoints: (a) minimum change potentially detectable by the instrument; (b) minimum change detectable given the measurement error; (c) observed change in a given population; (d) observed change in those deemed to have improved (estimated change), and/or (e) observed change in those deemed to have an important change[6]. The last 2 types of changes can be viewed from a number of perspectives, including those of the patient, clinician/researcher, payer, and/or society[6].

To assess existing information regarding detectable

From the Department of Medicine, University of Queensland, Queensland, Australia.

N. Bellamy, MD, MSc, FACP, FRCP(Glas,Edin), FRCPC, FAFRM, FRACP, Department of Medicine, University of Queensland; A. Carr, MPH, PhD, University of Nottingham, UK; M. Dougados, MD, Department of Rheumatology, Hôpital Cochin, Paris, France; B. Shea, MSc, Faculty of Medicine, University of Ottawa, Clinical Epidemiology Unit, Loeb Health Research Institute, Ottawa, Canada; G. Wells, PhD, Faculty of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada.

Address reprint requests to Prof. N. Bellamy, Department of Medicine, University of Queensland, C Floor, Clinical Sciences Building, Royal Brisbane Hospital, Queensland, Australia.
E-mail: nbellamy@medicine.uq.edu.au

differences in OA, a systematic literature search was conducted in MEDLINE, EMBASE, and Current Contents up to December 1999, using text words for OA and minimal clinically important difference, minimum observable or detectable difference, responsiveness, and improvement criteria. In addition, recent conference proceedings and journals were searched for additional relevant studies. The literature search identified 379 articles. Two independent reviewers assessed the titles and abstracts to determine eligibility. A total of 66 articles were considered potentially relevant and were retrieved for closer examination. Of these 66 articles, 33 were considered methodologically relevant to the definition and categorization of detectable differences in OA. The included articles were then evaluated to determine which concepts of the responsiveness cube were addressed in the publication.

The subsequent review noted that the musculoskeletal literature contains a wealth of information that relates to observed changes, much of which is derived from the clinical trials literature, but there have been relatively few methodologic studies that have systematically evaluated the nature, categorization, and relevance/consequence of the change[2]. Further, most of those that have been published take the perspective of an individual or group of experts other than that of the patient. The articles cited in the following paragraphs are considered relevant to the issue of defining various levels of difference in OA, and are for the most part based wholly or partly on the OMERACT/OARSI core set clinical measures of pain, function, and patient global assessment.

MINIMUM CHANGE POTENTIALLY DETECTABLE BY THE INSTRUMENT

The minimum change potentially detectable (MCPD) is a function of the subscale structure and scale length of the instrument. The smallest detectable difference would be one unit, which in the case of a visual analog scale is 1 mm, and in the case of a Likert scale is equivalent to the smallest numerical difference between adjacent grades defined by the scoring system. In the case of the WOMAC LK 3.1 Index, the scale ranges for the component subscales are as follows: pain 0–20, stiffness 0–8, physical function 0–68, total WOMAC 0–96[5]. Given an MCPD of 1 unit, the minimum percentage change potentially detectable (MPCPD) for the respective elements is as follows: pain 5%, stiffness 12.5%, physical function 1.5%, total WOMAC 1%. By comparison the WOMAC VA3.1 uses scale ranges as follows: pain 0–500, stiffness 0–200, physical function 0–1700, and total WOMAC score 0–2400. The MCPD is 1 mm and the MPCPD values are as follows: pain 0.2%, stiffness 0.5%, physical function 0.06%, total WOMAC score 0.04%. In contrast, the Indices of Clinical Severity[7] are scored on a 0–24 scale, with an option for differences of 0.5 in the physical function component to provide an MCPD of 0.5 and an MPCPD of 2%. It should be noted that the Indices of Clinical Severity are aggregated multidimensional indices and that the total WOMAC score would provide a comparable approach to aggregated measurement, albeit using a different weighting system. With the AUSCAN LK3.0 OA Hand Index[8], the length of the subscales are as follows: pain 0–20, stiffness 0–4, physical function 0–36, total AUSCAN score 0–60. The MCPD is 1 unit and the MPCPD values are as follows: pain 5%, stiffness 25%, physical function 2.8%, and AUSCAN total index score 1.7%. The Algofunctional Index contains 10 questions[9]. The scale range of the Algofunctional Index is 0–30, providing an MCPD of 1 unit and an MPCPD of 3.3%.

MINIMAL CHANGE DETECTABLE GIVEN THE MEASUREMENT ERROR

The measurement error can be subdivided according to several sources including the patient and any independent assessor. Circadian variation in pain and function has been observed in OA of the knee and hand using patient self-report methods and performance based measurement techniques[10,11]. Estimates of measurement error need to consider the volatility of the symptom complex and the way in which variations in a specified time frame might influence the determination[2]. As a result there are relatively few published studies that adequately address this issue.

.

OBSERVED CHANGE IN A GIVEN POPULATION

There are several sources for observed change in a given population. The majority come from either cohort/observational studies or from published clinical trials. Such studies need to be interpreted in the light of inclusion/exclusion criteria, the nature of the intervention, and the duration of the study. Relatively few clinical trial reports contain an exact description of the method of deriving the minimum clinically important difference sought and which was used in a sample sized calculation[2].

OBSERVED CHANGE IN THOSE DEEMED TO HAVE IMPROVED

The determination of change can be made by the patient, clinician/researcher, payer, or society[6]. It is to be anticipated that the perception of change might be different between different reference groups. In a recently published study evaluating minimum clinically perceptible improvement (MCPI) in OA patients, the MCPI for the WOMAC pain, function, and stiffness subscales (0–100 mm) were 9.7, 9.3, and 10 mm, respectively, while the MPCI for the investigator global assessment of disease status (0–4) was 0.42[12].

OBSERVED CHANGE IN THOSE DEEMED TO HAVE AN IMPORTANT CHANGE

The perceived importance of change may be different for

different stakeholders. In a group of studies published in *The Journal of Rheumatology*[13-15], a 3 round Delphi exercise was used to define minimum clinically important differences (MCID) for clinical trial purposes for a number of outcome measures used in prior OA clinical trials[2]. The median MCID for a comparative study of 2 nonsteroidal antiinflammatory drugs in a double-blind randomized control parallel trial, in the perception of 6 academic rheumatologists experienced in OA clinical trials and based on actual data from 60 patients, were as follows: Doyle Index 5.5, Physicians Overall Assessment of Pain (visual analog scale, VAS) = 15, Physicians Overall Assessment of Pain (Likert Scale, LK) = 0.78, Physicians Overall Assessment of Morning Stiffness (VAS) = 15, Physicians Overall Assessment of Morning Stiffness (LK) = 0.75, Duration of Morning Stiffness (time between arising and improvement in stiffness) = 0.23, Duration of Morning Stiffness (clock time from awaking to when stiffness begins to wear off) = 20, Duration of Morning Stiffness (time between awakening and when patient is limber) = 0.3, Grip Strength (FDA method) = 37.5, Grip Strength (Dictionary of the Rheumatic Diseases Method) = 37.5, Knee Range of Movement = 15, Intercondylar Distance = 6.5, Intermalleolar Distance = 8, Physicians Overall Assessment of Physical Disability (VAS) = 15, Physicians Overall Assessment of Physical Disability (LK) = 0.68, Investigators subject of opinion of Patients General Condition = 0.90, Physicians Estimate of Disease Activity = 0.78, Physicians Global Assessment of Disease Activity (VAS) = 15, Physicians Global Assessment of Disease Activity (LK) = 0.78, Soft Tissue Swelling = 1.50, Patient Pain at Rest (VAS) = 10.5, Patient Pain on Movement (VAS) = 17.5, Patient Overall Assessment of Pain (VAS) = 15, Patient Overall Assessment of Pain (LK) = 0.78, Subjective Pain Evaluation by Patient = 0.78, Patient's Overall Assessment of Morning Stiffness (VAS) = 17.5, Patient's Overall Assessment of Morning Stiffness (LK) = 0.80, Patient's Overall Assessment of Physical Disability (VAS) = 15, Patient's Overall Assessment of Physical Disability (LK) = 0.8, Lequesne Knee Index = 3, Patient Estimate of Disease Activity = 1, Patient's Opinion of General Condition = 0.9, Patient's Global Assessment of Disease Activity (VA) = 20, and Patient Global Assessment of Disease Activity (LK) = 1.

The recent OARSI Response Criteria Initiative (RCI) has permitted the development of response criteria for clinical trials in OA based on an analysis of 14 placebo controlled clinical trials (totaling 1886 patients). The criteria were presented at the OARSI International Conference in Vienna and use a tree format to categorize patients as responders or nonresponders according to 2 sets of class-specific criteria[16]. The first set of responder criteria are based on a high pain response, or alternatively a lower level of response on at least 2 of the 3 domains: pain, function, and patient global assessment. In contrast, the second set of responder criteria are based on a high level of response in pain or function, or alternatively, a lower level of response on at least 2 of the 3 domains: pain, function, and patient global assessment. These 2 different criteria sets accommodate the dynamic profiles of different classes of interventions. In both sets of criteria, a response is defined by a combination of both absolute and percentage change. As a consequence, they are applicable only to those patients whose symptom severity is such that they could qualify as a responder should their condition improve sufficiently. It is anticipated that the OARSI criteria will require further validation using additional data sets. Doubtless there will be further debate regarding the use of absolute and/or percentage change, the implications of incorporating initial and/or final values, and the implications of dichotomization. Nevertheless, the OARSI responder criteria represent an initial attempt to address the complex and challenging problem of dichotomizing continuous variables, in order to define clinically important changes in health status.

An alternative approach is to provide individual clinical profiles of OA patients to key informants and require them to categorize the patients according to whether they, the key informants, regard the change as being clinically important. Such a project was completed immediately prior to OMERACT 5. The study was based on the WOMAC Index and patient global assessments, and employed a 3-round Delphi exercise to facilitate consensus building. A report is pending.

PATIENTS' DEFINITIONS OF CHANGE

Most assessments of treatment efficacy within clinical trials, and to a lesser extent in clinical practice, are based on clinicians' definitions of clinically important change. Little is known about the degree to which clinicians' and patients' perceptions of clinically important change are concordant, but there is evidence from a number of studies that clinicians are poor judges of the degree of pain suffered by patients, their quality of life, and the relative importance of different treatment outcomes[17-21]. Qualitative research with patients with rheumatoid arthritis and their clinicians has highlighted differences in the ways in which patients and clinicians construct and evaluate disease activity, with patients focusing on the personal consequences in terms of pain and functional limitations and clinicians using biological indicators[22]. In OA, the criteria by which patients judge treatment efficacy appear to focus entirely on pain and function and are very specific, for example, being able to sit through one television program in comfort or being able to walk to a particular shop. There is also some suggestion that patients make "allowances" for treatments they particularly want to work, altering their efficacy criteria when the treatment fails their initial evaluation. These data suggest that patients' assessments of clinically important differences are highly individualized and inconsistent across different inter-

ventions. Quantifying patients' minimum clinically important differences to interpret the results of clinical trials, or as the basis of sample size calculations, may therefore be more complex than using differences derived mathematically from outcome measures or from groups of clinicians. Nevertheless, they are important predictors of health service use. Demand for medical care and treatment change is driven by patients' perceptions of treatment efficacy, and some attempt should be made to include them, particularly in clinical practice.

Issues surrounding the determination of the clinical importance and consequence of structural conservation have received little attention. It will be important to develop outcome measurement strategies for longterm studies. The issues are subtly different in situations where the progression of structural damage may be prevented, slowed, arrested, or reversed. Traditional measures of pain, patient global assessment and especially physical function will be relevant. So too may be the propensity for interventions to reduce the need for total hip replacement surgery[24], although the timing of this endpoint, while clinically relevant, is potentially subject to effects that relate more closely to the health care system in which treatment is being delivered than to actual health status of the individuals concerned.

These diverse sources of information go part way towards developing an understanding of detectable differences and their importance in the area of OA research and clinical practice. Stakeholder interests as well as factors that modulate perceptions of importance need to be taken into consideration. In particular, the patient's perspective of the importance of change at an individual level requires further evaluation. This area of clinical research is relatively underdeveloped, but there is considerable opportunity for progress.

REFERENCES

1. Bellamy N, Buchanan WW. Dr. Douglas Taylor and half a century of musculoskeletal Kelvinism. J Clin Rheumatol 1988:7:272-7.
2. Bellamy N. Musculoskeletal clinical metrology. Lancaster: Kluwer Academic Publishers; 1993.
3. Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997:24;799-802.
4. Osteoarthritis Research Society (OARS) Task Force Report: Design and conduct of clinical trials in patients with osteoarthritis; recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis Cartilage 1996;4:217-43.
5. Bellamy N. WOMAC user guide IV. Brisbane, Australia: 2000:43.
6. Beaton DE, Bombardier C, Katz J et al. Looking for important change/difference in studies of responsiveness. J Rheumatol 2001;28:400-5.
7. Lequesne MG, Mery C, Samson M, et al. Indexes of severity for osteoarthritis of the hip and knee. Validation — value of comparison with other assessment tests. Scand J Rheumatol 1987:65 Suppl:85-9.
8. Bellamy N, Campbell J, Haraoui B, et al. Development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index [abstract]. Arthritis Rheum 1997:40 Suppl:S110.
9. Dreiser RL, Maheu E, Guillou GB, et al. Validation of an algofunctional index for osteoarthritis of the hand. Rev Rheum (English ed.) 1995;62 Suppl 1:43S-54S.
10. Bellamy N, Sothern RB, Campbell J. Rhythmic variations in pain perception in osteoarthritis of the knee. J Rheumatol 1990; 17:364-72.
11. Bellamy N, Campbell J, Sothern R, et al. Circadian rhythm in pain, stiffness and manual dexterity in osteoarthritis of the hand: relationship between discomfort and disability. Chronobiology International 1997:24 Suppl 1:16.
12. Davies G, Ehrich E, Watson D, et al. Minimum perceptible clinical improvement in osteoarthritis: the WOMAC Questionnaire and Global Assessments [abstract]. Ann Rheum Dis 1999;P1078:255.
13. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. I. Effects of standardisation procedures on observer dependent outcome measures. J Rheumatol 1992;19:436-3.
14. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. II. Tables for calculating sample size for clinical trials. J Rheumatol 1992;19:444-50.
15. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials — results of a consensus development (delphi exercise). J Rheumatol 1992;19:451-7.
16. Dougados M, Leclaire P, van der Heijde D, et al. Response criteria for clinical trials on osteoarthritis of the knee and hip. Ostoearthritis Cart 2000 (In Press).
17. Donovan J. Patient education and the consultation: the importance of lay beliefs. Ann Rheum Dis 1991;50:418-21.
18. Kwoh K, Feinstein A, Gifford R, et al. Physician-patient agreement on patients' priorities for important treatment outcomes [abstract]. Arthritis Rheum 1987;30 Suppl 4:S101.
19. Pearlman RA, Uhlmann RF. Quality of life in chronic diseases: perceptions of elderly patients. J Gerontol 1988;43:M25-30.
20. Daltroy L, Eaton H, Hashimoto H, Liang MH. Rheumatoid arthritis patients' expectations for treatment outcomes [abstract]. Arthritis Rheum 1995;38 Suppl:S310.
21. Woolhead GM, Carr AJ, Wilkinson M, Bannister GC, Ackroyd CE. Expectations of treatment and outcome in patients awaiting joint replacement surgery and general orthopaedic outpatients [abstract]. Arthritis Rheum 1996;39 Suppl:S174.
22. Mitchell H, Stowers K, Hughes RA, Carr AJ. Patients' self-assessment of disease activity and severity in rheumatoid arthritis [abstract]. Arthritis Rheum 1999;42 Suppl:S274.
23. Carr AJ, Hughes RA, Stowers K. What influences placebo response to a complementary therapy, glucosamine sulfate? [abstract]. Arthritis Rheum 2000;43 Suppl:S223.
24. Dougados M, Gueguen A, Nguyen M, et al. Requirement for total hip arthroplasty: an outcome measure of hip osteoarthritis. J Rheumatol 1999;26:855-61.