

Minimal Clinically Important Differences: Review of Methods

GEORGE WELLS, DORCAS BEATON, BEVERLEY SHEA, MAARTEN BOERS, LEE SIMON, VIBEKE STRAND, PETER BROOKS, and PETER TUGWELL

ABSTRACT. Determining a minimal clinically important difference to be detected in a clinical trial is a critical methodological step in the design of a study. In this review, the different methods that have been used in detecting important changes or differences are considered and categorized according to a classification system consisting of which groups are contrasted, the setting of the results, and the type of change of difference being quantified. It was found that most methods and procedures for deriving minimal clinically important differences considered important changes from the viewpoint of a group of patients. Development of methods that focus on individuals should be a goal of the future. (J Rheumatol 2001;28:406–12)

Key Indexing Terms:

MINIMAL CLINICALLY IMPORTANT DIFFERENCE METHODOLOGY DISCRIMINATION
CLINICAL IMPORTANCE PATIENT PERSPECTIVE RESPONSIVENESS

INTRODUCTION

The development of a core set of outcome measures for a condition represents one advance in defining and standardizing the outcomes to be measured in clinical trials. However, this is not the end of the endeavor as it relates to the analysis and reporting of clinical trial results. Challenges exist in determining the clinical significance of any change or difference observed in an outcome measure and in developing a single definition of response that a patient has or has not improved based on the assessment of a core set of outcome measures¹. The focus here is on the determination of minimal clinically important differences (MCID).

A MCID can be considered as the smallest change or difference in an outcome measure that is perceived as beneficial and would lead to a change in the patient's medical management, assuming an absence of excessive side effects and costs. In determining a MCID for an outcome measure, several ingredients are needed: an indicator that change has occurred or that a difference exists; an important observed

change or difference based on a valid assignment of importance; and an appropriate method to determine the threshold level within the distribution of important change or difference scores.

The goal is to consider and classify the different methods that have been used in detecting important changes or differences for the purposes of developing the MCID for an outcome measure.

METHODS

An extensive literature search was conducted to retrieve all relevant articles related to specific topics on MCID, as well as any methodology articles published in the medical literature. The search of the bibliographic databases MEDLINE, EMBASE, and Current Contents up to December 1999 included keywords for MCID, minimum observable or detectable difference, responsiveness, and improvement criteria. Over 2000 abstracts were retrieved. Two reviewers independently completed a review of the search results. It was difficult to determine from the title and abstract which of the studies included the methodologies for determining MCID. A reference list of articles was assembled and circulated to content experts for further assessment. This process resulted in a small series of papers describing various methods for deriving MCID.

The Methods section of the retrieved articles was reviewed and the methodology used categorized according to the "cube" classification system for studies of responsiveness by Beaton, *et al*². This classification system consists of 3 axes: "which" groups are being contrasted (i.e., changes within a group over time, differences between groups at one point in time, or differences between groups on changes over time); the "setting" of the results (i.e.,

From the Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada.

G. Wells, PhD, Department of Epidemiology and Community Medicine, University of Ottawa; D. Beaton, PhD, Institute for Work & Health, Toronto; B. Shea, BScN, Clinical Epidemiology Unit, Loeb Health Research Institute, Faculty of Medicine, University of Ottawa; M. Boers, MD, PhD, Department of Clinical Epidemiology, VU University Hospital, Amsterdam, The Netherlands; L. Simon, Deaconess Medicine, Boston, MA, USA; V. Strand, MD, FACP, FACR, Department of Medicine, Stanford University; P. Brooks, MD, FRACP, Executive Dean, Faculty of Health Sciences, University of Queensland, Australia; P. Tugwell, MD, MSc, FRCPC, Faculty of Medicine, University of Ottawa.

Address reprint requests to Dr. G. Wells, Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Health Sciences Complex, 451 Smyth Road, Ottawa, Ontario, Canada.
E-mail: gwells@uottawa.ca

whether they are targeted for use as an “average” for a group of patients or as a “cutoff” for individual patients); and the type of change or difference being quantified. Although there are 5 types, the section of cube corresponding to “important change/difference” is the only source of information useful for the determination of MCID and classifies the different kinds of MCID that can be ascertained. Within this category the perspective taken (patient’s, clinician’s, payer’s, or society’s) on the determination of important change was also described.

RESULTS

Nine procedures were found from the literature review. The description of each procedure is given in the text and its placement in the classification system is given in Figure 1 and Table 1, with the latter providing a more detailed description of the 3 components needed for its classification.

Patient perspective I: Comparison to a global rating
As noted by Jaeschke, *et al*³, clinical experience with an outcome measure can be a valid method for determining the significance of changes in an outcome. That is, clinicians familiar with the outcome measure would have no difficulty

in identifying a change that is clinically important since they are able to come to these conclusions having observed a large number of patients and seeing the changes in function and in clinical course that correspond to the variations in outcome results. However, the significance of changes for unfamiliar assessments, or for those that capture patient-specific attributes, such as quality of life instruments, requires interpretation. They have developed an approach to elucidate the significance of changes in quality of life instruments and have applied it in chronic heart and chronic lung disease.

Procedure. Patients made global ratings on changes in the various domains of interest (15 point global rating scale from -7 to 7 for changes as follows): same (score 0); if worse then consider if almost hardly worse, little, somewhat, moderately, good deal, great deal, very great deal “worse” (score -7 to -1); and if better then consider if almost hardly better, little, somewhat, moderately, good deal, great deal, very great deal “better” (score 1 to 7). It was anticipated that: changes -3 to -1 or 1 to 3 represent small changes (MCID according to “definition”); changes -5, -4 or 4, 5 represent moderate changes; and changes -7, -6 or 6, 7 represent large changes. The relationship between global ratings of change and changes in the outcome measures was

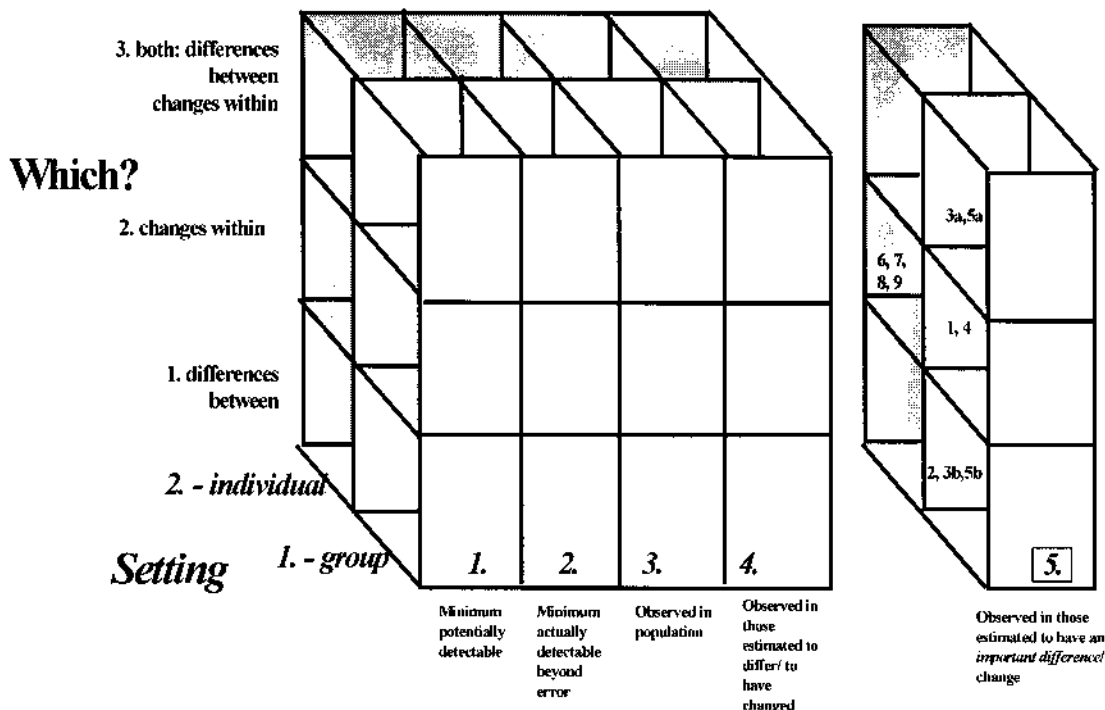


Figure 1. Classification of methods for determining minimal clinically important differences. Type of change/difference: 1. Patient perspective I: Comparison to a global rating. 2. Patient perception II: Patient conversation. 3a. Clinical perspective I: Consensus development (Delphi) a. 3b. Clinical perspective I: Consensus development (Delphi) b. 4. Clinician perspective II: Patient scenario scoring. 5a. Clinician perspective III: Patient scenario comparison a. 5b. Clinician perspective III: Patient scenario comparison b. 6. Clinician perspective IV: Prognostic rating scale. 7. Data driven approach. 8. Discerning important improvement I: Responder criteria. 9. Discerning important improvement II: Achieving treatment goals.

Table 1. Classification of methods for determining minimal clinically important differences.

	Which	Setting	Type = Important Description
Patient perspective I: Comparison to a global rating	Change within (global ratings on change by patients in various domains of interest and assessed changes in outcome measures)	Group (average clinical assessments for different global ratings)	Change determined by patients using specific global ratings on change with surveyed patients indicating whether change is worse, same, or better on a 15 point scale. Minimum important change determined using an indicator based on patient global rating on change corresponding to -3 to -1 or 1 to 3 on the 15 point scale
Patient perception II: Patient conversation	Differences between (differences in clinical outcome measures between patient groups based on subjective comparison ratings)	Group (average clinical assessments for different subjective comparison ratings)	Differences determined using differences in the mean of clinical outcome measures using patient subjective comparison ratings. Minimum important differences determined using differences in the mean of clinical outcome measures in which the patients rated themselves as "somewhat better" and those that rated themselves as "about the same"
Clinical perspective I: Consensus development (Delphi)	a. Differences between changes within (clinicians examine differences between within-group change summary statistics) b. Differences between (clinicians examine end of study summary statistics)	Group (summary statistics compared and considered at group level) Group (summary statistics compared and considered at group level)	Minimum clinically important differences proposed by clinicians for a hypothetical RCT comparing 2 treatments Minimum clinically important differences proposed by clinicians for a hypothetical RCT comparing 2 treatments
Clinician perspective II: Patient scenario scoring	Changes within (clinicians indicate change in outcome measure needed before recommending it)	Group (average change of the response considered)	Differences proposed by physicians by selecting from a number of options listed using both relative and absolute changes. Minimum important difference is the difference between the chosen option and the initial "average"
Clinician perspective III: Patient scenario comparison	a. Differences between changes within (clinicians contrast patients' change in outcome measure between adjacent scenarios) b. Differences between (clinicians contrast patients' outcome measure between adjacent scenarios)	Group (average assessment for different response levels) Group (average assessment for different response levels)	Differences are proposed by physicians by selecting from a number options listed. Minimum important difference determined using difference in outcome measures for pairs rated "a little less" or a "little more" Differences are proposed by physicians by selecting from a number options listed. Minimum important difference determined using difference in outcome measures for pairs rated "a little less" or a "little more"
Clinician perspective IV: Prognostic rating scale	Changes within (describes change from baseline to discharge)	Individual level (ROC analysis done to determine most accurate change score for discriminating between those who improved and those who did not)	Patients given a prognostic rating by treating physiotherapist on admission. Good or excellent prognosis was used as indicator of important improvement
Data driven approach:	Changes within (data used to validate the SEM were longitudinal change scores)	Individual level	SEM (defined as baseline SD × the square root of one minus Cronbach's alpha) considered a proxy for MCID. Similar in magnitude to Jaeschke's approach using same questionnaire
Discerning important improvement I: Improvement criteria	Changes within (patient baseline and end of study data considered by surveyed clinicians)	Individual (patients from clinical trials near expected thresholds for improvement)	Change determined by clinicians surveyed with patients characterized as improved if indicated by "vast" majority; important change determined by controlled randomized trials discriminating efficacious intervention from placebo
Discerning important improvement II: Achieving treatment goals	Changes within (patients followed from admission to discharge from physiotherapy)	Individual level (analysis was done to specifically determine the best cut point to define an important improvement in an individual patient)	Achievement of treatment goals set at the beginning of physiotherapy program used as an indication of an important improvement. Best cut point for individual patient change determined using ROC analysis

RCT: randomized controlled trial; ROC: receiver operating curve; SEM: standard error of measurement; SD: standard deviation.

examined with particular attention to changes in the outcome measures corresponding to the MCID. Absolute change in global ratings and outcome scores was calculated. Assessments were made on the orderly progression of results over the global rating categories and the consistency of the value corresponding to the MCID with an hypothesized value for the MCID based on a consensus among clinicians with extensive experience with the outcome measure.

Placement. The method is designed to consider important changes for a group of patients (2.1.5). For the groups contrasted, the focus is on change within a group over time, with patients making global ratings on changes in various domains of interest and being assessed on changes in outcome measures. The setting of the results is at the group level, with average clinical assessments considered for the different global ratings. The type of change is an “important change” derived by a combination of the patients’ rating of the amount they had changed (on the 15 point scale), and the researchers’ impression of the importance of those ratings, with 1–3 being minimum important change.

Variations. Three other studies have used this same approach. Juniper, *et al*⁴ used the same 15 point global rating, though using a cutoff of ± 2 –3 points as the range reflecting a MCID. Stratford, *et al*^{5,6} used a similar scale in 2 studies. In 1997 the rating scale was completed by patients alone, and in 1998 the average of patient and clinician rating was used. In both, scores greater than +5/7 were considered by the researchers to reflect an important improvement. The change score that was able to discriminate most accurately between those who had and did not have an important change was considered to be the MCID for the scale.

Patient perception II: Patient conversation

For this procedure, patients’ perceptions were used to evaluate when a difference was sufficiently large to be important to the individual patient. This has been termed the threshold of clinical importance. The method has been considered by Redelmeier, *et al*⁷ and Wells, *et al*⁸.

Procedure. A clinical assessment was conducted for each patient. After the assessment, patients were assigned in a random fashion to have a one-to-one conversation with each other. For each conversation, patients were instructed to discuss various specified dimensions of their condition (e.g., pain, function, overall condition). After each conversation and in private, they rated themselves relative to their conversational partner on the dimensions of specified interest (e.g., pain, function, overall condition) using categories “much better,” “somewhat better,” “about the same,” “somewhat worse,” and “much worse” (subjective ratings). For each conversation there were 2 types of comparative ratings: first, the difference between the 2 conversational partners based on their individual clinical assessments; and second, the subjective comparison rating. The threshold of

clinical importance was calculated as the difference in mean scores between conversations in which the participants rated themselves as “somewhat better” and those that rated themselves as “about the same” (alternatively, “somewhat worse” and “about the same”).

Placement. Important differences between results at a group level (1.1.5) are derived using this method. Differences between patients based on the mean of the outcome measure between patients’ conversations are considered. Using study patients, average clinical assessments for different subjective comparison ratings are considered based on patient clinical assessments and patient conversations. Differences are proposed by patients using the subjective comparison ratings, with the minimum important difference determined using an indicator based on differences in the mean of the outcome measure in which the patients rated themselves as “somewhat better” and those that rated themselves as “about the same.”

Clinician perspective I: Consensus development (Delphi)

The Delphi procedure is a well used method for obtaining consensus. This procedure was considered by Bellamy, *et al*⁹⁻¹¹ for deriving MCID for outcome measures for a number of rheumatological disorders.

Procedure. Summary data (proportions, means, and standard deviations) were prepared for each outcome measure of interest based on previous study results. In the first round of the Delphi exercise, these data were sent to each participating clinician. The clinicians examined the data and recommended a MCID for each outcome measure appropriate for a hypothetical randomized controlled trial (RCT) comparing 2 treatments for a particular patient group satisfying a given list of eligibility criteria. The MCID estimates were made independently and returned by each clinician, but they retained the summary data that they had received. The individual estimates of MCID from each clinician for each outcome measure were retabulated and displayed with measures taken to ensure anonymity of individual responses. In the second round, the clinicians were provided with the anonymous recommendations of their colleagues and given the opportunity to modify their recommendations. Following receipt of these data and retabulation, a third round was repeated under identical conditions, with the clinicians given a final opportunity to modify their recommendations.

Placement. The method is designed to determine important differences between changes with results considered at a group level (3.1.5). For a group of patients, differences between changes within the groups are considered by requesting clinicians to examine the data and recommend a MCID in the outcome measure for a hypothetical RCT comparing 2 treatments, thinking about differences between the within-group change scores provided. The results are targeted at a group level with summary statistics compared

for the outcome measure. Important differences are proposed by clinicians with the derived important difference determined using an indicator based on clinician judgment reached by consensus using the Delphi procedure.

This method can also be considered to provide important differences between groups of patients (1.1.5), if the clinicians are asked to examine end-of-trial group scores and to consider what difference in each score they would consider important in a hypothetical trial comparing treatments.

Clinician perspective II: Patient scenario scoring

Van Walraven, *et al*¹² surveyed clinicians using patient scenarios to establish MCID from both the perspective of the absolute risk reduction (ARR) (i.e., difference in event rate between patients on and not on the treatment) and the number needed to treat (NNT) (i.e., number of patients needing treatment to prevent one additional event).

Procedure. Clinicians were enrolled in the survey with questions posed to assess level of experience. Two types of clinical scenarios describing a patient were used. First, clinicians were provided with the average of an outcome measure of interest and asked that if a treatment were known to significantly affect the outcome and there were no contraindications to the treatment, then how much would the outcome measure have to change before they would recommend it. They would select from a number of options that were listed using both relative and absolute changes. The MCID was considered to be the difference between the chosen option and the initial “average.” Second, clinicians were provided with the typical proportion (or percentage) of patients facing a notable event over a specified time period and asked that if a treatment were known to significantly affect the course of this event and there were no contraindications to the treatment, then how much would the proportion (percentage) have to decrease before they would recommend it. They selected from a number of options that were listed using either the absolute risk reduction (ARR) (i.e., difference in event rate between patients taking and not taking the treatment) or the number needed to treat (NNT) (i.e., number of patients needing treatment to prevent one additional event). The MCID was the difference between the typical percentage and the clinician’s response when the ARR approach was taken; or the difference between the typical percent and the reciprocal of clinician’s response when the NNT approach was taken. All clinicians in the study received the first type of scenarios and were randomly allocated to receive the second scenarios in either the ARR or the NNT format.

Placement. This procedure considers important changes within patients with results at a group level (2.1.5). Changes are within patients with clinicians indicating how much the outcome measure would need to change before they would recommend it. The setting of the results is at the group level using averages of the responses to the ARR and NNT. The

differences are proposed from the perspective of the clinicians by having them select from a number of options listed using both relative and absolute changes. The MCID is the difference between the chosen option and the initial “average.”

Clinician perspective III: Patient scenario comparison

In an attempt to determine the clinical significance of differences in pain scores, Todd¹³ conducted a descriptive study to establish this difference by referring pain estimates assigned by clinicians on a visual analog scale (VAS) to categorical measures of pain intensity differences. The objective of the research was to determine the MCID in physician-assigned VAS pain scores.

Procedure. Clinicians were enrolled in the survey with questions posed to assess their level of experience. A number of written scenarios describing patients’ conditions were developed and randomly ordered. The clinicians estimated the patient’s outcome measure and contrasted this outcome with that of the previous scenario (for ‘n’ scenarios will have ‘n–1’ contrasts). As an example, for an outcome measure recorded on VAS with anchors least possible to worst possible, the contrasts can be chosen from 1 of 5 responses (much less, a little less, about the same, a little more, much more). The MCID was defined as the average difference in outcome scores for pairs rated “a little less” or “a little more.”

Placement. This method is designed to provide important differences between changes when clinicians would contrast patients’ change scores in the outcome measure between adjacent scenarios (3.1.5). The setting of the results is at the group level, with average clinical assessments for the different response levels considered. Differences are proposed by clinicians by selecting from a number of options listed; a minimum important difference is the difference in outcome measures for pairs rated “a little less” or “a little more.”

This method can also be considered as providing important differences between patient groups (1.1.5) when clinicians would contrast patients’ scores in the outcome measure between adjacent scenarios at one point in time.

Clinician perspective IV: Prognostic rating scale

Stratford, *et al*¹⁴ reported on the MCID for the Neck Disability Index (NDI) using a prognostic rating scale as the criterion for change. The NDI is a self-report measure of functional status and symptoms related to soft-tissue neck injuries.

Procedure. At baseline clinicians rated the prognosis of each patient on a 5 point scale: little or no change expected in impairment or function, some improvement expected, moderate improvement expected, good improvement expected, excellent improvement expected. This rating was made using the clinician’s judgment based on the patient

history (age, duration of problem), degree of impairment, response to prior treatment, radiographic findings (when available), and natural history of the condition. Prognostic ratings of good or excellent were used to define those patients who underwent important change. Receiver operator characteristic curves were used to find the most accurate change score that discriminated between those that had an important improvement and those that did not.

Placement. This approach would be considered within-person change, and the type of change is important (2.2.5). The results are presented to be applied to an individual patient, with supporting information on the sensitivity (0.78) and specificity (0.80) of the MCID in that patient sample.

Data driven approach

Recently, Wyrwich, *et al*^{15,16} have suggested that the MCID can be determined using the standard error of measurement, which they describe as a “proxy” for the MCID.

Procedure. The standard error of measurement (SEM) is calculated by the authors as being the standard deviation at baseline (s_1) \times the square root of one minus the internal consistency of the scale (Cronbach’s alpha coefficient). $SEM = s_1 * \sqrt{1-r_{xx}}$, where r_{xx} = alpha coefficient. They suggest that this is a fixed estimate for the instrument and, through comparisons with Jaeschke’s approach to the MCID, have found it to be similar to those cutpoints for MCID. This formula is a variation of the formula for the minimal detectable change ($MDC = 1.96 * 2 * SEM$), although in the latter, test-retest reliability is used rather than Wyrwich’s use of alpha (see also McHorney, *et al*¹⁷).

Placement. The authors would place this in the within-person, important change at an individual level (2.2.5).

Discerning important improvement I:

Improvement criteria

The goal of the ACR20 (American College of Rheumatology 20% Improvement criteria; Felson, *et al*¹⁸) and the EULAR response criteria, using the DAS (Disease Activity Score) at different cutoff points (van Gestel, *et al*¹⁹, 1996) was to derive a single definition of response that a patient with rheumatoid arthritis has improved or not based on the assessment of a core set of outcome measures. The interest is to compare the percentage of patients improving, which would resolve the deficiencies associated with a comparison of an average improvement on a specific measure (i.e., an average improvement can occur in many different ways and testing for each measure in the core set increases the type I error). The approach used by Felson, *et al*^{18,20} for evaluating the ACR20 can provide a method for elucidating important changes.

Procedure. The first step was to conduct a survey of clinicians, providing them with information on randomly selected patients from actual clinical trials who were near expected thresholds for improvement. For the outcome

measures, data at baseline and end of study, as well as the percentage change, were provided for each patient, and the surveyed clinicians indicated whether each patient had improved. The analysis focused on patients characterized as improved by a “vast” majority of the surveyed clinicians. The next step was a consideration of the statistical analysis of clinical trial data. Data sets were assembled of appropriate placebo controlled trials with interventions that offered as large an efficacy difference as possible between intervention and placebo and included the outcome measures. The improvement definition was selected that best discriminated an efficacious intervention from placebo. After selecting the definition of improvement based on its performance in placebo controlled trials, it was then evaluated in large comparative trials. Finally, the improvement definition selected was based on ease of use and credibility, with a group of experienced trialists ranking the face validity.

Placement. This procedure provides information on responsiveness, focusing on important changes within individual patients (2.2.5). Changes within patients are considered, with the surveyed clinicians indicating whether a patient had improved based on patient data at baseline and end of study. The setting of the results is at the individual patient level with a definition of improvement. Occurrence of change is determined by the physicians surveyed, with patients characterized as improved if a “vast” majority of the surveyed physicians so indicate. Occurrence of important change is determined using an indicator based on the results of several appropriate placebo controlled randomized trials that discriminate an efficacious intervention from placebo.

Discerning important improvement II: Achieving treatment goals

Riddle²¹ quantified important improvements also without specifically pursuing the MCID. Their work focused on patients with low back pain undergoing physiotherapy treatment, and an important improvement was defined by the achievement of treatment goals.

Procedure. Patients and the clinicians set treatment goals at enrolment (beginning of treatment). At discharge the goals were reviewed and those who had achieved their goals considered to have had an important improvement. The sensitivity and specificity of several different change scores to the occurrence of this important change was evaluated, and the most accurate reported. Of note was that the threshold varied depending on the baseline score for the patient.

Placement. This study would be an individual level study of within-person change focusing on important change (2.2.5). The importance of the change was defined by the clinician and research team as being the achievement of goals. Clinicians and patients decided if they had achieved this threshold.

DISCUSSION

This review of different methods and procedures for deriving MCID found that most of the methods consider important changes from the viewpoint of a group of patients. The change/difference considered is from all perspectives, that is, change within groups, differences between group, and difference between change within group. A few methods (clinician perspective prognostic rating scale, data driven approach, discerning important improvement with responder criteria, and discerning important improvement by achieving treatment goals) consider changes within individual. Development of methods that focus on individuals should be the goal of future research.

REFERENCES

1. Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. *J Rheumatol* 1993;20:561-5.
2. Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, Strand V, Shea B, for the OMERACT MCID Working Group. Looking for important change/differences in studies of responsiveness. *J Rheumatol* 2001;28:400-5.
3. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertain the minimal clinically important difference. *Controlled Clin Trials* 1989;10:407-5.
4. Juniper EF, Guyatt GH, Willan A, Griffith L. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 1994;47:81-7.
5. Stratford PW, Binkley JM. Measurement properties of the RM-18: a modified version of the Roland-Morris Disability Scale. *Spine* 1997;22:2416-21.
6. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris back pain questionnaire. Part 1. *Phys Ther* 1998;78:1186-96.
7. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements: an illustration in rheumatology. *Arch Intern Med* 1993;153:1337-42.
8. Wells GA, Tugwell P, Kraag GR, Baker PRA, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993; 20:557-60.
9. 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.
10. Bellamy N, Anastassiades TP, Buchanan WW, et al. Rheumatoid arthritis antirheumatic drug trials. III. Setting the delta for clinical trials of antirheumatic drugs — results of a consensus development (Delphi) exercise. *J Rheumatol* 1991;18:1908-15.
11. Bellamy N, Buchanan WW, Esdaile JM, et al. Ankylosing spondylitis antirheumatic drug trials. III. Setting the delta for clinical trials of antirheumatic drugs — results of a consensus development (Delphi) exercise. *J Rheumatol* 1991;18:1716-22.
12. van Walraven C, Mahon J, Moher D, Bohm C, Laupacis A. Surveying physicians to determine the minimal important difference: implications for sample size calculation. *J Clin Epidemiol* 1999;52:717-23.
13. Todd KH, Funk JP. The minimum clinically important difference in physician-assigned visual analog pain scores. *Acad Emerg Med* 1996;3:142-6.
14. Stratford PW, Riddle DL, Binkley JM, Spadoni G, Westaway MD, Padfield B. Using the neck disability index to make decisions concerning individual patients. *Physiother Can* 1999;Spring:107-12.
15. Wyrwich KW, Nienaber MA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999;37:469-78.
16. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting standard error of measurement based criterion for identifying meaningful intra-individual change in health-related quality of life. *J Clin Epidemiol* 1999;52:861-73.
17. McHorney CA, Tarlov AR. Individual patient monitoring in clinical practice: Are available health status surveys adequate? *Qual Life Res* 1995;4:293-307.
18. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727-35.
19. van Gestel AM, Prevoo MLL, van't Hof MA, van Rijswijk MH, van de Putte LBA, van Riel PLCM. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. *Arthritis Rheum* 1996;39:34-40.
20. Felson DT, Anderson JJ, Lange MLM, Wells GA, LaValley MP. Should improvement in rheumatoid arthritis clinical trials be defined as fifty percent or seventy percent improvement in core set measures, rather than twenty percent? *Arthritis Rheum* 1998;41:1564-70.
21. Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris back pain questionnaire: 2. *Phys Ther* 1998;78:1197-207.