Understanding the Minimum Clinically Important Difference: A Review of Concepts and Methods

Article *in* The Spine Journal · September 2007

Impact Factor: 2.43 · DOI: 10.1016/j.spinee.2007.01.008 · Source: PubMed

5 authors, including:

David W Polly
University of Minnesota Twin Cities
285 PUBLICATIONS **5,375** CITATIONS
SEE PROFILE

Thomas C Schuler
The Spinal Research Foundation
28 PUBLICATIONS **845** CITATIONS
SEE PROFILE

ELSEVIER

The Spine Journal 7 (2007) 541–546

THE SPINE JOURNAL

Understanding the minimum clinically important difference: a review of concepts and methods

Anne G. Copay, PhD[a,*], Brian R. Subach, MD[a,b], Steven D. Glassman, MD[c], David W. Polly, Jr., MD[d], Thomas C. Schuler, MD[a,b]

[a]*The Spinal Research Foundation, 1831 Wiehle Avenue, Suite 200, Reston, VA 20190, USA*
[b]*Virginia Spine Institute, 1831 Wiehle Avenue, Suite 200, Reston, VA 20190, USA*
[c]*Department of Orthopaedic Surgery, University of Louisville School of Medicine and the Kenton D. Leatherman Spine Center, 210 East Gray Street, Suite 900, Louisville, KY 40202, USA*
[d]*Department of Orthopaedics, University of Minnesota, 2450 Riverside Avenue, SR 200, Minneapolis, MN 55454, USA*

Received 7 September 2006; accepted 24 January 2007

Abstract

BACKGROUND CONTEXT: The effectiveness of spinal surgery as a treatment option is currently evaluated through the assessment of patient-reported outcomes (PROs). The minimum clinically important difference (MCID) represents the smallest improvement considered worthwhile by a patient. The concept of an MCID is offered as the new standard for determining effectiveness of a given treatment and describing patient satisfaction in reference to that treatment.

PURPOSE: Our goal is to review the various definitions of MCID and the methods available to determine MCID.

STUDY DESIGN: The primary means of determining the MCID for a specific treatment are divided into anchor-based and distribution-based methods. Each method is further subdivided and examined in detail.

METHODS: The overall limitations of the MCID concept are first identified. The basic assumptions, statistical biases, and shortcomings of each method are examined in detail.

RESULTS: Each method of determining the MCID has specific shortcomings. Three general limitations in the accurate determination of an MCID have been identified: the multiplicity of MCID determinations, the loss of the patient's perspective, and the relationship between pretreatment baseline and posttreatment change scores.

CONCLUSIONS: An ideal means of determining the MCID for a given intervention is yet to be determined. It is possible to develop a useful method provided that the assumptions and methodology are initially declared. Our efforts toward the establishment of a MCID will rely on the establishment of specific external criteria based on the symptoms of the patient and treatment intervention being evaluated.

Keywords: Outcomes measures; Metrics; Minimum clinically important difference; Disability; Functional assessment

Introduction

In everyday discourse, a significant difference is understood as a change that is important or meaningful. In the world of statistics, a significant difference is simply a difference that is unlikely to be caused by chance or happenstance and has a mathematical basis for such a claim. In the realm of health care, a difference may be statistically significant based on a simple numerical value, yet may at the same time be of little or no importance to the health or quality of life of patients afflicted by a certain disease. Furthermore, the size of the sample being tested will often contribute to the statistical significance of a given variable, such that a seemingly unimportant detail may gain apparent

FDA device/drug status: not applicable.

Authors acknowledge a financial relationship (Medtronic), which may indirectly relate to the subject of this research.

* Corresponding author. Virginia Spine Institute, 1831 Wiehle Avenue, Suite 200, Reston, VA 20190. Tel.: (703) 709-1114; fax: (703) 709-1117.
E-mail address: acopay@spinemd.com (A.G. Copay)

1529-9430/07/$ – see front matter
doi:10.1016/j.spinee.2007.01.008

statistical significance. For example, most clinical trials and multicenter studies rely on large patient samples; small treatment effects may be identified as statistically significant. As a result, clinicians are left to decipher the importance of seemingly statistically significant results to their own patients.

The concept of a "clinically important difference" evolved as a way to overcome the shortcomings of the "statistically significant difference." A clinically important difference represents a change that would be considered meaningful and worthwhile by the patient such that he/she would consider repeating the intervention if it were his/her choice to make again. The minimum clinically important difference (MCID) is a threshold value for such a change. Any amount of change greater than the MCID threshold is considered to be meaningful or important. Any patient whose answers allow them to reach the MCID threshold are considered "responders" [1]. The proportion of responders to total patients involved in a given treatment indicates to a clinician the likelihood of his/her patients also responding favorably to the same treatment [2].

The definition of an MCID would be particularly helpful in the evaluation of patient-reported outcomes (PROs). PROs refer to the patient's evaluation of a health condition and its treatment. PROs are generally used in medical research for three reasons [1]. First, the patient may be the only source of information. Some treatment effects can only be experienced by the patient. For instance, we rely entirely on a patient to assess physiological responses such as pain and nausea, for which there are no adequate observable or physical measures. Second, clinical measurements do not always match a patient's evaluation. Improvement in a clinical measure does not always correspond to an improvement in patient's pain and disability. For instance, the radiological evidence of bridging bone after spinal fusion surgery indicates a radiographically successful surgery. However, patients may still report pain and disability despite such evidence of bony fusion [3,4]. Third, clinicians may obtain information known only to patients by directly asking questions of them. This is commonly done in clinical evaluations. PROs provide a structured and reproducible way to obtain this information. PROs are not biased by third-party interpretation and often remain reliable in studies in which significant interobserver bias may occur.

Patient reporting of outcomes has become an integral part of the evaluation of spinal interventions because back pain treatment primarily affects a patient's pain level, functional status, and overall quality of life [5]. Commonly used PRO measurements in spinal surgery research are pain scales, the Oswestry Disability Index, and the Short Form of the Medical Outcomes Study. No definitive MCID values have been established for these three PRO instruments in patients undergoing spinal surgery.

MCID was originally defined as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" [6]. This definition was later simplified to "the smallest change that is important to patients" [7]. MCID has recently been called "the new metric on the block" [8]. This review will describe the methods used to calculate MCID and their shortcomings. As it will be described, many issues are still unresolved in the determination of MCID, which render its new metrics status decidedly premature.

Determinations of MCID

Two general approaches have been used to determine MCID: anchor-based methods and distribution-based methods. All approaches measure a quantifiable change in outcomes, but the specific choice of approach will decide the type of change measured [9,10].

Anchor-based approaches

Anchor-based approaches compare the change in PRO score to some other measure of change, considered an anchor or external criterion. Very few studies have relied on an objective external criterion. For instance, Farrar et al. [2] compared patients' ratings on pain scales to the objective amount of ingested pain medication. Most commonly, studies compare PRO scores to the patients' answers to another subjective assessment, typically a Global Assessment Rating in which the patients rate themselves to some extent as "better," "unchanged," or "worse." The choice of a subjective assessment as an external criterion is not ideal but is due to the lack of satisfying objective assessment, a situation that spurred the use of PRO in the first place. The use of global ratings is debated due to their unknown validity and reliability [7]. Possible interference by "recall bias" with such global ratings has also been described [11]. However, global assessment scales have been shown to be very sensitive to change, both positive and negative [12]. Some attempts have been made to substantiate the patients' subjective assessment by combining the clinicians' assessment with the patients' assessment [7,13] or by using the physical therapists' reports of the reason for discharging patients [14]. Regardless of the instrument chosen as external criterion, there often needs to be an established association between the external criterion and the PRO measurement to make any meaningful inference about PRO scores. It is not known how small the association may be and still allow for useful and meaningful inferences [15].

Although the use of an external criterion is the common characteristic of all anchor-based approaches, many differences remain between approaches. Four variations may be identified among the anchor-based approaches.

"Within-patients" score change

The first anchor-based approach [6,16] defines MCID as the change in PRO scores of a group of patients selected according to their answers to a global assessment scale. In the classic anchor-based studies [6,16], patients rated their change on a 15-point global scale (−7="much worse" to 0="no change" to +7="much better"). The MCID was defined as the average change of the patients who exhibited small changes (ie, who scored themselves with±1, 2, or 3 [6]). Later, a score of 1 was considered equivalent to 0, and only the patients who scored±2 or 3 were used to determine MCID [16]. Others have used a similar technique with different scales. For instance, van der Roer et al. [17] used a 6-point scale: 1="completely recovered," 2="much improved," 3="slightly improved," 4="no change," 5="slightly worsened," or 6="much worse." The MCID was the mean change in scores of the "much improved patients." In this first approach, the selection of a group of patients as markers of MCID is arbitrary. The arbitrariness stems from the number of levels in the original scales and from the combination of levels to form the selected patient group.

"Between-patients" score change

A second approach is to compare the PRO scores [18] or the PRO change scores of groups of patients [19] with different responses to a global assessment scale. A cross-sectional study defines a MCID as the score difference between two adjacent levels on a global rating: "not at all impaired" patients and "very mildly impaired" patients [18]. In a longitudinal study, patients rated themselves as "much better," "better," "unchanged," or "worse." The MCID was defined as the difference in the change score of the "better" and "unchanged" patients [19]. Conceptually, a minimum difference should be a difference between two adjacent levels of a scale. In this approach, the selection of the two adjacent levels of the scale is arbitrary. Furthermore, the levels may be formed by arbitrarily combining several original scale levels.

Sensitivity- and specificity-based approach

A third approach is to select as MCID, a score that allows for the best discrimination between groups of patients (ie, a score that produces the greatest sensitivity and/or specificity). In diagnostic tests, sensitivity is the proportion of true positives, or patients with the condition, who have a positive test result. Specificity is the proportion of true negatives, or patients without the condition, who have a negative test result [20]. Used in conjunction with MCID and applied to both functional and quality-of-life scales, sensitivity is the proportion of patients who report an improvement on the external criterion and whose PRO scores are above the threshold MCID value. Specificity is the proportion of patients who do not report an improvement on the external criterion and whose PRO scores are below the threshold MCID value. A sensitivity of 1 indicates that all true positives are identified, whereas a specificity of 1 indicates that all true negatives are identified. Unfortunately, a desirable MCID sensitivity or specificity level has yet to be determined. Typically, researchers have chosen to have equal sensitivity and specificity for their MCID value. Similarly to the previous approaches, there remains a degree of arbitrary error inherent in the definition of patients who either report or conversely fail to report an improvement.

Receiver operating characteristic (ROC) curves have also been used to identify the PRO score with equal sensitivity and specificity [7,14,17]. For instance, in the previously mentioned van der Roer [17] study with six possible response levels, patients were combined into 1 and 2 "improved"; 3, 4, and 5 "unchanged"; or 6 "deteriorated." ROC curves were used to determine the score with equal sensitivity and specificity to discriminate between the "improved" and "unchanged" patients [17] (ROC curves also rely on the arbitrary patient groups selection). Additionally, the area under the curve of an ROC curve represents the probability that scores will correctly discriminate between improved and unimproved patients (as classified by the external criterion). Probabilities range from 0.5 to 1, with 0.5 representing the ability to discriminate because of chance and 1 representing the ability to correctly discriminate all the patients. An area of 0.7 to 0.8 is considered acceptable and an area of 0.8 to 0.9 excellent [21].

Social comparison approach

A fourth, but not widely used, approach has patients compare themselves with other patients. Patients are paired with other patients to discuss their health situation. After the discussion, patients rate themselves as the same or to varying degrees of worse or better than the patient with whom they spoke. The MCID is the difference in scores of patients who rate themselves as "a little better" or "a little worse" instead of "about the same" as compared to the other patient [22].

Distribution-based approaches

Distribution-based approaches compare the change in PRO scores to some measure of variability such as the standard error of measurement (SEM), the standard deviation (SD), the effect size, or the minimum detectable change (MDC).

SEM

The SEM is the variation in the scores due to the unreliability of the scale or measure used. A change, smaller than the identified SEM, is the likely result of measurement error rather than a true observed change. Wyrwich et al. noticed that the value of 1 SEM corresponded to the MCID value when defined with the classic anchored-based method

[23]. Those authors then defined 1 SEM as the MCID in heart failure and respiratory disease patients [23,24]. On the other hand, Ware et al. [25] used 2 SEM to classify 5 groups of patients as better, worse, or unchanged, but this fact has escaped the discussion of MCID. There is no agreement yet that 1 SEM could be a general MCID value across instruments. At the least, 1 SEM may be used as the yardstick of true change for individual change scores and possibly for mean group change scores.

MDC

Associated with the SEM is the MDC, also called the smallest detectable change. The MDC is the smallest change that can be considered above the measurement error with a given level of confidence (usually 95% confidence level) [5]. Clearly, a valid MDIC should then be at least as large as the observed MDC [5,19]. A related concept, the reliable index change (RCI) is obtained by dividing the individual patient change score by the square root of the SEM. If the RCI is greater than 1.96, the change in the patient is considered to be a true change with 95% confidence [26,27]. RCI by itself does not constitute a potential MCID, but it has been used in conjunction with other methods in studies of MCID [26].

SD

The SD is the variation among a group of scores. Norman et al. [28] found that the value of 0.5 SD corresponded to the MCID across a variety of studies.The authors attributed their finding to the fact that 0.5 SD represents the limit of the human mental discriminative capacity, a limit that would appear in most patient-reported outcomes. They also noted that 0.5 SD is equivalent to 1SEM for a reliability of 0.75 [28].

Effect size

Effect size is a standardized measure of change obtained by dividing the difference in scores from baseline to post-treatment by the SD of the baseline scores. The value of the effect size represents the number of SDs by which the scores have changed from baseline to posttreatment. By convention, an effect size of 0.2 is considered small, 0.5 moderate, and 0.8 large [29]. Used in conjunction with an external criterion, effect size ascertains the responsiveness of the external criterion. For instance, the effect size should be small in patients reporting no change and large in patients reporting a great improvement [30]. In regard to MCID, the change in scores corresponding to the small effect size is considered the MCID [18,30–32]. To calculate the change score equivalent to the MCID, one multiplies the SD of the baseline scores by 0.2 (the small effect size) [31].

Limitations of MCID determinations

Three main limitations remain in the methods of MCID definitions: each method produces a MCID value different from the other methods, MCID definitions do not take into account the cost of treatment to the patient, and the change in PRO scores depends on the patient initial baseline status.

Multiple MCID values

The thrust of MCID studies is the search for a unique threshold value, whereas, ironically, the different methods produce a variety of MCID values. Anchor-based methods will produce different MCID depending on the criterion scale and the arbitrary selection or grouping of scale levels. Conceptually, a minimal difference is a difference between two adjacent levels on a scale, such as ‘‘unchanged’’ and ‘‘slightly better.’’ MCID would then depend on the number of levels on a scale: the larger the number of levels, the smaller the difference between two adjacent levels, and the smaller the MCID. The proximity of two scale levels makes it more likely that there would be no statistical difference between adjacent scale levels and that MCID would be small enough to fall within the boundaries of measurement error. Combining levels on a scale is a common but arbitrary procedure in MCID studies. Because ROC curves require a dichotomous variable, arbitrary selection or grouping of scale levels is necessary for studies relying on ROC curves.

Distribution-based methods also yield different values of MCID depending on the measure of statistical variability. Although methods relying on SEM and MCD ensure the statistical soundness of a MCID value, other methods do not. More importantly, distribution-based approaches do not address the question of clinical importance and ignore the purpose of MCID, which is to distinctly separate clinical importance from statistical significance. Another limitation of distribution-based approaches is the fact that they are sample specific in the sense that the MCID value depends on the variability of the scores in the studied samples.

The cost of the treatment is not taken into account

The original definition of MCID stated that a patient might decide whether a change was clinically significant or not by weighing the costs and benefits of the change. Most studies (even the original one) rely on a global assessment scale as a measurement of the clinically significant change. A global assessment scale indicates whether a change has occurred but does not take into account the cost of that change. For example, a patient might agree that he/she has improved, but based on the cost, may also consider that the benefit gained was not worth it [32]. ‘‘If there were no side effects and no costs, then any improvement in symptoms would be worth having—no matter what’’ [33].

The changes are associated with the baseline scores

PRO changes are associated with baseline level: patients with greater level of disability show greater improvement [7,14]. There are several potential reasons for the association between baseline scores and change scores: regression to the mean, floor and ceiling effect, and use of noninterval scales. Some solutions have been proposed to take into account the association with baseline scores: the use of statistical control, the use of percent change, and the creation of a range of MCID values.

Potential reasons for the association with baseline scores

Regression to the mean

At follow-up, patients will score closer to the average score if their initial score was extreme because of chance (such as an unusually good or bad day).

Floor and ceiling effects

Patients whose initial scores are close to the ends of the scale are not able to register a large change because such a change would exceed the span of the scale.

Use of noninterval scales

PRO scales are not true interval scales (ie, the meaning or amount of change differs according to the position on the scale). Intuitively, the amount and quality of change would differ between 1 and 2 and 9 and 10 on a 10-level pain scale. Similarly, the amount and quality of change is likely to be different for improvement than deterioration [2,15]. This fact particularly influences the determination of MCID [18].

Proposed solution to the association with baseline scores

Statistical control

Statistical methods may be used to eliminate the effect of baseline scores.These statistical controls assume that extreme baseline scores are because of chance or error. However, patients involved in medical research would be expected to have more extreme scores because of their medical condition. Statistical control of baseline scores would thus mask true variation [34,35].

Use the percentage of change rather than raw change scores

Using percent change scores will correct for high baseline scores when high baseline scores carry the possibility of large change.This is the case for pain scales and the Oswestry Disability Index. On the other hand, using percent change scores will increase the association with baseline scores when high baseline scores carry the possibility of small change. This is the case in instruments in which a high score indicates a better health status such as the Short Form of the Medical Outcomes Study. Hence, using percent change score will not correct for the baseline scores in all instruments. A secondary advantage of percent change is the fact that they allow for comparison across different instruments [26].

Define a range of MCID rather than an absolute MCID

Because the meaning of change differs according to the level of a scale, MCID could be defined at different levels of a scale. Studies have arbitrarily divided the range of their PRO scores into two [17] or five [7,14] sections and determined a MCID for each section. This method relies on an arbitrary division of PRO scores into sections and, ultimately, detracts from the advantage of a single MCID threshold value.

Summary

Clearly, clinicians need a systematic way to assess the perceived benefit of a certain treatment based on individual patient improvement relative to both cost and risk of complications. MCID would ideally provide a specific threshold to serve as a treatment goal and is already used in that regard. For instance, MCID is included in the evaluation of clinical trials, but the Food and Drug Administration admits its need for further information about MCID [1]. The potential usefulness of MCID is to serve as a benchmark for improvement of individual patients. The success of treatment would then be measured by the proportion of patients who reach MCID as opposed to the average change of a group of patients. Beyond the values presented thus far in the spinal research literature, additional studies are needed to define MCID thresholds for specific patient populations.

To establish MCID as a useful measure of treatment outcomes, two major steps have to be taken. First, researchers have to come to an agreement regarding the appropriate method to determine MCID. Second, efforts have to be made to translate PRO and MCID scores into concrete changes for the patients. The most useful concept of MCID should carry the certainty that treatment is effective and should alert the physician and the patient to its impact on patient's life.

Acknowledgment

The authors thank the members of the Lumbar Spine Study Group for their supported.

References

[1] U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Rockville, MD: U.S. Department of Health and Human Services, 2006.

[2] Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. Pain 2000;88:287–94.
[3] Fritzell P, Hägg O, Wessberg P, Nordwall A, TSLSSG. Lumbar fusion versus nonsurgical treatment for chronic low back pain. A multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. Spine 2001;26:2521–34.
[4] Laasonen EM, Soini J. Low-back pain after lumbar fusion. Spine 1989;14:210–3.
[5] Beaton DE. Understanding the relevance of measured change through studies of responsiveness. Spine 2000;25:3192–9.
[6] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407–15.
[7] 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.
[8] Schwartzstein RM, Harver A. Interactive textbook on clinical research, ch 23. Dyspnea: NIH 2003. URL: http://symptomresearch.nih.gov/chapter_23/sec29/cahs29pg1.htm. Accessed June 14, 2006.
[9] Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of responsiveness. J Rheumatol 2001;28:400–5.
[10] Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status. Mayo Clin Proc 2002;77:371–83.
[11] Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol 1997;50:869–79.
[12] Hagg O, Fritzell P, Oden A, Nordwall A, the Swedish Lumbar Spine Study Group. Simplifying outcome measurement. Evaluation of instruments for measuring outcome after fusion surgery for chronic low back pain. Spine 2002;27:1213–22.
[13] Walsh TL, Hanscom B, Lurie JD, Weinstein JN. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Owestry Disability Index, MODEMS, and the SF-36. Spine 2003;28:607–15.
[14] Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part 2. Phys Ther 1998;78:1197–207.
[15] Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, the Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. Mayo ClinProc 2002;77:371–83.
[16] 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.
[17] van der Roer N, Ostelo RWJG, Bekkering GE, van Tulder MW, de Vet HCW. Minimally clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. Spine 2006;31:578–82.
[18] Kulkarni AV. Distribution-based and anchor-based approaches provided different interpretability estimates for the Hydrocephalus Outcome Questionnaire. J Clin Epidemiol 2006;59:176–84.
[19] Hägg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J 2003;12:12–20.
[20] Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? JAMA 1994;271:703–7.
[21] Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley, Inc, 2000.
[22] Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. J Clin Epidemiol 1996;49:1215–9.
[23] Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky F. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care 1999;37:469–78.
[24] Wyrwich KW, Tierney WM, Wolinsky F. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 1999;52:861–73.
[25] Ware JE, Kosinski M, Keller SK. SF-36 physical and mental health summaries scales: a user's manual. Boston, MA: The Health Institute, 1994.
[26] Bolton JE. Sensitivity and specificity of outcome measures in patients with neck pain: detecting clinically significant improvement. Spine 2004;29:2410–7.
[27] Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59:12–9.
[28] Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life. The remarkable universality of half a standard deviation. Med Care 2003;41:582–92.
[29] Cohen J. Statistical power: analysis for the behavioural sciences. New York: Academic Press, 1977.
[30] Taylor SJ, Taylor AE, Foy MA, Fogg AJB. Responsiveness of common outcome measures for patients with low back pain. Spine 1999;24:1805–12.
[31] Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures. A general approach with illustration to the Health Utilities index Mark II. Pharmacoeconomics 1999;15:141–55.
[32] Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. Pharmacoeconomics 2000;18:419–23.
[33] Kirwan JR. Minimum clinically important difference: the crock of gold at the end of the rainbow? J Rheumatol 2001;28:439–44.
[34] Jamieson J. Dealing with baseline differences: two principles and two dilemmas. Int J Psychophysiol 1999;31:155–61.
[35] Beaton DE, Boers M, Wells G. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Curr Opinion Rheumatol 2002;14:109–14.