Minimal Clinically Important Rehabilitation Effects in Patients with Osteoarthritis of the Lower Extremities

FELIX ANGST, ANDRÉ AESCHLIMANN, BEAT A. MICHEL, and GEROLD STUCKI

ABSTRACT. ***Objective.*** To estimate minimal clinically important differences (MCID) of effects measured by the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) in patients with osteoarthritis (OA) of the lower extremities undergoing a comprehensive inpatient rehabilitation intervention.

Methods. A prospective cohort study assessed patients' health by the WOMAC at baseline (entry into the clinic) and at the 3 month followup, and by a transition questionnaire asking about the change of "health in general related to the OA joint" during that time period. The WOMAC section score differences between the "equal" group and the "slightly better" and "slightly worse" groups resulted in the MCID for improvement and for worsening.

Results. In total 192 patients were followed up. The MCID for improvement ranged from 0.80 to 1.01 points on the continuous WOMAC numerical rating scale from 0 to 10, reflecting changes of 17 to 22% of baseline scores. The MCID for worsening conditions ranged from 0.29 (6%) to 1.03 points (22%). In the transition reply subjectively unchanged patients reported a "pessimistic bias" of 0.35 to 0.51 points, except for the stiffness section. Both MCID and pessimistic bias showed regression to the mean and baseline dependency.

Conclusion. The assessment of MCID using the transition method is a heuristic and valid strategy to detect particular rehabilitation effects in patients with OA of the lower extremities with the use of the WOMAC, and it is worth implementing. The size of the MCID and of the systematic bias is comparable to that assessed by other methods and in other therapeutic settings. (J Rheumatol 2002;29:131–8)

Key Indexing Terms:
OSTEOARTHRITIS REHABILITATION BIOMETRY WOMAC SF-36

Inpatient rehabilitation intervention is an increasingly important therapeutic strategy consisting of multifactorial and multidisciplinary medical practice where a set of particular interventions together improve patients' state of health and quality of life. Within this overall improvement particular effects can be very small and difficult to discern, but assessment of the effects is essential for evidence based therapeutic interventions. There is a need for responsive and sensitive instruments to assess overall quality of life as well as specific symptoms and (dis)abilities[1-5].

The WOMAC (Western Ontario and McMaster Universities Osteoarthritis Index) is the most widely used and best proven condition-specific health assessment instrument for patients with osteoarthritis (OA) of the lower extremities to measure condition-specific dimensions as well as overall state of health[6-12].

While many reports on instruments rely on statistically detectable differences, clinical importance and impact of effects are rarely examined and discussed. Often, this issue is simply summarized by a statement like "in our experience, the changes seen are clinically important [or clinically trivial]"[13], which lacks an exact quantified assessment. The minimal clinically important difference (MCID) can be defined generally as the smallest difference in score (i.e., the effect) that patients perceive as beneficial and which would then mandate, in the absence of troublesome side effects and excessive costs, a change in the patient's management[13]. When assessing interventional effects that ought to be clinically relevant, a primary requirement of an instrument is that it gives an estimate of the MCID. For example, this can be important in a clinical situation to determine the sample size for a clinical trial by the estimation of an expected effect. The planning of a study in which one wants to prove a clinically relevant effect will be more precise if given the MCID from a pilot study or from research in the literature.

For patients with joint arthroplasty there is well documented data on the size of effects, but there is a lack of estimation of smaller effects, such as those reported by rehabilitation intervention[7,14-18]. Further, it is more precise to determine MCID in interventions resulting in small effects. The assess-

From the Clinic of Rheumatology and Rehabilitation, Zurzach; Department of Physical Medicine and Rehabilitation, University of Zurich, Zurich, Switzerland; and Department of Physical Medicine and Rehabilitation, University of Munich, Munich, Germany.

Supported by the Zurzach Rehabilitation Foundation SPA.

F. Angst, MD, MPH; A. Aeschlimann, MD, Professor Clinic of Rheumatology and Rehabilitation; B.A. Michel, MD, Professor, Department of Physical Medicine and Rehabilitation, University of Zurich; G. Stucki, MD, MS, Professor, Department of Physical Medicine and Rehabilitation, University of Munich.

Address reprint requests to Dr. F. Angst, Schiedhaldenstrasse 1, CH-8700 Küsnacht, Switzerland. E-mail: fangst@datacomm.ch

Submitted January 24, 2001; revision accepted July 26, 2001.

ment of positive and negative effects enables the determination not only of a minimal clinically important improvement, but also of a minimal clinically important worsening, thus providing a more balanced estimate of MCID.

Our overall objective was to estimate the size of MCID for the WOMAC in patients with OA of hip or knee using patients' subjective perceptions of transition effects before and after a comprehensive rehabilitation intervention. The specific aims were first, to determine "systematic bias," defined as the mean effect within the group of patients who assessed themselves as unchanged by a transition reply; and second, to estimate the dependency of the MCID on the baseline scores. Alternative ways to assess clinically important effects were compared.

MATERIALS AND METHODS

Patients and data collection. Consecutive patients referred for a 3 to 4 week comprehensive inpatient rehabilitation intervention to the Zurzach Rheumatology and Rehabilitation Clinic were assessed by the WOMAC at entry (baseline examination) and by the WOMAC and a "transition" questionnaire at 3 month followup after baseline examination. On the day of entry to the clinic, a trained physician performed the baseline interview and examination, which determined eligibility for the study. The comprehensive rehabilitation intervention of usually 3 to 4 weeks' duration consisted of passive physical therapy, such as electrotherapies, hydrotherapies, thermotherapies, massage, and others, and of especially active physical therapy to strengthen and stretch the musculature and passive structures, and to reestablish regular joint mobility. The use of nonsteroidal antiinflammatory drugs (NSAID) and analgesics was minimized as far as possible.

According to the American College of Rheumatology (ACR) guidelines, inclusion criteria for knee OA were: knee pain for more than 25 of the last 30 days, morning stiffness < 30 min, and crepitating in the knee, or pain for more than 25 of the last 30 days and osteophytes on radiographs indicating knee OA[19]. Patients with hip OA were included when there was pain for more than 25 of the last 30 days and at least 2 of the following 3 criteria: erythrocyte sedimentation rate < 20 mm/h, osteophytes on radiograph, or obliteration of joint space[19]. Patients were excluded if they did not fulfill the ACR criteria, had a history of medication abuse or noncompliance, had difficulty completing questionnaires, or had severe illness or arthroplasty of the joint in question.

Measures. We used 2 concepts to measure changes in the patients' health status. The first was the 2 point measure in WOMAC scores resulting in effects defined by the difference of the score between the 3 month followup and the baseline examination[6]. The second was the patient's self-assessment (at the 3 month followup) of the global change in health status between baseline and the 3 month followup measured by the "transitional" scale[13].

The condition-specific WOMAC measures quality of life and particular disabilities in a multidimensional set of 24 items graded in a numerical rating scale ranging continuously from 0 ("no symptoms") to 10 ("extreme symptoms")[6]. Its validity and reliability have been proved in different settings[2,7-12]. The unweighted mean of 5 predetermined items results in the pain score, the mean of 2 items in the stiffness score, and the mean of 17 items in the physical function score[6]. Patients included in the analysis had filled out the questionnaires in accordance with the missing rules of the WOMAC user guide, which specifies completion of at least 4 of the 5 pain items, one of the 2 stiffness items, and 14 of the 17 function items in the WOMAC[6].

To detect clinically relevant changes in the patient's subjective perception, the concept of the "transition" method was used. This concept was originally developed by Jaeschke, *et al*[13], introduced into rheumatology by Redelmeier and Lorig[22], and since applied in different settings[4,13,21-24]. The so-called transition questionnaire investigated the current state of health in general related to the OA joint at the 3 month followup compared to the state 3 months earlier (at baseline examination) by the question: "Please imagine how you would have described your health status three months ago. How do you feel in general today as compared to three months earlier as far as your osteoarthritis is concerned?" The possible replies were "much worse," "slightly worse," "equal," "slightly better," and "much better." The questionnaire was sent by mail to the patients at the 3 month followup and assessed the change of health status in written format independently of the WOMAC or any other health assessment instrument.

Analysis. The WOMAC score at the 3 month followup minus the score at baseline examination prior to the intervention defined the effect measured by WOMAC. The self-perceived change at the 3 month followup compared to baseline was assessed by the transition reply. To determine MCID and systematic bias, the WOMAC effects were related to the transition replies. The difference between the mean effects measured by WOMAC of the "slightly better" group and the "equal" group was defined as the MCID of improvement, and that of the "slightly worse" group and the "equal" group as the MCID for worsening. This method has been proposed and applied in different settings[4,13,21-24]. We graphed the mean effects of the transition categories on bar charts in order to compare them visually. An alternative illustration was done by the cumulative frequency distribution of the effects stratified by transition categories, as shown by Ehrich, *et al*[25].

Within the "equal" transition group, one would expect that the effects measured by the WOMAC would be zero if patients in this group correctly assessed their health status subjectively as unchanged. If not, the WOMAC effect can be considered a bias caused by a subjective misjudgment of the change in health status between baseline and the 3 month followup. Assuming this misjudgment was in the same direction and was the same size among the patients of the other transition groups, this bias could be considered "systematic," as proposed by Redelmeier and Lorig[22]. In this case, the mean within-effect of the "equal" group can be subtracted from the mean effects of the other transition groups, resulting in "pure" within-effects for the transition groups and, by definition, the MCID for improvement ("pure" mean effect of the "slightly better" group) and for worsening ("pure" mean effect of the "slightly worse" group). This step is illustrated in Figures 1, 2, and 3.

The "systematic" bias is (subjectively) pessimistic if the mean WOMAC effect of the "equal" group is negative, indicating improvement of health status, and, analogously, is (subjectively) optimistic if the mean WOMAC effect is positive, indicating worsening of health status[22]. However, there is no further assessment of other possible systematic bias or of classical (random) error of measurement occurring in a specific setting with a specific measure by this concept as classically defined in epidemiology.

To examine whether effects were dependent on the baseline score, effects were graphed, stratified by baseline scores. In order to yield at least 10 subjects and thus a good estimate of the mean effect per stratum, this stratification had to end after splitting the sample in 2 halves by the baseline median.

RESULTS

Patients at baseline examination. In the period December 1997 to December 1999, 225 patients fulfilling the ACR criteria were included in the study. Between the baseline examination and the 3 month followup, 2 patients had to be excluded due to severe illness and 9 due to arthroplasty. Eight more did not want to participate further, and 14 either did not return complete sets of questionnaires according to the missing rules for analysis of the WOMAC or did not complete the transition questionnaire sufficiently.

The remaining 192 patients (85.3% of 225) were 40 to 86 years old (median 65.5); 135 (70.3%) patients were female and 57 (29.7%) were male; 113 (58.9%) had knee OA and 79 (41.1%) hip OA. For highest level of education, 71 (37.0%) patients had completed the basic school of 9 years, another 77

(40.1%) had completed vocational training, and 31 (16.1%) had completed high school and/or university. No significant between-group differences were observed for these sociodemographic characteristics between the included patients and the 33 who dropped out of the analysis between baseline and the 3 month followup.

The mean baseline scores listed in Table 1 ranged from 4.62 to 4.65 in the middle of the 10 point scale, and were comparable across the sections for pain, stiffness, function, and global. The baseline scores varied widely within the sections, as indicated by standard deviations of around 2.1. However, there were few patients with minimal and maximal scores, i.e., there were low floor and low ceiling effects.

Patients at the 3 month followup, MCID. The followup scores and the raw mean effects stratified by the transition reply are listed in Table 1. Compared to baseline, the overall mean effect at the followup was –0.70 (standard deviation 2.00) for pain, –0.19 (2.54) for stiffness, –0.44 (1.75) for function, and

Table 1. Patients with OA of the lower extremities (n = 192) before and after an inpatient rehabilitation intervention: WOMAC scores and effects (difference of baseline to 3 month followup) by transition reply.

	Score, Mean (SD)		Mean Effect Within Transition Group, WOMAC Score Points (SD)						MCID for WOMAC Score Points (SD) = % of Baseline Score
	Baseline	3 Mo Followup	Much Better	Slightly Better	Equal	Slightly Worse	Much Worse	Worsening	Improvement
n	192	192	26	47	64	38	17	(38 vs 64)	(47 vs 64)
Pain	4.62 (2.13)	3.92 (2.37)	–2.19 (1.66)	–1.34 (1.82)	–0.51 (1.74)	0.13 (2.01)	0.88 (1.72)	0.64 (2.01) = 14% (44%)	–0.83 (1.72) = 18% (37%)
Stiffness	4.65 (2.60)	4.46 (2.49)	–1.56 (2.41)	–0.89 (2.06)	0.12 (2.41)	0.41 (3.11)	1.35 (1.63)	0.29 (3.11) = 6% (67%)	–1.01 (1.63) = 22% (35%)
Function	4.63 (2.15)	4.19 (2.38)	–1.66 (1.20)	–1.19 (1.52)	–0.39 (1.39)	0.64 (1.88)	0.87 (1.82)	1.03 (1.88) = 22% (41%)	–0.80 (1.82) = 17% (39%)
Global	4.63 (2.05)	4.16 (2.31)	–1.73 (1.20)	–1.17 (1.49)	–0.35 (1.42)	0.61 (1.98)	0.84 (1.71)	0.96 (1.98) = 21% (43%)	–0.82 (1.71) = 18% (37%)

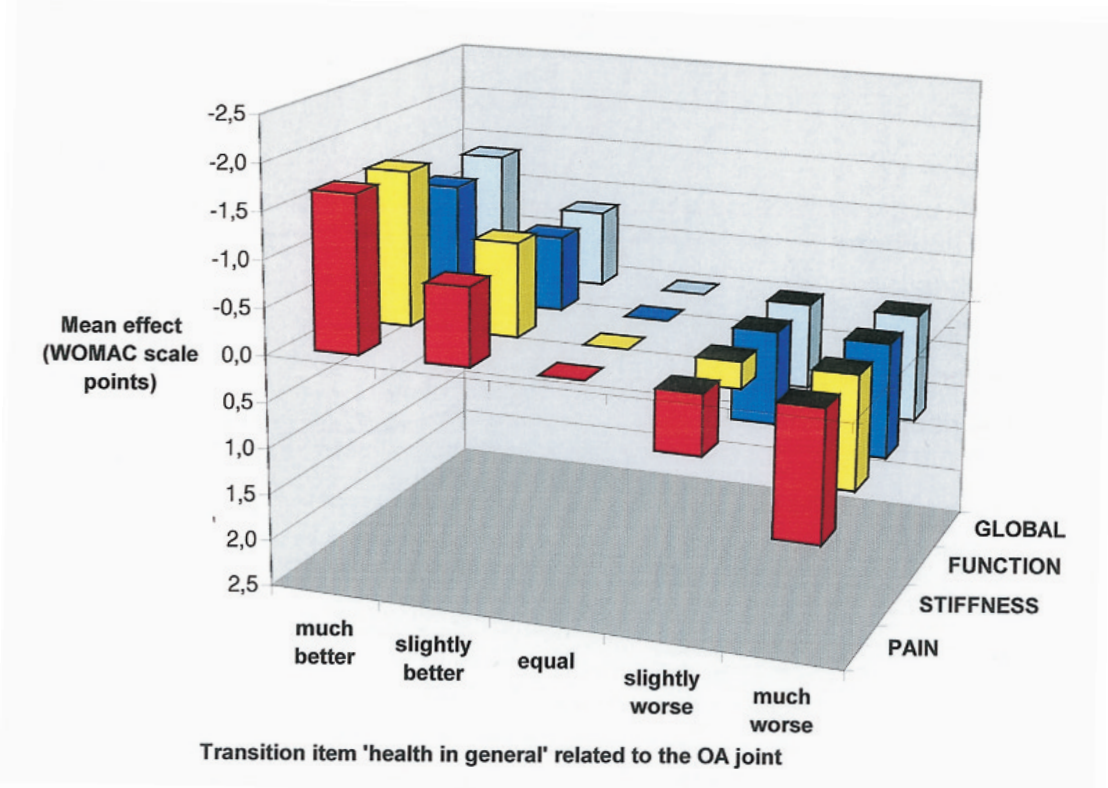

Figure 1. Mean WOMAC effects (3 month followup vs baseline) within transition groups of OA patients (n = 192) after rehabilitation: Without correction (subtraction) of the "equal" effect.

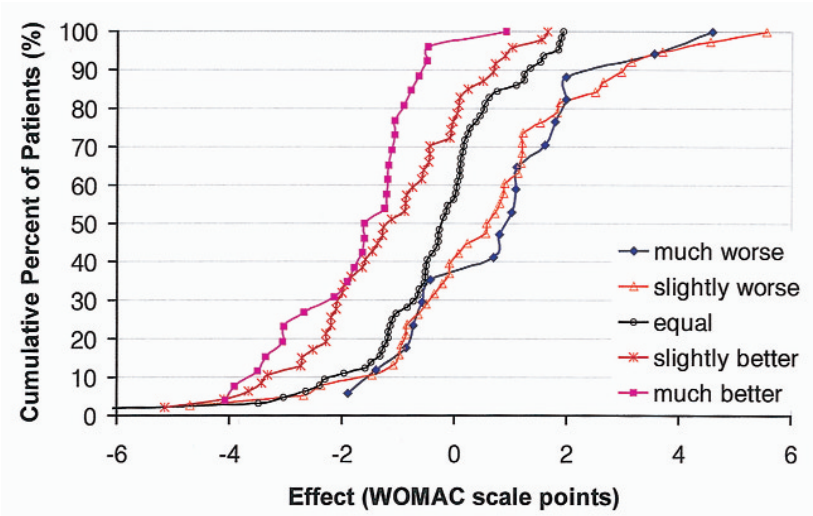

Figure 2. Cumulative frequency distribution of WOMAC effects of OA patients (n = 192) after rehabilitation, stratified by transition reply.

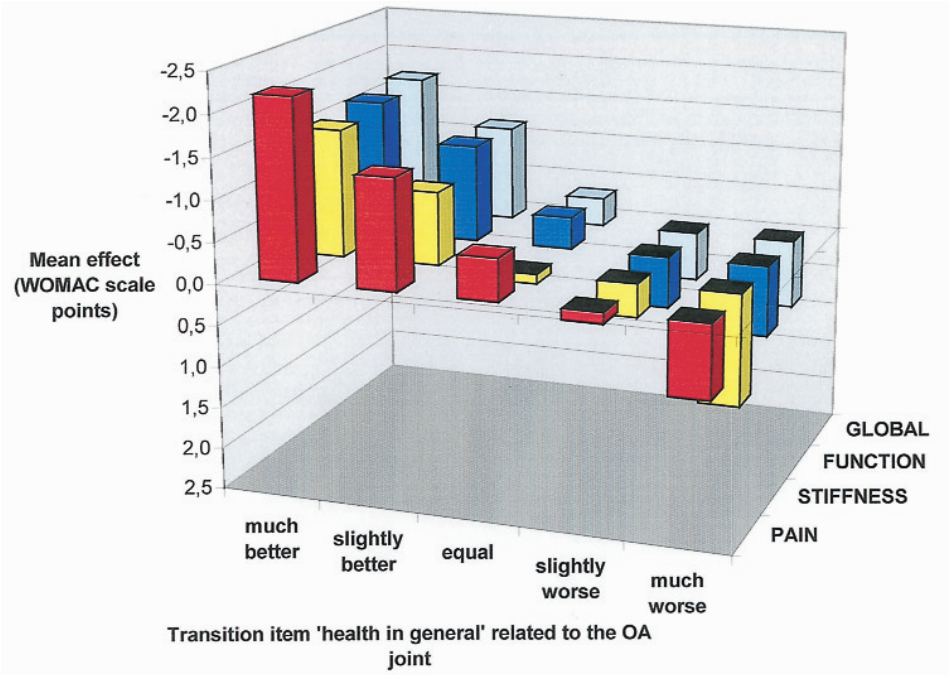

Figure 3. Mean WOMAC effects (3 month followup vs baseline) within transition groups of OA patients (n = 192) after rehabilitation: After correction (subtraction) of the "equal" effect.

–0.47 (1.76) for the global score, reflecting improvement in all sections. The effects within the transition categories listed in Table 1 are illustrated by bar charts in Figure 1 and by their cumulative frequency distribution in Figure 2. Overall, the effects within the transition categories were well correlated to the transition reply, i.e., the better the self-assessment by transition reply, the greater was the resulting improvement as measured by the WOMAC. The mean effects of the transition "equal" group, which is considered a "systematic bias," showed improvement between –0.35 (1.42) and –0.51 (1.74) WOMAC scale points, except for the stiffness section, which worsened slightly [+0.12 (2.41)].

Figure 3 shows the effects of the transition categories corrected (subtracted) by the mean effect of the "equal" group (systematic bias). There was good symmetry of the effects in the higher transition categories "much better" and "much worse," with values between (±) 1.2 and 1.7 WOMAC points. (Bars in Figure 1, data not shown; these values can be determined by subtracting the effects of the "equal" group from Table 1.) When comparing the effects of the categories

Table 2. Mean WOMAC global effects totally, within the transition "equal" group and MCID of OA patients after rehabilitation (n = 192) stratified by the median of the baseline score: WOMAC global score.

	WOMAC Global Score, Mean (SD)		Effect Within Transition, Mean (SD)	MCID for WOMAC Score Points (SD) = % of Baseline Score (SD)	
	Baseline	3 Mo Followup	Equal	Worsening	Improvement
n below/above median	96/96	96/96	26/38	(24 vs 26/14 vs 38)	(22 vs 26/25 vs 38)
Below median	2.93 (1.05)	2.79 (1.76)	–0.03 (1.20)	1.04 (1.92) = 35% (66%)	–0.60 (1.30) = 20% (44%)
Above median	6.37 (1.16)	5.55 (1.95)	–0.57 (1.52)	0.49 (1.96) = 8% (31%)	–1.08 (1.51) = 17% (24%)

"slightly worse" and "slightly better" there was also good symmetry for pain (+0.64, –0.83 points) and for function (+1.03, –0.80), but not for stiffness (+0.29, –1.01). The stiffness section represents the mean of only 2 WOMAC items and showed higher variance (see standard deviations) than the pain and the function section. The later listed figures represent the MCID for worsening and improvement and are listed as absolute values and relative values (percentages) in relation to the baseline score (= 100%) in Table 1. The MCID for worsening varied more (0.29 to 1.03; 6 to 22% of baseline score) than the MCID for improvement (0.80 to 1.03; 17 to 22% of baseline score) due to the data in the stiffness section.

Baseline dependency, regression-to-the-mean effect. Figure 4 and Table 2 illustrate the baseline dependency of the global WOMAC effects after correction by the effect of the "equal" group, because the analysis was done within the 2 groups that are separated by the median. The more severely ill group (above the median, effect of "equal": –0.57, SD 1.52) showed higher positive as well as negative effects than the less severely ill group (below the median, effect of "equal": –0.03, SD 1.20), except within the "slightly worse" group.

The course of patient groups is illustrated in Figure 5 for the mean WOMAC global score within the 10 classes separated by the 10% percentiles of baseline score. The less severely ill patients (with low baseline scores) showed lower improvement (or even worsening) than the more severely ill patients, whose improvement grew with the level of the baseline score so that the 10 graphs are focused on the followup assessment, illustrating a regression-to-the-mean effect.

DISCUSSION

We used the concept of the "transition" method to define MCID by the patient's self-assessment of "health in general related to the OA joint," by asking not only about slight improvement but also about slight worsening, in order to obtain a more balanced estimate of MCID compared to other studies[4,13,21-24]. To date, MCID has never been assessed in rehabilitation settings.

The MCID for improvement showed relatively symmetric values compared to the MCID for worsening. This was increasingly valid as the number of WOMAC items by which the WOMAC section was determined increased. Thus, the MCID of the global score, that is, the mean of all 24 WOMAC items, was quite symmetric — 0.82 points for improvement and 0.96 for worsening. The MCID of function, the mean of 17 items, was still moderately symmetric (–0.80, 1.03), whereas the MCID of stiffness, the mean of only 2 items, showed asymmetric figures: –1.01 for improvement and 0.29 for worsening. This may be due to a statistical effect of small numbers, or more probably it is due to the hypothesis that stiffness is not as important as pain and function in the subjective assessment of health status by the transition query, or it may be due to the low responsiveness of the stiffness section[17,21].

It is important to note that the overall perception of change using a transition question is experienced in a sophisticated way in a variety of problem areas such as pain, stiffness, and function that are addressed in a multidimensional rehabilitation intervention. This is emphasized by the fact that determination of MCID for the stiffness section, which is known from other studies to be of low responsiveness[17,21], was not possible in all cases. In comprehensive interventions such as rehabilitation, it may therefore be preferable to measure not only general health effects, but also problem-area-specific transition dimensions similar to the problem-area-specific measurement, i.e., with the WOMAC.

The baseline scores of our 192 patients centered around 4.6, i.e., in the middle of the 10 point WOMAC scale. To quantify MCID relative to baseline score is more appropriate than by absolute points since MCID are baseline dependent, as shown in our data. The size of MCID for improvement of 17 to 22% of the baseline scores is slightly higher than, but comparable and consistent with, the MCID determined by the transition method applied to patients with chronic respiratory disease or chronic heart failure, where dyspnea, fatigue, and emotional function were assessed by Jaeschke, *et al*[13]. In that study, on a 7 point scale the MCID had an average value of 0.5, resulting in a relative MCID of 14% of the middle score of 3.5 points.

As well, in patients with hip or knee OA and using the transition concept, Ehrich, *et al* determined the MCID for improvement as 0.97 WOMAC points/15% of baseline score

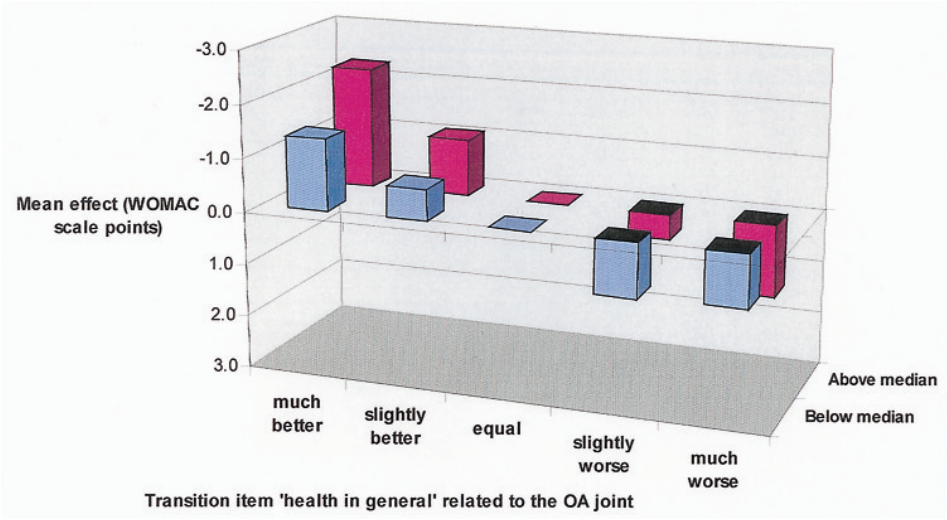

Figure 4. Mean WOMAC global effects (3 month followup vs baseline) within transition groups of OA patients (n = 192) after rehabilitation stratified by baseline score: After correction (subtraction) of the "equal" effect.

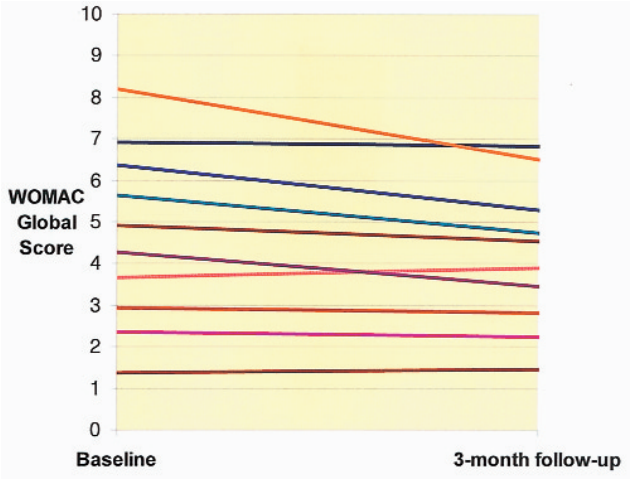

Figure 5. Course of OA patients (n = 192) after rehabilitation measured by the WOMAC global score within groups separated by the 10% percentiles of the baseline score. A negative slope of the graphs reflects improvement, a positive slope, worsening.

for pain, 1.00/16% for stiffness, and 0.93/14% for function after a treatment with NSAID, but they did not give an assessment of the MCID for worsening[25].

Kosinski, *et al* assessed patients with rheumatoid arthritis before NSAID treatment and 6 weeks after baseline by the SF-36, the Health Assessment Questionnaire (HAQ) for patients' global health and pain rating and a physician's global health rating using the levels "very good," "good," "fair," "poor," and "very poor"[26]. The "minimal important changes" for improvement were defined by an increase of one level in the health rating. Most of the mean changes for improvement in the SF-36 scales and the HAQ were around 20% per level of the health ratings. For example, there were changes of 20% of baseline in SF-36 physical function (per one level in the patient's rating), 39% in SF-36 bodily pain (per one level in patient rating), and 18% in HAQ (per one level in patient rating), and these were therefore partly comparable to our findings.

In our data, the size of the MCID for improvement was also comparable with the MCID determined by the method where patients with arthritis assessed each other's state of health, as described by Redelmeier and Lorig[22]. The 2 scores of the HAQ functional health assessment instrument were compared when one patient was considering his overall health state as the same as that of another patient in one-to-one conversations. A difference of 0.19 units, i.e., 21% of the baseline score (0.92 units), was necessary for an assessment of "somewhat better" in that study.

Another method to describe clinically meaningful changes in health status can be applied by the effect size involving the disturbance of information by statistical variability[4,27]. The effect size (ES) is the effect between 2 measurements divided by the standard deviation of the baseline measurement. Arbitrarily, an ES of 0.20, i.e., an effect that is a fifth of the baseline SD, has been defined as "small," representing a minimal clinically important difference[27]. Further, Deyo, *et al* used ES combined with the transition method to distinguish patient groups with unchanged and with improved followup data[4]. From their data the minimal clinically important ES can be determined as 0.33. In our data the MCID for improvement of pain, –0.83 points, and function, –0.80 points, can be related to the baseline standard deviations of 2.13 and 2.15 (Table 1), resulting in an ES of 0.39 for pain and 0.37 for function, which reveals good comparability of our method to the concept reported by Deyo, *et al*[4] and Kazis, *et al*[27].

The effects measured as WOMAC score differences were "systematically pessimistic" compared to the patient's percep-

tion of unchanged state of health as measured by the transition method, except for the stiffness section. The patients in our study showed a pessimistic bias of –0.51 WOMAC points/11.0% of the baseline score for pain, optimistic 0.12/2.6% for stiffness, and pessimistic –0.39/8.4% for function. Thus, correction by this "systematic bias" of the "equal" group of patients, who considered themselves as unchanged, results in a symmetry of effects for the "much better" and the "much worse" groups, as well as a similar distribution for the "slightly better" and "slightly worse" groups.

A systematic pessimistic bias was also found in patients with Crohn's disease using the transition method[28]. Within the "equal" group, the effects ranged from 2.4% to 12.1% of the baseline score using a generic measure, and they were 5.5% using a condition-specific measure, indicating results comparable to our findings of 8.4% for function and 11.0% for pain[28]. The followup examination of patients with chronic low back pain who received an outpatient rehabilitation intervention resulted in a pessimistic bias of 17.9% improvement of the baseline score for the patients who judged themselves as clinically unchanged[4], a figure that is higher than ours. In contrast with these findings of pessimistic bias was Ehrich's study of hip or knee OA, in which the observed bias was optimistic with 0.21 points/3.2% of the baseline value for pain, 0.42/6.6% for stiffness, and 0.37/5.6% for function[25].

In Redelmeier, *et al*, a systematic optimistic bias was shown. The arthritis patient's HAQ score was 0.15 (5% on the 3 point scale) points worse than the partner's HAQ score when the patient considered his/her state of health as "about the same" compared to that of the partner in overall functional status, meaning that patients did underestimate their functional disability[22]. While the size of the systematic bias of 5.1 to 6.5% (on the 10 point scale) in our study is similar to Redelmeier's, our bias was pessimistic and was therefore in the opposite direction.

It is not surprising that patients overestimate their health when faced with other patients, as Redelmeier, *et al* reported, as it resembles the well known cognitive phenomenon of overestimating oneself relative to others[29]. It remains unclear why the transition method leads to a pessimistic bias in general. A possible explanation may be that the pessimistic transition question represents some kind of dissatisfaction with respect to unrealistically high expectations prior to the rehabilitation intervention.

When faced with a systematic bias, one can try to adjust for it as suggested by several authors[22,25,29]. However, the strong baseline dependency of effects that was also found by Redelmeier and Lorig[22] and Bindman, *et al*[30] requires adjustment by baseline strata. The importance of adjusting effects by baseline scores is emphasized by the fact that only stratified correction for the systematic bias led to symmetry in both directions of change, i.e., worsening and improvement. While one would ideally stratify by as many strata as possible, in a sample of 192 patients with 5 transition categories, it seems reasonable to limit stratification to obtain at least n = 10 in each of the cells in order to improve validity of the mean effects of the strata. The effect's baseline dependency is caused not only by the regression-to-the-mean phenomenon but also by the properties of effect measurement by a limited/"closed" scale such as the WOMAC from 0 to 10 points. Patients with lower baseline scores have a higher probability of increasing the score at the followup, reflecting worsening of their health status, than patients with higher baseline scores, and vice versa. In the extreme case, at the end of the scale, only one direction of an effect $\neq 0$ is possible: worsening when baseline = 0 (ceiling effect) and improvement when baseline = 10 (floor effect)[30]. With respect to this, MCID should be related to baseline measurement in every study.

Our results may have some further limitations. First, the concept of determination of MCID by the transition query in relation to WOMAC effects can be considered a weakness of the method since the one point transition measurement of subjective change in global health status based on patient's memory stays in contrast to the more objective 2 point assessment of the WOMAC. One could hypothesize that a 2 point measure of the global health status and its effect (baseline vs followup) could give different results/distribution of the categories "much better," "slightly better," etc., than the transition query. Second, although the transition question was posed concerning the change of overall health status, one can argue that this remains a single criterion measure that may not assess the patient's outcome characteristics comprehensively enough. Therefore, the patient's self-assessment determined a minimally important change rather than a minimally important clinical criterion measure. However, the query asked about the change in health in general, implying a large variety of different health aspects and points of view, even if they are not listed explicitly. Third, due to small numbers (n < 10) within the transition categories, specific MCID for hip and for knee OA could not be determined. The same is true for the assessment of baseline dependency that had to be done in only 2 groups separated by the median value.

Assessment of MCID using the transition method is a heuristic and valid strategy to detect particular rehabilitation effects in patients with OA of the lower extremities with the use of the WOMAC, and it is worth implementing. The size of the MCID and of the systematic bias is comparable to that assessed by other methods and in other therapeutic settings.

ACKNOWLEDGMENT

We thank Stephan Mariacher, MD, and Susanne Lehmann for the planning, management, and implementation of the data base, and Robin Kyburg and Diane Fassett for help in preparing the manuscript.

REFERENCES

1. Dieppe P. Therapeutic targets in osteoarthritis. J Rheumatol 1995;22 Suppl 43:136-9.
2. Bellamy N. Outcome measurement in osteoarthritis clinical trials. J Rheumatol 1995;22 Suppl 43:49-51.

3. Lequesne M, Brandt K, Bellamy N, et al. Guidelines for testing slow acting drugs in osteoarthritis. J Rheumatol 1994;21 Suppl 41:65-73.
4. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Controlled Clin Trials 1991;12:142S-158S.
5. Lesquesne M, Méry C, Samson M, Gérard P. Indexes of severity for osteoarthritis of the hip and knee. Scand J Rheumatol 1987;65 Suppl:85-9.
6. Bellamy N. WOMAC Osteoarthritis Index. A user's guide. London, Canada: University of Western Ontario; 1995.
7. Bellamy N, Buchanan WW. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clin Rheumatol 1986;5:231-41.
8. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
9. Bellamy N. Pain assessment in osteoarthritis: experience with the WOMAC Osteoarthritis Index. Semin Arthritis Rheum 1989;18 Suppl 2:14-7.
10. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Duku E. Signal measurement strategies: are they feasible and do they offer any advantage in outcome measurement in osteoarthritis? Arthritis Rheum 1990;33:739-45.
11. Bellamy N, Kean WF, Buchanan WW, Gerecz-Simon E, Campbell J. Double blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): post validation reapplication of the WOMAC Osteoarthritis Index. J Rheumatol 1992;19:153-9.
12. Stucki G, Meier D, Stucki S, et al. Evaluation of a German version of the WOMAC (Western Ontario and McMaster Universities) Osteoarthritis Index. Z Rheumatol 1996;55:40-9.
13. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controlled Clin Trials 1989;10:407-15.
14. Hawker G, Melfi C, Paul J, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) instrument in the measurement of outcomes after knee replacement surgery. J Rheumatol 1995;22:1193-6.
15. Kantz ME, Harris WJ, Levitsky K, Ware JE, Davies AR. Methods for assessing condition-specific and generic functional status outcomes after total knee replacement. Med Care 1992;5:MS240-52.
16. Wright JG, Young NL. A comparison of different indices of responsiveness. J Clin Epidemiol 1997;50:239-46.
17. Ruof J, Sangha O, Stucki G. Comparative responsiveness of 3 functional indices in ankylosing spondylitis. Rheumatol 1999; 26:1959-63.
18. Laupacis A, Bourne R, Rorabeck C, et al. The effect of elective total hip replacement on health-related quality of life. J Bone Joint Surg Am 1993;75:1619-26.
19. Altman RD, Bloch DA, Bole GG Jr, et al. Development of criteria for osteoarthritis. J Rheumatol 1987;14 Suppl 14:3-6.
20. Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. J Clin Epidemiol 1995;48:1369-78.
21. Stucki G, Daltroy L, Liang MH, Lipson SJ, Fossel AH, Katz JN. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. Spine 1996;21:796-803.
22. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. Arch Intern Med 1993;153:1337-42.
23. Stucki G, Liang MH, Stucki S, Katz JN, Lew RA. Application of statistical graphics to facilitate selection of health status measures for clinical practice and evaluative research. Clin Rheumatol 1999;18:101-5.
24. Corzillius M, Fortin P, Stucki G. Responsiveness and sensitivity to change of SLE disease activity measures. Lupus 1999;8:655-9.
25. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities Osteoarthritis Index questionnaire and global assessments in patients with osteoarthritis. J Rheumatol 2000;27:2635-41.
26. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum 2000;43:1478-87.
27. Kazis ES, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27 Suppl:S178-89.
28. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: A clarification. J Clin Epidemiol 1989;42:403-8.
29. Weinstein ND. Optimistic biases about personal risk. Science 1989;246:1232-3.
30. Bindman AB, Keane D, Lurie N. Measuring health changes among severely ill patients. The floor phenomenon. Med Care 1990; 28:1142-51.