

ORIGINAL ARTICLE

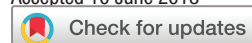
Multidimensional minimal clinically important differences in knee osteoarthritis after comprehensive rehabilitation: a prospective evaluation from the Bad Zurzach Osteoarthritis Study

Felix Angst,¹ Thomas Benz,¹ Susanne Lehmann,¹ André Aeschlimann,¹ Jules Angst²

To cite: Angst F, Benz T, Lehmann S, *et al.* Multidimensional minimal clinically important differences in knee osteoarthritis after comprehensive rehabilitation: a prospective evaluation from the Bad Zurzach Osteoarthritis Study. *RMD Open* 2018;**4**:e000685. doi:10.1136/rmdopen-2018-000685

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/10.1136/rmdopen-2018-000685>).

Received 21 March 2018
Revised 5 June 2018
Accepted 10 June 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Research Department, Rehabilitation Clinic ('RehaClinic'), Bad Zurzach, Switzerland

²Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric Hospital Burghölzli, University of Zurich, Zurich, Switzerland

Correspondence to
Dr Felix Angst;
fangst@vtxmail.ch

ABSTRACT

Objective To determine minimal clinically important differences (MCIDs) for improvement and worsening in various health dimensions in knee osteoarthritis under conservative therapy.

Methods Health, symptoms and function were assessed by the generic Short Form 36 and the condition-specific Western Ontario and McMaster Universities Osteoarthritis Index in n=190 patients with knee osteoarthritis before and after comprehensive rehabilitation intervention (3-month follow-up). By means of construct-specific transition questions, MCIDs were defined as the difference between the 'slightly better/worse' and the 'almost equal' transition response categories according to the 'mean change method'. The bivariate MCIDs were adjusted for sex, age and baseline score to obtain adjusted MCIDs by multivariate linear regression. They were further standardised as (baseline) effect sizes (ESs), standardised response means (SRMs) and standardised mean differences (SMDs) and compared with the minimal detectable change with 95% confidence (MDC95).

Results Multivariate, adjusted MCIDs for improvement ranged from 2.89 to 16.24 score points (scale 0–100), corresponding to ES=0.14 to 0.63, SRM=0.17 to 0.61 and SMD=0.18 to 0.72. The matching results for worsening were –5.80 to –12.68 score points, ES=–0.30 to –0.56, SRM=–0.35 to –0.52 and SMD=–0.35 to –0.58. Almost all MCIDs were larger than the corresponding MDC95s.

Conclusions This study presents MCIDs quantified according to different methods over a comprehensive range of health dimensions. In most health dimensions, multivariate adjustment led to higher symmetry between the MCID levels of improvement and worsening. MCIDs expressed as standardised effect sizes (ES, SRM, SMD) and adjusted by potential confounders facilitate generalisation to the results of other studies.

Key messages

What is already known about this subject?

- Changes in health or symptoms should not only be detectable by statistical significance tests, they have also to be perceived by the person affected. This led to the development of minimal clinically important differences (MCIDs).
- In knee osteoarthritis, there are no data on MCIDs beyond pain and general physical function.

What does this study add?

- The present study extended the determination of MCIDs to overall and leg-specific function, standing/walking, mobility/stiffness, physical role performance, activity/vitality, social functioning, affective/mental health, and general health perception in knee osteoarthritis.
- The MCIDs were quantified by three different effect sizes and adjusted by confounders.

How might this impact on clinical practice?

- Multidimensional, generalisable MCIDs enable us to rate the clinical impact of measured effects in any study of knee osteoarthritis, with effect sizes specifically adjusted according to the type of study design.

INTRODUCTION

Minimal clinically important differences (MCIDs) play an increasingly important role in evidence-based medical practice and outcome measurement.^{1–5} Changes in health or symptoms should not only be detected by statistical significance tests; they also have to be perceived by the person affected. The patient's

perspective of health is integral to understanding health outcomes.^{1 3 5} The subjective perception of outcome effects is the key element of the MCID concept.

In clinical trials, every outcome difference becomes statistically significant provided the sample size is large enough, as has been demonstrated.⁵ However, we can safely assume that, despite reaching statistical significance, a very small effect difference in a very large sample will not be subjectively perceptible to the person affected, that is, is not 'clinically important'. In contrast to the concept of statistical effect significance, an effect measured that is larger or equal to the MCID indicates that the patients in that setting subjectively perceive their improvements as beneficial.

It has become accepted practice to use 'anchor'-based estimates to determine the MCID because the patient's viewpoint is the key characteristic and predictor for patient-rated outcome measures.^{1 3 5 6} Anchor-based methods use an external indicator, the 'transition item', to assess changes in health status; the transition item asks patients to rate any change in their health between baseline and a specific follow-up point.^{1 3 5 7 8} Today, the most important and most frequently used anchor-based method is the 'mean change method' originated by Redelmeier and Lorig in 1993.⁸ The MCID for improvement, for example, equals the mean of an instrument's score difference between baseline and follow-up (eg, of pain) of the 'slightly better' transition response group minus that of the 'almost equal' group results.

While a number of studies have been published assessing MCIDs in knee osteoarthritis (for example: 3 9–13), none, to our knowledge, has examined MCIDs in health dimensions beyond pain and function in general. The present study seeks to fill this gap by extending the determination of MCIDs to other, specific functional abilities and psychosocial domains. The evaluated MCIDs will provide future studies dealing with therapy effects in knee osteoarthritis with a basis for comparison of their measured effects. Furthermore, this is the first report on the application of our recently proposed methodology using multivariate adjustment by potential confounders to minimise bias in the estimated MCID. This method maximises the generalisability of the estimated MCID levels to other testing settings.⁵

Using construct-specific transition questions (anchors), this evaluation study aimed to determine the MCIDs for improvement and worsening in patients with knee osteoarthritis in the following dimensions: generic and condition-specific pain, overall function, standing and walking, mobility/stiffness, physical role performance, activity/vitality, social functioning, affective/mental health and general health perception.

METHODS

Patients and data collection

The data derive from the 'Zurzach Osteoarthritis Study', an observational, prospective cohort study, which examined

Table 1 Sociodemographic and disease-relevant data (n=190)

Characteristic	n	%
Female	149	78.4
Education (%)		
Basic school (8–9 years)	69	36.3
Vocational training	92	48.4
College/high school/university	29	15.3
Comorbidities (%)		
None	1	0.5
1	12	6.3
2	35	18.4
3	54	28.4
4	35	18.4
≥5	53	27.9
Age (years)	Mean: 66.1	SD: 10.2

health and quality of life before and after comprehensive rehabilitation for hip and knee osteoarthritis.¹⁴ All patients fulfilled the American College of Rheumatology criteria for knee osteoarthritis and signed written informed consent to participate.¹⁵ Clinical outcome data were earlier reported for controlled short-term effects in knee patients.^{14 16}

The rehabilitation intervention consisted chiefly of active therapeutic entities (individual and group, land-based and water-based physiotherapy) and passive modalities (massage, packs, educational measures, coping instructions), which have been previously described in detail.¹⁴ For this study, all patients with knee osteoarthritis treated in hospital (2–3 weeks' inpatient stay) and ambulatory (6–9 weeks' outpatient sessions) settings were included. Data were collected at the start of the intervention (baseline) and at follow-up 3 months later, when all therapies had been completed (n=190).

Measures

MCIDs were determined on the basis of two instruments: the well-known and best-tested generic Medical Outcomes Study 36-Item Short Form Health Survey (SF-36), version 1, and the condition-specific Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC).^{17–22} In the medical literature, the SF-36 is the most widely used and most important patient-rated outcome instrument for healthy people and any health-affecting condition.^{6 10–14} The WOMAC is the most frequently used specific tool for hip and knee.^{6 9–14} Both instruments require license payments. While the manual of the SF-36 can be obtained for unlimited use for approximately US\$300 (as can the German version and manual), the use of the WOMAC is limited by the number of applications (persons and follow-ups).^{17 19} However, all the WOMAC items are included in the Knee injury and Osteoarthritis Outcome Score (KOOS), which is available free of charge, and the German WOMAC version 1 used in this study is in the public domain.^{6 20}

The SF-36 consists of 36 items distributed over eight scales, four physical and four psychosocial. The WOMAC is a 24-item instrument with three subscales: five items for pain, two for stiffness and 17 for function. By Rasch analysis, four functional factor subdimensions can be further defined.^{21 22} Of those, the WOMAC standing and walking factor, was included in the present study. The complex rationale for the construction and validation of this factor dimension has been published before.^{20–22}

Table 2 summarises the 11 construct domains of the SF-36 and WOMAC scales used in our analysis and the corresponding transition item anchors, with their specific contents. Seven of the eight scales of the SF-36 were included (role emotional was omitted) together with the three original WOMAC subscales and the additional WOMAC factor standing/walking.

The WOMAC factor standing/walking comprises one pain and one function item for standing together with one pain and one function item for walking on flat surface, that is, four items in total. SF-36 role physical consists of four items: limits in the amount of time spent on work or other activities, accomplished less than would have liked, limits in the kind of work or other activities, and difficulties performing work or other activities. SF-36 social functioning has two items: extent and time of interference of physical or emotional problems with normal social activities.

ANALYSIS

All analyses were performed using the statistical software package IBM SPSS V.23.0 for Windows. For all steps described below, an exemplary calculation is outlined in the online supplementary appendix 1.

Descriptive data

Descriptive data at baseline and score differences to 3-month follow-up were compiled using arithmetic means and metric SD. For ease of comparison, all scores were scaled from 0=worst (most pain, no function, worst health) to 100=best (no pain, best function, best health).^{14 17}

Assessment of appropriateness for determining the MCID

As a preliminary analysis to test a scale's appropriateness for calculation of the MCID, floor and ceiling proportions were determined for each scale, with the floor being the percentage of subjects with the minimum score=0, or worst health, and the ceiling the percentage with the maximum score=100, or best health. A high floor at baseline means that many patients can experience no further deterioration to the follow-up. Vice versa, a high ceiling phenomenon at baseline indicates that many patients cannot further improve. Both the floor and the ceiling affect the ability to calculate changes between baseline and follow-up on a closed scale (0–100) and, consequently, also the quantification of valid MCIDs.

In addition, Spearman rank correlation coefficients between the five transition ratings (much worse, slightly worse, almost equal, slightly better, much better) and the score differences (baseline to follow-up) were calculated.

'Revicki's criterion' was applied^{1 13}; that is, a proposed correlation of ≥ 0.30 to prove a sufficient level of construct convergence between the scale and the anchor.

Specific anchors

The content of the transition item questions was construct specific for all scales (see table 2). For the constructs of function and role physical (role limitations due to physical health), two transition items were combined (table 2): walking plus dressing and personal hygiene; household chores plus work. At least one of the items had to be rated 'slightly better' for the combined item to score 'slightly better'. The same applied to the rating 'slightly worse'. However, the response to both items had to be 'almost equal' for the rating 'almost equal' to be attributed.

Bivariate, unadjusted MCIDs

Bivariate, unadjusted MCIDs were calculated according to the 'mean change method'.^{5 8} The mean score change or difference of the transition category 'slightly better' minus that of the transition category 'almost equal' defined the MCID for improvement. By analogy, the mean score change and difference of the transition category 'slightly worse' minus that of the transition category 'almost equal' defined the MCID for worsening. A positive MCID reflects improvement, a negative MCID worsening.

The interval for 95% confidence of the MCID was calculated by $\pm t \times \text{square root of } (1/n_1 + 1/n_2) \times \text{pooled SD}$.²³ The pooled SD is equal to $\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$, where the 'slightly better' group had n_1 =number of patients and s_1 =SD of the score changes (baseline to follow-up), and the 'almost equal group' had correspondingly n_2 and s_2 . The multiplier 'square root of $(1/n_1 + 1/n_2)$ ' gives the SE of the MCID. The t-value comes from Student's t distribution with two-sided type I error of 0.05 and $df = n_1 + n_2 - 2$ and can be calculated on the internet.²⁴

Multivariate, adjusted MCIDs

In order to obtain a less biased, more generalisable estimate of the MCID, multivariate, confounder-adjusted MCIDs were calculated using the regression model as follows⁵: score change (baseline to follow-up; in score points) = $b_1 \times \text{transition group (binary coded: 1='slightly better', 0='almost equal')}$ + $b_2 \times \text{baseline score (in score points)}$ + $b_3 \times \text{sex (m/f)}$ + $b_4 \times \text{age (in years)}$ + constant term. The coefficient b_1 is equal to the adjusted MCID and the SE of b_1 is the SE of the MCID. This adjusted, multivariable MCID can also be divided by the specific SD to obtain adjusted effect sizes as described above. The interval for 95% confidence can be calculated by $\pm t \times \text{SE of the MCID}$.^{23 24}

Standardisation of MCIDs by effect sizes

All MCIDs were further standardised, that is, divided by specific SD to obtain a parameter belonging to the family of 'effect sizes'.^{5 23} This results in a dimensionless parameter that is less biased, especially by different baseline

Table 2 Construct domains of the scales and their relation to the transition items, and floor, ceiling and reliability data (n=190)

Instrument	Scale	Items	Domain/transition item content	Floor %	Ceiling %	Correlation	ICC
WOMAC	Pain	5	Pain	–	–	0.43	0.70
SF-36	Bodily pain	2	Pain	–	0.5	0.42	0.72
WOMAC	Function	17	Walking plus dressing/personal hygiene	–	–	0.49	0.74
SF-36	Physical functioning	10	Walking plus dressing/personal hygiene	–	–	0.36	0.86
WOMAC	Factor standing/walking	4	Walking	–	1.1	0.39	0.79
WOMAC	Stiffness	2	Joint mobility	–	3.7	0.37	0.55
SF-36	Role physical	4	Household chores plus work	65.8	5.4	0.26	0.44
SF-36	Vitality	4	Fatigue	–	1.1	0.18	0.03
SF-36	Social functioning	2	Social activities (performance)	–	23.7	0.20	0.80
SF-36	Mental health	5	Mood/mental health	–	1.1	0.31	0.70
SF-36	General health	6	Health in general	–	0.5	0.35	0.88

Ceiling, proportion of patients with best possible score (%) at baseline; correlation, Spearman rank correlation between the score difference (baseline to follow-up) and the transition item; floor, proportion of patients with worst possible score (%) at baseline; ICC, intraclass correlation coefficient for test–retest reliability out of the literature^{20 21 27}; SF-36, Short Form 36; WOMAC, Western Ontario and McMaster Universities Index.

scores, and is more generalisable to the findings of other studies.⁵ A positive effect size (ES) reflects improvement, a negative ES a worsening of health.

The following three most important and frequently used parameters were calculated: (1) the (baseline) ES according to Kazis *et al*, which is the crude MCID divided by the baseline SD of the whole sample.²⁵ (2) The standardised response mean (SRM) according to Liang, which is the division of the crude MCID by the SD of the score differences of the whole sample.²⁶ (3) The SMD according to Borenstein, which is the crude MCID divided by the pooled SD of the ‘slightly better’ transition category and the ‘almost equal’ transition category (see above) and multiplied by a correction factor J to reduce bias for small sample sizes.²³

MDC95 for comparison with MCID

For each scale, the intraclass correlation coefficients (ICCs) were sourced from the literature.^{20 21 27} The ICC measures test–retest reliability and is needed to determine the minimum detectable changes with 95% confidence (MDC95). The MDC95 reflects the measurement error of two different measurements on the same scale.²⁵ If an MCID is larger than the MDC95, it reflects a difference (between baseline and follow-up) larger than a difference that may occur due to an error of measurement with 95% confidence. The MDC95 is calculated by the t-values multiplied by the SE of measurement (SEM), which is equal to the SE of the MCID multiplied by the square root of (1–ICC).²⁵ The MDC95 is a CI which applies to single parameter estimates (the MCID) but not to CIs of a parameter.²⁵

RESULTS

Setting and descriptive data

The sample consisted of n=190 patients with knee osteoarthritis with complete baseline and follow-up data (tables 1–3). Of those, 149 (78.1%) were women. The mean age was 66.1 years and the SD 10.2 years (table 1).

The construct overlap (correlation) was ≥ 0.30 for all scales, with the exception of SF-36 role physical, vitality and social functioning, which failed the ‘Revicki’s criterion’ (table 2).¹ However, those three scales were kept in the analysis for the sake of completeness; they are marked in lighter font. The scores at baseline (see also table 3), showed a very high floor rate for SF-36 role physical (65.8%) reflecting n=125/190 subjects with score 0 (worst health); no other scale had a floor phenomenon (table 2). Ceiling (ie, 100=best health) percentage scores were as follows: 23.7% (n=45) for SF-36 social functioning, 5.4% (n=10) for role physical and 3.7% (n=7) for WOMAC stiffness, but were very low or absent on all the other scales. The test–retest ICCs were found in three studies of hip and knee osteoarthritis.^{20 21 27}

Baseline and follow-up scores for the whole sample (n=190) and within the three transition groups relevant for the MCID are shown in table 3. All baseline

Table 3 Descriptive score data before (baseline) and 3 months after rehabilitation (follow-up) of knee osteoarthritis (n=190)															
All participants (n=190)						Transition group 'slightly better'			Transition group 'almost equal'			Transition group 'slightly worse'			
Baseline		Change to follow-up			Change to follow-up			Change to follow-up			Change to follow-up				
m	s	m	s	n	m	s	n	m	s	n	m	s	n	m	s
WOMAC	Pain	50.93	21.48	7.60	21.13	49	13.51	21.58	62	4.77	15.62	44	-1.32	20.81	
SF-36	Bodily pain	24.84	16.34	12.83	21.85	46	18.72	22.71	62	8.31	14.02	44	4.05	19.05	
WOMAC	Function	50.99	20.45	4.39	18.61	45	14.88	16.82	65	0.40	14.61	37	-5.87	18.87	
SF-36	Physical functioning	33.46	19.30	3.91	20.77	52	7.17	20.16	65	2.94	15.50	43	-2.44	22.42	
WOMAC	Factor standing/ walking	53.70	22.90	6.34	24.55	40	15.75	24.67	77	5.55	21.59	39	-5.06	22.61	
WOMAC	Stiffness	49.53	25.85	1.71	26.57	48	15.52	22.37	72	-4.72	22.47	43	-10.93	29.08	
SF-36	Role physical	16.58	27.97	10.46	39.32	46	19.02	43.20	68	9.56	38.88	53	1.42	27.91	
SF-36	Vitality	41.03	20.69	3.63	16.45	45	6.44	16.98	89	2.28	15.14	29	1.72	13.58	
SF-36	Social functioning	62.90	29.90	2.64	25.22	37	8.43	30.54	94	1.47	24.65	35	-3.14	22.75	
SF-36	Mental health	63.46	20.52	2.06	16.91	65	5.91	16.91	47	0.68	14.93	41	-3.39	15.19	
SF-36	General health	53.49	19.63	-1.80	16.49	42	3.93	15.28	72	-2.07	17.51	37	-8.59	13.71	

All means are indicated in bold.
m, arithmetic mean (bold); positive values reflect improvement, negative worsening); n, number of subjects; s, SD; SF-36, Short Form 36; WOMAC, Western Ontario and McMaster Universities Index.

score levels were around the middle of the scale 0–100, except SF-36 bodily pain, physical functioning and role physical (16.58), reflecting relatively great pain and low physical function and role (table 2). This means that changes between the baseline and follow-up were detectable in both directions, that is, improvements (positive differences) and worsening (negative differences). The mean score at the 3-month follow-up can be calculated by adding the change to the baseline score, example SF-36 general health: $53.49 + (-1.80) = 51.69$.

Consistent with the anchor rating, all changes showed improvements in the ‘slightly better’ category. In the ‘almost equal’ category, on average most changes showed improvements (maximum on SF-36 bodily pain, mean = +8.31), with the exception of WOMAC stiffness and SF-36 general health (table 3). In the ‘slightly worse’ group, 8/11 scales showed worsening, except SF-36 bodily pain, role physical and vitality.

Minimal clinically important differences

In the domain of pain, MCIDs for improvement ranged from 7.09 to 10.41 score points and those for worsening from –4.26 to –7.07; symmetry (improvement and worsening) was higher in the adjusted parameters: 7.09 and 8.19 versus –7.07 and –6.08 (table 4). In the domain of function, WOMAC function showed relatively large MCIDs for improvement (14.48 bivariate and 11.25 multivariate) compared with SF-36 physical functioning (4.23 and 3.81), whereas the levels for worsening were comparable (–5.38 to –7.29). Consistent with these findings, the MCIDs for the WOMAC factor standing/walking were large (improvement: 10.20, 5.93; worsening: –10.62, –12.87). WOMAC stiffness showed very large MCIDs for improvement (20.24 bivariate, 16.24 multivariate) and lower levels for worsening (–6.21, –9.91). Moderate and comparable MCIDs were found for SF-36 mental health (improvement: 5.23, 2.89 vs worsening: –4.07, –8.94) and SF-36 general health (6.00, 7.15 vs –6.53, –5.80).

The three scales with low correlations (<0.30) with the transition item or having high floor/ceiling effects, namely SF-36 role physical, SF-36 vitality and SF-36 social functioning, showed partly lopsided results (eg, SF-36 social functioning, multivariate: +2.57 for improvement, –18.50 for worsening). Moreover, most of those three MCIDs were lower than the corresponding MCD95.

The MCIDs of all the other scales were mostly much higher than the corresponding MCD95, for example, WOMAC function improvement, bivariate: 14.48 versus 3.05; the very few exceptions included the bivariate MCID for worsening of WOMAC stiffness (table 4).

Expressed as ES, SRM and SMD, 12 of 21 of the bivariate and 8/21 of the multivariate MCIDs for improvement were in the range of small effect levels of 0.30–0.50 (table 4). This was also the case for 10/21 (bivariate) and 13/21 (multivariate) MCIDs for worsening. The highest effect sizes were for improvement in WOMAC function (bivariate up to SMD 0.93).

DISCUSSION

This study presents the MCIDs for improvement and worsening in various health dimensions in patients with knee osteoarthritis. The MCIDs were quantified by the classical mean change method (unadjusted, bivariate) and adjusted for confounders by multivariate linear regression modelling. The MCIDs were expressed as raw score changes/differences in score points and in the number of SD by effect sizes, namely the (baseline) ES according to Kazis (applicable to pilot studies, which have only baseline data), the SRM according to Liang (for longitudinal cohort studies without a control group) and the SMD according to Borenstein (for randomised controlled trials).^{5 23 25 26} Valid MCID estimates could be determined for pain, function, standing/walking, stiffness, mental health and general health. All multivariate MCIDs and almost all bivariate MCIDs were higher than the corresponding MCD95, that is, beyond the instruments’ error of measurement with 95% confidence.²⁵

The adjusted, multivariate MCIDs were more evenly balanced between improvement and worsening than the crude, bivariate levels (eg, WOMAC pain: +7.09/–7.07 vs +8.74/–6.09 score points; SF-36 bodily pain: +8.19/–6.08 vs +10.41/–4.26). The same is true for WOMAC function and WOMAC stiffness but not for the other scales. However, there was little difference between the multivariate and bivariate MCIDs in most cases.

In other words, the bivariate is a good estimate of the multivariate MCID. Nevertheless, the adjusted, multivariate MCID is to be preferred for application to the results of other studies.^{5 28 29} The variables sex, age and baseline score substantially confound health changes, but since they are available in every study setting, adjustment can be made for them. Adjustment provides more valid and generalisable MCIDs, which can be transferred and applied to different settings and patient groups for the purpose of assessing the clinical relevance of the outcome effects measured. If an effect measured in a specific study is larger or equal to the MCID, this means that the patients in that setting subjectively perceive their improvements as beneficial both on average and at group level.

Our MCIDs are in line with those of a recent multidisciplinary outpatient programme for patients with knee osteoarthritis (OA) (table 5).¹³ In that study, the MCIDs for improvement on the function scale of the KOOS, which is identical to the WOMAC function scale, were 8.93 score points for the transition question regarding ‘walking on level ground’ and 7.64 for the transition question regarding ‘my knee in general’. The corresponding MCIDs for worsening were –6.57 and –4.00. MCIDs for patients with knee OA after 6 weeks’ treatment with non-specific, non-steroidal anti-inflammatory drugs were calculated for the WOMAC total score (sum of 5 pain, 2 stiffness, 17 function items) and resulted in an MCID for improvement of 11 and for worsening of –16 score points.¹²

Table 4 Bivariate (unadjusted) and multivariate (adjusted) MCIDs

		Improvement					Worsening							
		MDC95	MCID	95% CI	ES	SRM	SMD	MDC95	MCID	95% CI	ES	SRM	SMD	
WOMAC	Pain	b	3.83	8.74	1.73 to 15.74	0.41	0.41	0.47	-3.84	-6.09	-13.11 to 0.92	-0.28	-0.29	-0.34
		m	3.37	7.09	0.93 to 13.25	0.33	0.34	0.38	-3.62	-7.07	-13.67 to -0.47	-0.33	-0.33	-0.39
SF-36	Bodily pain	b	3.72	10.41	3.38 to 17.44	0.64	0.48	0.57	-3.37	-4.26	-10.63 to 2.11	-0.26	-0.20	-0.26
		m	3.31	8.19	1.93 to 14.44	0.50	0.37	0.45	-3.18	-6.08	-12.09 to -0.07	-0.37	-0.28	-0.37
WOMAC	Function	b	3.05	14.48	8.51 to 20.46	0.71	0.78	0.93	-3.39	-6.27	-12.91 to 0.38	-0.31	-0.34	-0.38
		m	3.10	11.25	5.17 to 17.32	0.55	0.60	0.72	-3.34	-7.29	-13.84 to -0.75	-0.36	-0.39	-0.44
SF-36	Physical functioning	b	2.44	4.23	-2.30 to 10.77	0.22	0.20	0.24	-2.71	-5.38	-12.61 to 1.85	-0.28	-0.26	-0.29
		m	2.26	3.81	-2.23 to 9.84	0.20	0.18	0.21	-2.46	-6.64	-13.21 to -0.08	-0.34	-0.32	-0.36
WOMAC	F.Standing/walking	b	4.01	10.20	1.44 to 18.95	0.45	0.42	0.45	-3.91	-10.62	-19.16 to -2.08	-0.46	-0.43	-0.48
		m	3.53	5.93	-1.76 to 13.63	0.26	0.24	0.26	-3.51	-12.87	-20.53 to -5.22	-0.56	-0.52	-0.58
WOMAC	Stiffness	b	5.55	20.24	11.97 to 28.52	0.78	0.76	0.90	-6.44	-6.21	-15.80 to 3.39	-0.24	-0.23	-0.25
		m	4.76	16.24	9.15 to 23.34	0.63	0.61	0.72	-5.21	-9.91	-17.67 to -2.15	-0.38	-0.37	-0.39
SF-36	Role physical	b	11.51	9.46	-5.92 to 24.85	0.34	0.24	0.23	-9.37	-8.14	-20.67 to 4.38	-0.29	-0.21	-0.23
		m	10.85	7.47	-7.03 to 21.96	0.27	0.19	0.18	-9.61	-12.68	-25.53 to 0.17	-0.45	-0.32	-0.36
SF-36	Vitality	b	5.62	4.16	-1.55 to 9.87	0.20	0.25	0.26	-6.16	-0.56	-6.82 to 5.70	-0.03	-0.03	-0.04
		m	5.05	4.91	-0.21 to 10.03	0.24	0.30	0.31	-5.60	-2.25	-7.94 to 3.44	-0.11	-0.14	-0.15
SF-36	Social functioning	b	4.54	6.96	-3.18 to 17.11	0.23	0.28	0.26	-4.23	-4.61	-14.07 to 4.85	-0.15	-0.18	-0.19
		m	3.68	2.57	-5.66 to 10.80	0.09	0.10	0.10	-3.74	-18.50	-26.85 to -10.15	-0.62	-0.73	-0.76
SF-36	Mental health	b	3.35	5.23	-0.89 to 11.34	0.25	0.31	0.32	-3.50	-4.07	-10.47 to 2.32	-0.20	-0.24	-0.27
		m	2.85	2.89	-2.32 to 8.10	0.14	0.17	0.18	-3.22	-8.94	-14.82 to -3.06	-0.44	-0.53	-0.59
SF-36	General health	b	2.23	6.00	-0.44 to 12.43	0.31	0.36	0.36	-2.27	-6.53	-13.07 to 0.02	-0.33	-0.40	-0.40
		m	1.95	7.15	1.51 to 12.78	0.36	0.43	0.42	-1.94	-5.80	-11.41 to -0.19	-0.30	-0.35	-0.35

All MCIDs are indicated in bold.

b, bivariate; CI, 95% CI of the MCID; ES, effect size according to Kazis; F, factor; m, multivariate; MCID, minimal clinically important difference (bold; positive values reflect improvement, negative worsening); MDC95, minimal detectable difference with 95% confidence (two-sided) based on the test-retest reliability of the scale and the based on the SE of measurement; SF-36, Short Form 36; SMD, standardised mean difference; SRM, standardised response mean according to Liang; WOMAC, Western Ontario and McMaster Universities Index.

Table 5 Unadjusted MCIDs of comparative literature

		Improvement			Worsening		
		Present	Escobar <i>et al</i> 2007 ¹⁰	Mills <i>et al</i> 2016 ¹³	Present	Escobar <i>et al</i> 2007 ¹⁰	Mills <i>et al</i> 2016 ¹³
WOMAC	Pain	8.74	10.77	5.56	-6.09	-4.39	-8.08
SF-36	Bodily pain	10.41	9.33	–	-4.26	-16.00	–
WOMAC	Function	14.48	9.55	7.64	-6.27	-9.19	-4.00
SF-36	Physical functioning	4.23	7.93	–	-5.38	-8.86	–
WOMAC	F.Standing/walking	10.20	–	8.93	-10.62	–	-6.57
WOMAC	Stiffness	20.24	7.11	–	-6.21	-10.71	–
SF-36	Role physical	9.46	9.39	–	-8.14	-11.92	–
SF-36	Vitality	4.16	5.35	–	-0.56	-10.91	–
SF-36	Social functioning	6.96	11.66	–	-4.61	-12.50	–
SF-36	Mental health	5.23	0.12	–	-4.07	-16.38	–
SF-36	General health	6.00	1.73	–	-6.53	-9.94	–

F., factor; present, results of the present study; SF-36, Short Form 36; WOMAC, Western Ontario and McMaster Universities Index.

In an earlier study, detailed MCIDs for the SF-36 and WOMAC were calculated 6 months after patients had undergone total knee replacement surgery. They showed score point differences comparable with the bivariate MCIDs of our study on most scales, ranging from 0.12 to 11.66 for improvement and from -4.39 to 10.62 for worsening (table 5).¹⁰

On half of the scales, our MCIDs expressed as ESs were in the range of 0.30 to 0.50, a generalisable range according to the literature.¹ However, it seems that the MCID increased with the responsiveness of the scale. In other words, a highly responsive/sensitive scale measured relatively large differences between the baseline and follow-up until the change reached the level subjectively measurable by the transition rating. This was especially true for WOMAC function improvement, which is more responsive than SF-36 physical functioning (multivariate, adjusted SMD=+0.72 vs +0.21).^{9 30 31} Furthermore, the threshold at which improvement in stiffness became subjectively perceptible was high, and much higher than that for deterioration (multivariate SMD=+0.72 vs -0.39). Finally, subjectively perceived changes of rehabilitation in mental and general health were as expected small, reflected in the relatively small MCIDs.¹⁴

The MCIDs of SF-36 role physical, vitality and social functioning are presented but for various reasons do not appear to provide valid estimates: low construct convergence to the anchor (correlations<0.30), high floor or high ceiling effects, and MCID <MDC95 (respectively >for worsening).^{1 5} MCID data of those scales were presented for completeness, interest and to give an idea of their levels, but caution should be used in applying them for comparison with the effect data of other studies.

One strength of our study is that it is the first to examine MCIDs in health dimensions beyond pain

and general function in patients with knee OA. The anchor rating was not global but construct specific, which improved the specificity of the MCIDs. Construct convergence to the anchor rating and floor/ceiling effects were taken into account to rate the validity of the MCID estimate for each scale. All MCIDs were related to the corresponding MCD95s. Quantification of the MCID by ESs improves their generalisability and the application of the results to testing studies, especially by the SMD for randomised controlled trials.⁵ Multivariate adjustment further improved the generalisability to enable comparison with other study results.^{5 28 29}

The most important limitation is that the setting examined conservative therapy of knee osteoarthritis and it may not be possible to generalise the MCID to knee surgery.

Acknowledgements This study was supported by the Zurzach Rehabilitation Foundation SPA, Bad Zurzach, Switzerland. The authors thank Elizabeth Kyrke for the linguistic editing of the paper.

Contributors FA planned the study, analysed and interpreted the data, and wrote the draft study report. TB, as a clinical expert, helped in the literature search and contributed to interpreting the data and to putting them into a clinical context. SL helped to plan and to carry out the study, collected the data and helped to interpret them. AA provided the resources for carrying out the study and helped in planning the study and in interpreting the data. JA helped to interpret the data, to put them into a clinical context and to finalise the study report.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent Obtained.

Ethics approval EK AG 2008/026, Switzerland.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is

properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Revicki D, Hays RD, Cella D, *et al.* Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102–9.
2. Kamath CC, Sloan JA, Cappelleri C. Clinical significance. In: N Balakrishnan, *Methods and applications of statistics in clinical trials: concepts principles, trials, and designs*. Hoboken, New Jersey, USA: John Wiley & sons Inc, 2014:Vol 1. 170–90.
3. Bellamy N, Hochberg M, Tubach F, *et al.* Development of multinational definitions of minimal clinically important improvement and patient acceptable symptomatic state in osteoarthritis. *Arthritis Care Res* 2015;67:972–80.
4. Angst F. MCID - The minimal clinically important difference assigns significance to outcome effects. *J Rheumatol* 2016;43:258–9.
5. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol* 2017;82:128–36.
6. Katz PP. Introduction to special issue: patient outcomes in rheumatology, 2011. *Arthritis Care Res* 2011;63:S1–S3.
7. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
8. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. *Arch Intern Med* 1993;153:1337–42.
9. Angst F, Aeschlimann A, Michel BA, *et al.* Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. *J Rheumatol* 2002;29:131–8.
10. Escobar A, Quintana JM, Bilbao A, *et al.* Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage* 2007;15:273–80.
11. de Vet HC, Terluin B, Knol DL, *et al.* Three ways to quantify uncertainty in individually applied "minimally important change" values. *J Clin Epidemiol* 2010;63:37–45.
12. Hmamouchi I, Allali F, Tahiri L, *et al.* Clinically important improvement in the WOMAC and predictor factors for response to non-specific non-steroidal anti-inflammatory drugs in osteoarthritic patients: a prospective study. *BMC Res Notes* 2012;5:58.
13. Mills KA, Naylor JM, Eyles JP, *et al.* Examining the minimal important difference of patient-reported outcome measures for individuals with knee osteoarthritis: a model using the Knee Injury and Osteoarthritis Outcome Score. *J Rheumatol* 2016;43:395–404.
14. Angst F, Verra ML, Lehmann S, *et al.* Effects of inpatient rehabilitation in hip and knee osteoarthritis: a naturalistic prospective cohort study with intraindividual control of effects. *Arch Phys Med Rehabil* 2013;94:2139–45.
15. Altman R, Asch E, Bloch D, *et al.* Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. *Arthritis & Rheumatism* 1986;29:1039–49.
16. Weigl M, Angst F, Stucki G, *et al.* Inpatient rehabilitation for hip or knee osteoarthritis: 2 year follow up study. *Ann Rheum Dis* 2004;63:360–8.
17. Ware JE, Snow KK, Kosinski M. *SF-36 Health Survey: manual and interpretation guide*. 3rd edition. USA: Lincoln, RI, 2004.
18. Bullinger M, Kirchberger I. *SF-36 Fragebogen zum Gesundheitszustand. Handanweisung. (The SF-36 questionnaire to assess health status. Manual)*. Germany, Hogrefe: Göttingen, 1998.
19. Bellamy N. *WOMAC osteoarthritis index. A user's guide*. London, Ontario, Canada: University of Western Ontario, 1995.
20. Stucki G, Meier D, Stucki S. Evaluation einer deutschen Version des WOMAC (Western Ontario and McMaster Universities) Arthroseindex (Evaluation of a German version of the WOMAC). *Z Rheumatol* 1996;55:40–9.
21. Ryser L, Wright BD, Aeschlimann A, *et al.* A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. *Arthritis Care Res* 1999;12:331–5.
22. Angst F, Ewert T, Lehmann S, *et al.* The factor subdimensions of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) help to specify hip and knee osteoarthritis. a prospective evaluation and validation study. *J Rheumatol* 2005;32:1324–30.
23. Borenstein M. Effect sizes for continuous data. In: Cooper H, Hedges LV, Valentine JC, *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation. 2nd ed. New York, 2009:12. 222–36.
24. Internet t-value calculator (input: 0.05, two-sided and degrees of freedom). <http://www.danielsoper.com/statcalc/calculator.aspx?id=10>
25. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;27(3 Suppl):S178–S189.
26. Streiner DL, Norman GR, Cairney J. Reliability: issues of interpretation. In: Streiner DL, Norman GR, Cairney J, *Health measurement scales. A practical guide to their development and use*. 5th ed.. Oxford University press: Oxford, 2015:180–2.
27. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990;28:632–42.
28. Darvey RC, Edwards SM, Cochrane T. Test-retest reliability of lower extremity functional and self-reported measures in elderly with osteoarthritis. *Adv Physiother* 2003;5:155–60.
29. Louie GH, Ward MM. Sex disparities in self-reported physical functioning: true differences, reporting bias, or incomplete adjustment for confounding? *J Am Geriatr Soc* 2010;58:1117–22.
30. Hedderich J, Sachs L. Statistische Modellbildung (Statistical modelling). In: Hedderich J, Sachs L, eds. *Angewandte Statistik (Applied statistics)*. 15th ed. Berlin, Heidelberg, Germany: Springer, 2016: 8. 758–860.
31. Angst F, Aeschlimann A, Steiner W, *et al.* Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. *Ann Rheum Dis* 2001;60:834–40.
32. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Rheum* 2001;45:384–91.