What is an acceptable outcome of treatment before it begins? Methodological considerations and implications for patients with chronic low back pain

Henrik Hein Lauridsen · Claus Manniche ·
Lars Korsholm · Niels Grunnet-Nilsson ·
Jan Hartvigsen

Received: 20 August 2008 / Revised: 27 April 2009 / Accepted: 1 June 2009 / Published online: 23 June 2009
© Springer-Verlag 2009

Abstract Understanding changes in patient-reported outcomes is indispensable for interpretation of results from clinical studies. As a consequence the term “minimal clinically important difference” (MCID) was coined in the late 1980s to ease classification of patients into improved, not changed or deteriorated. Several methodological categories have been developed determining the MCID, however, all are subject to weaknesses or biases reducing the validity of the reported MCID. The objective of this study was to determine the reproducibility and validity of a novel method for estimating low back pain (LBP) patients’ view of an acceptable change (MCID_{pre}) before treatment begins. One-hundred and forty-seven patients with chronic LBP were recruited from an out-patient hospital back pain unit and followed over an 8-week period. Original and modified versions of the Oswestry disability index (ODI), Bourne-mouth questionnaire (BQ) and numeric pain rating scale (NRS_{pain}) were filled in at baseline. The modified

questionnaires determined what the patient considered an acceptable post-treatment outcome which allowed us to calculate the MCID_{pre}. Concurrent comparisons between the MCID_{pre}, instrument measurement error and a retrospective approach of establishing the minimal clinically important difference (MCID_{post}) were made. The results showed the prospective acceptable outcome method scores to have acceptable reproducibility outside measurement error. MCID_{pre} was 4.5 larger for the ODI and 1.5 times larger for BQ and NRS_{pain} compared to the MCID_{post}. Furthermore, MCID_{pre} and patients post-treatment acceptable change was almost equal for the NRS_{pain} but not for the ODI and BQ. In conclusion, chronic LBP patients have a reasonably realistic idea of an acceptable change in pain, but probably an overly optimistic view of changes in functional and psychological/affective domains before treatment begins.

Keywords Minimal clinically important difference (MCID) · Retrospective · Prospective · Patient-reported outcome · Clinical change

H. H. Lauridsen (✉) · L. Korsholm · N. Grunnet-Nilsson ·
J. Hartvigsen
Clinical Locomotion Science,
Institute for Sports Science and Clinical Biomechanics,
University of Southern Denmark, Campusvej 55,
5230 Odense, Denmark
e-mail: hlauridsen@health.sdu.dk

J. Hartvigsen
Nordic Institute of Chiropractic and Clinical Biomechanics,
Part of Clinical Locomotion Science, Odense, Denmark

L. Korsholm
Department of Statistics, University of Southern Denmark,
Odense, Denmark

C. Manniche
Backcenter Funen, Part of Clinical Locomotion Science,
Ringe, Denmark

Introduction

Patient-reported outcomes provide the patient’s perspective on the effectiveness of the treatment, and this has become an important source of health outcome endpoint data also for patients with low back pain (LBP) [2]. Using such measures requires a detailed understanding of the meaning of the differences observed when used in a longitudinal setting. In 1989 Jaeschke et al. [22] coined the term “minimal clinically important difference” describing the smallest difference in score which patients perceive as beneficial and which would mandate a change in the patient’s management.

So far three methodological categories have been used to determine a meaningful change in measured health status. By far the most common are the a posteriori anchor-based approaches where observed change scores on the outcome measure whose interpretation is under question is compared to some independent measure serving as an aid for interpretation. Of these approaches, the single anchor method employing the patients' global rating of change is the most widely used as it specifies a threshold between important and trivial change ($MCID_{post}$) [9, 19]. Employing patients' global rating of change are, however, limited by recall bias [18, 31, 33] and do not, in themselves, account for the measurement error of the instrument [10]. Other weaknesses have been mentioned such as present-state bias where the patients' global rating correlates more to the present health state rather than to the experienced change during the treatment [39], motivational bias where patients going through a cumbersome treatment have a tendency to overestimate their improvement [1] and the complexity of the global rating of change question compared to questions in multi-item scales [30]. Finally, little information exists on the reliability and validity of the global ratings [9, 10].

Two a priori methods for estimating the minimal important difference include the between-patient method developed by Redelmeier et al. [34] and the clinician-based prognostic rating method by Westaway et al. [40]. Several limitations have been highlighted for the between-patient method. First, it is questionable if the results of this method are comparable to within-patient methods. Second, the method poses practical difficulties of assembling a representative group of patients. Finally, the issue of generalisability of the results across disease states have been challenged [42]. A drawback for the prognostic rating method is that it focuses on the clinician as the arbiter of good prognosis and not the patient. Thus, the question "how do we determine a clinically acceptable change before treatment begins?" remains unanswered. This is unfortunate as aligning clinician expectations to what is acceptable to the patient is important for the clinical outcome and satisfaction [23, 29].

We therefore developed a novel a priori method of estimating low back pain patients' acceptable outcome of treatment by modifying the standard outcome instruments. A comparison of the modified questionnaire score to standard instrument score allowed us to calculate the pre-treatment acceptable change score ($MCID_{pre}$) and compare it to $MCID_{post}$, post-treatment acceptable change and measurement error.

The overall objective of this study was to develop a novel method to estimate LBP patients' view of an acceptable outcome of treatment before it begins. We did this by: (a) developing new questionnaires measuring

acceptable outcomes by modifying well-known questionnaires, (b) testing the modified questionnaires for reproducibility, (c) comparing the results of the prospective acceptable outcome method to a well-established retrospective method and measurement error, and (d) determining if patients with chronic LBP can determine an acceptable change before treatment commences.

Methods

Study approval

The study was reported to and accepted by The Danish Data Protection Agency.

Patient selection

Patients suffering from chronic low back pain and/or leg pain were recruited from an out-patient hospital back pain clinic in 2005. Inclusion criteria were: (1) age between 18 and 60 years, (2) presence of low back pain and/or leg pain, and (3) able to read and understand Danish. Patients were excluded if: (1) a pathological disorder of the spine (e.g. fractures, spinal infections or malignancy, ankylosing spondylitis, rheumatoid arthritis, or other inflammatory diseases) was suspected, (2) they had received a prior back operation, (3) they showed signs and symptoms of a progressive neurological disorder, (4) pending action for damages/litigation were recorded in the case notes, and (5) they had been diagnosed with a psychiatric disorder. Included patients received oral and written information about the project and gave their informed consent.

Design

Pilot study

The modified questionnaires (see outcome measures) were tested for face validity prior to the main study. Twenty-five consecutive patients were interviewed by the first author after filling in the baseline questionnaire booklet. The semi-structured interview focused on difficulty and comprehension in answering the modified questionnaire and resulted in minor changes.

Main study

Patients fulfilling the inclusion criteria were followed over an 8-week period when they received standard conservative treatment according to the Danish national guidelines for the management of LBP [27]. They received a questionnaire booklet at baseline before commencing the

treatment. Questionnaire booklets were mailed to all patients at 1-week (test–retest study for the modified questionnaires) and 8-week follow-up.

Post-hoc study

In order to clarify whether patients can distinguish between what is acceptable and what are their expectations/hopes to the treatment we randomised another 133 chronic LBP patients from the same out-patient hospital back pain clinic as the main study into three groups. Group A filled in three 11-box numeric rating scales for: (1) pain intensity over the past week, (2) their expectations/hopes to the treatment, and (3) their acceptable result of the treatment. Group B and C received question 1 and *either* question 2 or question 3, respectively, and these groups were compared to group A.

Outcome measures

At baseline, patients provided standard sociodemographic information and completed a questionnaire booklet, including standard and modified pain and functional/psychological measures.

Standard LBP outcome measures

1. The Oswestry disability index (ODI) version 2.1 is a self-administered questionnaire measuring “pain-related function” on a 10-item scale with six response categories each. Each item scores from 0 to 5 and the score is subsequently transformed into 0–100 [24, 25, 37].
2. The Bournemouth questionnaire (BQ) measures seven dimensions of back pain: (1) pain intensity, (2) day-to-day physical function, (3) day-to-day social activity, (4) anxiety, (5) depression, (6) work-related fear avoidance, and (7) pain locus of control. Each subscale is scored on an 11-box numeric rating scale (0–10) resulting in a scale range from 0 to 70 with a high

score representing increasing severity of the combined dimensions [6, 7, 21].

3. The 11-box numeric pain rating scale (NRS_{pain}) measures pain intensity over the past week with 0 being “no pain” and 10 being “worst possible pain” [8, 41].

The summary score of the ODI and the BQ and the score of the NRS_{pain} at baseline was termed the pre-treatment standard score.

Modified LBP outcome measures

The standard LBP outcome measures were modified to allow patients (before treatment start) to reflect on what would be acceptable for each item after cessation of treatment. First, patients were asked to differentiate between what they considered an acceptable result and their expectations/hopes to the treatment outcome in the introduction. Second, all the questions in each questionnaire were modified to include the following basic question: “Please indicate what you consider to be (e.g. an acceptable level pain) after completion of the treatment if you had to accept some (e.g. pain)?” (Fig. 1). The summation of all the items of the modified outcome measures was termed the pre-treatment acceptable score. These were subsequently transformed into 0–100 scales to allow for comparison.

At 1-week follow-up, all patients completed both the standard and modified LBP outcome measures including questions indicating whether (a) their condition had changed or (b) their opinion on what constituted an acceptable outcome of the treatment had changed since baseline. Only patients answering “unchanged” to question (a) and “I have not changed my opinion” to question (b) were considered stable and included in the reproducibility study.

At 8-week follow-up, patients were asked to complete the standard LBP outcome instruments. Furthermore, the patient assessed the treatment result by completing a 7-point global rating of change (transition question) [16].

(a) “How would you describe your average back/leg pain over the past week?”

No pain											Worst possible pain
0	1	2	3	4	5	6	7	8	9	10	
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

(b) “Please indicate what you consider to be an acceptable level of pain after completion of the treatment, if you have to accept some pain?”

No pain										Worst possible pain
0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Fig. 1 Sample question of the modified questionnaire. Question **a** is the first question from the Bournemouth questionnaire; question **b** is the modified question

Focus on the *change* in health rather than the *present* health state was optimised by informing the patients of their baseline global rating of pain severity before answering the transition questions (TQ) [17, 20].

Statistical methods

Descriptive statistics were used to summarise patient demographic and clinical data and frequency distributions of the pre-treatment acceptable scores were generated for each of the modified outcome measures. Pre-treatment standard scores were also correlated with the pre-treatment acceptable scores to ensure score reliability for the validity study [36].

Reproducibility

Test–retest reproducibility of the modified questionnaires was carried out on 55 stable patients using the limits of agreement (LOA) method as outlined by Bland and Altman [3]. This method plots the difference between the measurements (modified scores at baseline minus modified scores at 1-week follow-up) against the mean of the same measurements with 95% limits calculated as the mean difference ± 1.96 SD. Thus, 95% of the differences between the two measurements lie between these limits [5].

Second, the internal consistency was tested for the modified ODI using Cronbach's alpha and values between 0.7 and 0.9 was considered acceptable [15, 39]. Alpha was not calculated for the modified BQ and NRS_{pain} as each instrument dimension is represented by only one item [4]. The original questionnaire scale range was used.

Concurrent validity

The MCID_{pre} was compared to a post-treatment anchor-based method of establishing the minimal clinically important difference (MCID_{post}) and measurement error. The minimal detectable change at the 95% confidence level (MDC_{95%}) was used as a marker for measurement error.

The MCID_{pre} was calculated by taking the mean of the pre-treatment acceptable score minus the mean pre-treatment standard score. Thus, the MCID_{pre} represents the change score acceptable to the included patients determined before commencement of treatment.

The MCID_{post} was established using the anchor-based receiver operating characteristic (ROC) curve method [35]. It determines the minimal clinically important difference retrospectively at the individual level of interpretation (changes within patients over time). A ROC curve analysis was used to determine sensitivity and specificity for classifying patients as having experienced an “important improvement” or “no change” was carried out. Patients

classified as having experienced an “important improvement” had to rate themselves as either “much better” or “better” on the TQ. The optimal cut-off change score was identified as the cut-point with equally balanced sensitivity and specificity [14], and this was considered an expression of the MCID_{post} at the individual level. Confidence intervals for the MCID_{post} were estimated using STATA's programming function to calculate the optimal cut-point and a bootstrap procedure using 200,000 samples with replacement.

The MDC_{95%} expresses the degree of change required in an individual's score, in order to establish it (at a 95% confidence level) as being a “real change” over and above measurement error [2, 11]. The standard error of the measurement (SEM_{consistency}) was used to indicate the MDC_{95%} and was defined as the square root of the residual variance computed with ANOVA for random effects [12, 28]. At the 95% confidence level, the MDC was calculated as $1.96 \times \sqrt{2} \times \text{SEM}_{\text{consistency}}$ which is equivalent to $2.77 \times \text{SEM}_{\text{consistency}}$.

Acceptable treatment outcome

To establish whether our cohort of chronic patients was able to determine an acceptable outcome of treatment before it began, the MCID_{pre} was compared to: (a) the MCID_{post} and (b) the post-treatment acceptable change.

The *post-treatment acceptable change* was defined as the mean serial change score in patients who rated themselves as “better” or “much better” on the transition question.

Statistical significance between the MCID_{pre} and the post-treatment acceptable change was tested using Wilcoxon rank-sum test.

All statistical calculations were performed using the statistical package STATA v. 10.0 IC (STATA Corp., College Station, TX, USA) and statistical significance was accepted at the $P < 0.05$ level.

Results

Sample and pre-treatment acceptable score

A total of 225 consecutive patients were eligible for inclusion in the study. Seventy-eight patients (34%) refused to participate or never returned the baseline questionnaires. Thus, 147 patients were available at baseline. At 1- and 8-week follow-ups the response rates were 83.7% ($n = 123$) and 81.0% ($n = 119$), respectively. Table 1 provides characteristics of the participants at baseline and 8-week follow-up.

Table 1 Baseline and 8-week follow-up descriptive data

Characteristic	Study population	
	Baseline (<i>n</i> = 147)	8 weeks (<i>n</i> = 119)
Instrument scores are on their original scale: ODI 0–100; BQ 0–70; NRS _{pain} 0–10		
BQ Bournemouth questionnaire;		
ODI Oswestry disability index;		
NRS _{pain} numeric 11-box pain rating scale		
^a The low back is defined as the area between the lower costal margin and the gluteal folds		
^b The pre-treatment <i>standard</i> score is the pain and disability scores at baseline		
^c The pre-treatment <i>acceptable</i> score is the pain and disability scores of the modified outcome measures		
Age, mean (min, max)	46 (19–82)	47 (19–82)
No. of males (<i>n</i>) (%)	65 (44.2)	52 (43.7)
Pain location (LBP only ^a) (<i>n</i>) (%)	42 (28.6)	35 (29.4)
Duration of current problem (≤ 6 months) (<i>n</i>) (%)	54 (36.7)	46 (38.7)
Other previous treatment for same problem (<i>n</i>) (%) yes	94 (65.7)	
Pre-treatment <i>standard</i> scores ^b , mean (95% CI)		
ODI	34.9 (32.5; 37.3)	29.3 (26.0; 32.5)
BQ	37.0 (34.9; 39.2)	28.9 (25.8; 32.0)
NRS _{pain}	6.2 (5.9; 6.5)	4.9 (4.4; 5.3)
Pre-treatment <i>acceptable</i> scores ^c , mean (95% CI)		
ODI	8.7 (7.2; 10.2)	
BQ	11.4 (10.1; 12.7)	
NRS _{pain}	1.9 (1.7; 2.1)	

A dropout analysis showed higher median days of sick leave during the last 12 months and a higher proportion of patients with light physical workload for the dropouts, however, this was not statistically significant. All other baseline characteristics were identical in the two groups.

Correlation between the pre-treatment standard scores and the pre-treatment acceptable scores were 0.39 for the ODI, 0.34 for the BQ and 0.38 for the NRS_{pain}.

The distribution of the pre-treatment acceptable score (0–100 scales) is shown in Fig. 2. Approximately 90% of the scores were below 32 for all instruments. The pre-treatment acceptable score varied according to the type of instrument. 84% scored less than 16 points on the modified ODI, 56% on the modified BQ and 37% on the modified NRS_{pain}. The NRS_{pain} was the instrument which had most patients with a pre-treatment acceptable score between 17 and 32 points (55%).

Reproducibility

The median (25, 75 percentiles) time interval between baseline and 1-week follow-up was 11 (9, 20) days. The LOA plots show negligible systematic bias for all the modified questionnaires and acceptable 95% LOA intervals. The systematic difference and 95% LOA were 0.8 [−6.6; 8.2] for the modified ODI, −0.2 [−8.8; 8.4] for the modified BQ, and 0.0 [−1.9; 1.9] for the modified NRS_{pain} (data not shown). Cronbach's alpha was 0.84 for the modified ODI.

Concurrent validity

Table 2 provides a concurrent comparison of measurement error (MDC_{95%}), MCID_{pre} and MCID_{post} for all patients.

Fig. 2 Frequency distribution of pre-treatment acceptable score in the pain and functional outcomes. ODI Oswestry disability index; BQ Bournemouth questionnaire; NRS_{pain} numeric 11-box pain rating scale. The pre-treatment acceptable scores were first transformed to 0–100 scales and then grouped into seven categories

Acceptable change determined before treatment (MCID_{pre}) varies between the chosen outcome measures. Thus, the MCID_{pre} for chronic LBP patients scoring the ODI is a 26.1 points reduction (26%) whereas this figure is 25.6 points (37%) for the BQ and 4.2 points (42%) for the NRS_{pain}. Further, the MCID_{pre} appears approximately 4.5 times larger compared to the MCID_{post} for the ODI, however, only 1.5 times larger for the BQ and NRS_{pain}. Third, the MCID_{pre} values are well above measurement error (MDC_{95%}) when looking at the instrument sumscores. However, five out of seven of the BQ subscales showed MCID_{pre} values smaller than measurement error. In contrast, the MCID_{post} values are all smaller than measurement error except the NRS_{pain}.

Table 2 Concurrent comparison of measurement error and patient determined relevant change prospectively and retrospectively

Instrument	Measurement error		Pre-treatment change		Post-treatment change	
	<i>n</i>	<i>MDC</i> _{95%}	<i>n</i>	<i>MCID</i> _{pre} (95% CI)	<i>n</i>	<i>MCID</i> _{post} (S and S)
ODI	55	14.8	119	26.1 (23.9; 28.4)	95	6 (81.8; 67.7)
BQ	55	17.4	119	25.6 (23.6; 27.7)	95	16 (66.7; 88.7)
Pain intensity	55	2.9	119	4.2 (3.9; 4.5)	95	3 (66.7; 87.1)
Physical function	55	3.5	119	4.1 (3.7; 4.4)	95	3 (66.7; 75.8)
Social function	55	4.6	119	3.8 (3.4; 4.2)	95	2 (75.8; 75.8)
Anxiety	55	4.3	119	3.6 (3.2; 4.0)	95	3 (51.5; 75.8)
Depression	55	4.8	119	3.3 (2.8; 3.7)	95	2 (63.6; 72.6)
Fear avoidance	55	4.3	119	3.3 (2.9; 3.7)	95	2 (57.6; 72.6)
Pain locus of control	55	5.1	119	3.3 (2.9; 3.7)	95	1 (78.8; 62.9)
<i>NRS</i> _{pain}	55	2.8	119	4.2 (3.9; 4.5)	91	3 (68.8; 94.9)

The *MCID*_{pre} is calculated as the mean pre-treatment acceptable score subtracted from the mean pre-treatment standard score. Instrument scores are on their original scale: ODI 0–100; BQ 0–70; BQ subscales 0–10; *NRS*_{pain} 0–10

*MDC*_{95%} minimal detectable change at 95% confidence level; *MCID*_{pre} minimal clinically important difference determined prospectively; *MCID*_{post} minimal clinically important difference determined retrospectively; *S* and *S* sensitivity and specificity; *ODI* Oswestry disability index; *BQ* Bournemouth questionnaire; *NRS*_{pain} numeric 11-box pain rating scale

Acceptable change during treatment

A comparison of *MCID*_{pre} to *MCID*_{post} and post-treatment acceptable change (mean change score of those patients who rated themselves as “better” or “much better” on the external criterion) is provided in Fig. 3. The *MCID*_{pre} for the ODI and BQ (including subscales) did not match the post-treatment acceptable change ($P < 0.05$). However, this was not true for the pain measure where *MCID*_{pre} and post-treatment acceptable change were almost comparable (4.2 [3.9; 4.5] vs. 3.8 [2.9; 4.7], $P = 0.33$). Thus, the mean pain change score of patients who rated their improvement as “better” or “much better” after treatment was in fact almost identical to what was an acceptable change in pain before commencing treatment.

Post-hoc study

The post-hoc study showed no difference in baseline data (age, sex, pain duration, disability and pain scores) between the three groups. The median baseline pain score was 7–8 in the three groups. Group A ($n = 41$) received all three questions and expected/hoped to become pain free (median pain score of 0) after the treatment; however, found a median pain score of 2 acceptable. Group B ($n = 46$) received the question about expectations/hopes to the treatment and scored this to a median of 1 after the treatment. Contrary, group C ($n = 46$) received the question about what was an acceptable result after the treatment and found a pain rating of 2 to be acceptable after the treatment.

Discussion

This study applied the prospective acceptable outcome method to two well-established LBP questionnaires (ODI and BQ) and a pain rating scale (*NRS*_{pain}) and compared it to a commonly used retrospective method of establishing an important change and instrument measurement error. We found the pre-treatment acceptable scores reproducible and the *MCID*_{pre} outside instrument measurement error for the instrument sumscores. Furthermore, the *MCID*_{pre} was between 1.5 and 4.5 times larger compared to the *MCID*_{post}. Interestingly, patients seemed to overestimate an acceptable change in functional and psychological/affective aspects but less so for pain *before treatment* when compared to serial improvement in patients who rated themselves as “better” or “much better” *after treatment*.

Our results showed a gap between an acceptable change determined before treatment (*MCID*_{pre}) and the retrospective *MCID*_{post} for all included measures (Table 2). Several reasons may, at least in part, explain why we observed such large discrepancies between what patients consider acceptable before and after treatment. First, it is likely that a response shift has taken place during the treatment. This has been defined as changes in the meaning of the patient’s self-evaluation of the instrument resulting from changes in (a) conceptualisation (i.e. meaning of the item content), (b) values (i.e. a change in the relative importance of the item as an indicator of the measured dimension), or (c) internal standards (i.e. a change in the meaning of the response options) [32, 38]. What is an acceptable outcome before treatment may change during the course of treatment as

Fig. 3 A comparison of $MCID_{pre}$, $MCID_{post}$ and post-treatment acceptable change. *ODI* Oswestry disability index; *BQ* Bournemouth questionnaire; *NRS_{pain}* numeric 11-box pain rating scale. Post-treatment acceptable change is the mean change score in patients who improved on a global retrospective assessment of treatment effect question. Instrument scores are all on 0–100 scales. Confidence intervals for the $MCID_{post}$ were estimated using STATA's programming function and a bootstrap procedure

patients continuously receive information on treatment efficacy, and any mismatch between this and patient-established acceptable outcome is likely to have been reset. Thus, the direction of the response shift is probably of an overestimation of the $MCID_{pre}$ which, during the course of treatment, was adjusted to more realistic benchmarks for treatment outcome and therefore a relatively smaller $MCID_{post}$. Second, it can be questioned whether patients in fact can distinguish between what is acceptable and what are their expectations/hopes to the treatment. If our patients experienced difficulties differentiating between the concepts of “acceptable result of the treatment” and “expectations/hopes to the treatment”, this would result in an underestimation of what is acceptable and an overestimation of the $MCID_{pre}$. The post-hoc study results suggest that chronic LBP patients are able to distinguish the dimensions of acceptable treatment results from their expectations/hopes as group B and C rate these two concepts differently (i.e. group B expected/hoped for no pain after treatment whereas group C could accept a pain rating around 2). Similar findings are reported by Yelland et al. [43] who also found a disparity between what LBP patients regard as a minimum worthwhile reduction in pain and disability and their expectations/hopes to the treatment. Therefore, we believe that patients are able to differentiate between an acceptable outcome and their expectations/hopes to the result of the treatment. Thus, this factor is less likely to

have resulted in an overestimation of the $MCID_{pre}$. A last point to be remembered is that difference scores (i.e. $MCID_{pre}$ and $MCID_{post}$) often show lower reliability compared to both single scores [36].

In summary, we found that the prospective acceptable outcome method yielded results which are not comparable to the retrospective $MCID_{post}$ method and that the disparity is possibly influenced by a response shift during the treatment.

In Fig. 3, the $MCID_{pre}$ is compared to the $MCID_{post}$ and the post-treatment acceptable change. We defined post-treatment acceptable change as the mean change score of those patients who rated themselves as “better” or “much better” on the external criterion. However, some patients may consider less of an improvement acceptable resulting in an overestimation of the post-treatment acceptable change using this method. This is illustrated in the disparity between the post-treatment acceptable change and the $MCID_{post}$. Nevertheless, Fig. 3 clearly demonstrates the disproportion between $MCID_{pre}$ and an acceptable change after the treatment for the ODI and BQ (including subscales) but less so for the *NRS_{pain}*. Accordingly, it may be more difficult for back pain patients to establish an acceptable score before treatment for functional and psychological/affective constructs (ODI and BQ) compared to pain intensity. The ODI and BQ include various aspects of complex cognitive and affective concepts such as social

life and depression which require difficult mental estimations of what is an acceptable level of these dimensions. In contrast, estimating an acceptable pain level is probably easier as this is often the main reason for seeking care. We hypothesise that the difference in cognitive/affective complexity of the constructs included in the instruments accounts, at least in part, for some of the disparity observed between the instruments.

In summary, chronic LBP patients probably overestimate the size of an acceptable change in functional and psychological/affective aspects. In contrast, the same patients seem to have a clearer understanding of what is an acceptable change in pain intensity before treatment begins. Clinically, this highlights the importance of matching the patient established benchmarks for what are acceptable functional and psychological/affective outcomes and the anticipated treatment efficacy during a rehabilitation programme.

Many authors advocate that estimates of the minimal clinically important difference should fall outside measurement error of the instrument in question [10, 13, 26]. Our results show that measurement error ($MDC_{95\%}$) was smaller compared to the $MCID_{pre}$, however, larger when compared to the $MCID_{post}$. Whether this invalidates the $MCID_{post}$ is probably a matter of perspective. We agree that any values of the $MCID_{post}$ lower than measurement error at the group level of interpretation is probably invalid. However, interpreting the $MCID_{post}$ at the individual level could indeed be valid as a patient may well have experienced an important improvement when the change is equal to or above the $MCID_{post}$ value but below measurement error.

In conclusion, the prospective acceptable outcome method offers a benchmark by which the patients' acceptable scores on well known and validated clinical outcome measures can be scrutinised before treatment commences. It yields results which are 1.5–4.5 times larger compared to the retrospective anchor-based method of determining the minimal clinically important difference which possibly can be explained by a response shift. Our results suggest that chronic LBP patients probably overestimate an acceptable change in functional and psychological/affective domains but have a clearer understanding of what is an acceptable change in pain intensity before treatment commences. This has implications for matching patients' acceptable outcome to the expected treatment efficacy.

Acknowledgments We thank Jytte Johannesen and Ida Bhandari for administering the questionnaires. Furthermore, we would like to thank the management and staff at Backcenter Funen for their enthusiastic participation in the project. A special thanks to the seven chiropractic clinics for their involvement in recruiting patients for the study. The study was supported by the Foundation of Chiropractic

Research and Postgraduate Education, The Faculty of Health Science at the University of Southern Denmark and The European Chiropractic Union.

Conflict of interest statement The funding bodies have no control over design, conduct, data, analysis, review, reporting, or interpretation of the research conducted with the funds.

References

1. Aseltine RH, Carlson KJ, Fowler FJ Jr, Barry MJ (1995) Comparing prospective and retrospective measures of treatment outcomes. *Med Care* 33:AS67–AS76
2. Beaton DE (2000) Understanding the relevance of measured change through studies of responsiveness. *Spine* 25:3192–3199. doi:10.1097/00007632-200012150-00015
3. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
4. Bland JM, Altman DG (1997) Cronbach's alpha. *BMJ* 314:572
5. Bland JM, Altman DG (2003) Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol* 22:85–93. doi:10.1002/uog.122
6. Bolton JE, Breen AC (1999) The Bournemouth questionnaire: a short-form comprehensive outcome measure. I. Psychometric properties in back pain patients. *J Manipulative Physiol Ther* 22:503–510. doi:10.1016/S0161-4754(99)70001-1
7. Bolton JE, Humphreys BK (2002) The Bournemouth questionnaire: a short-form comprehensive outcome measure. II. Psychometric properties in neck pain patients. *J Manipulative Physiol Ther* 25:141–148. doi:10.1067/mmt.2002.123333
8. Childs JD, Piva SR, Fritz JM (2005) Responsiveness of the numeric pain rating scale in patients with low back pain. *Spine* 30:1331–1334. doi:10.1097/01.brs.0000164099.92112.29
9. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC (2007) Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 7:541–546. doi:10.1016/j.spinee.2007.01.008
10. Crosby RD, Kolotkin RL, Williams GR (2003) Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 56:395–407. doi:10.1016/S0895-4356(03)00044-1
11. Davidson M, Keating JL (2002) A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 82:8–24
12. de Vet HC, Terwee CB, Knol DL, Bouter LM (2006) When to use agreement versus reliability measures. *J Clin Epidemiol* 59:1033–1039. doi:10.1016/j.jclinepi.2005.10.015
13. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM (2006) Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 4:54–62. doi:10.1186/1477-7525-4-54
14. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL (2000) Defining the clinically important difference in pain outcome measures. *Pain* 88:287–294. doi:10.1016/S0304-3959(00)00339-0
15. Fayers PM, Machin D (2000) Multit-item scales. In: Fayers PM, Machin D (eds) *Quality of life: assessment, analysis and interpretation*. Wiley, pp 72–90
16. Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H (1999) Capturing the patient's view of change as a clinical outcome measure. *JAMA* 282:1157–1162. doi:10.1001/jama.282.12.1157

17. Guyatt GH, Berman LB, Townsend M, Taylor DW (1985) Should study subjects see their previous responses? *J Chronic Dis* 38:1003–1007. doi:[10.1016/0021-9681\(85\)90098-0](https://doi.org/10.1016/0021-9681(85)90098-0)
18. Guyatt GH, Norman GR, Juniper EF, Griffith LE (2002) A critical look at transition ratings. *J Clin Epidemiol* 55:900–908. doi:[10.1016/S0895-4356\(02\)00435-3](https://doi.org/10.1016/S0895-4356(02)00435-3)
19. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR (2002) Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 77:371–383. doi:[10.4065/77.4.371](https://doi.org/10.4065/77.4.371)
20. Guyatt GH, Townsend M, Keller JL, Singer J (1989) Should study subjects see their previous responses: data from a randomized control trial. *J Clin Epidemiol* 42:913–920. doi:[10.1016/0895-4356\(89\)90105-4](https://doi.org/10.1016/0895-4356(89)90105-4)
21. Hartvigsen J, Lauridsen HH, Ekstrom S, Nielsen MB, Lange F, Kofoed N et al (2005) Translation and validation of the danish version of the Bournemouth questionnaire. *J Manipulative Physiol Ther* 28:402–407. doi:[10.1016/j.jmpt.2005.06.012](https://doi.org/10.1016/j.jmpt.2005.06.012)
22. Jaeschke R, Singer J, Guyatt GH (1989) Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 10:407–415. doi:[10.1016/0197-2456\(89\)90005-6](https://doi.org/10.1016/0197-2456(89)90005-6)
23. Kalauokalani D, Cherkin DC, Sherman KJ, Koepsell TD, Deyo RA (2001) Lessons from a trial of acupuncture and massage for low back pain: patient expectations and treatment effects. *Spine* 26:1418–1424. doi:[10.1097/00007632-200107010-00005](https://doi.org/10.1097/00007632-200107010-00005)
24. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N (2006) Danish version of the Oswestry disability index for patients with low back pain. Part 1: Cross-cultural adaptation, reliability and validity in two different populations. *Eur Spine J* 15:1705–1716. doi:[10.1007/s00586-006-0117-9](https://doi.org/10.1007/s00586-006-0117-9)
25. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N (2006) Danish version of the Oswestry disability index for patients with low back pain. Part 2: Sensitivity, specificity and clinically significant improvement in two low back pain populations. *Eur Spine J* 15:1717–1728. doi:[10.1007/s00586-006-0128-6](https://doi.org/10.1007/s00586-006-0128-6)
26. Lydick E, Epstein RS (1993) Interpretation of quality of life changes. *Qual Life Res* 2:221–226. doi:[10.1007/BF00435226](https://doi.org/10.1007/BF00435226)
27. Manniche C, Ankjær-Jensen A, Olesen A, Fog A, Williams K, Biering-Sørensen F (1999) Statens Institut for Medicinsk Teknologivurdering: Ondt i ryggen. Forekomst, behandling og forebyggelse i et MTV-perspektiv. *Medicinsk Teknologivurdering Ser B* 1(1)
28. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1:30–46. doi:[10.1037/1082-989X.1.1.30](https://doi.org/10.1037/1082-989X.1.1.30)
29. McGregor AH, Hughes SP (2002) The evaluation of the surgical management of nerve root compression in patients with low back pain: Part 2: patient expectations and satisfaction. *Spine* 27:1471–1476. doi:[10.1097/00007632-200207010-00019](https://doi.org/10.1097/00007632-200207010-00019)
30. Middel B, Goudriaan H, de Greef M, Stewart R, van Sonderen E, Bouma J et al (2006) Recall bias did not affect perceived magnitude of change in health-related functional status. *J Clin Epidemiol* 59:503–511. doi:[10.1016/j.jclinepi.2005.08.018](https://doi.org/10.1016/j.jclinepi.2005.08.018)
31. Norman GR, Stratford P, Regehr G (1997) Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 50:869–879. doi:[10.1016/S0895-4356\(97\)00097-8](https://doi.org/10.1016/S0895-4356(97)00097-8)
32. Oort FJ, Visser MR, Sprangers MA (2005) An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. *Qual Life Res* 14:599–609. doi:[10.1007/s11136-004-0831-x](https://doi.org/10.1007/s11136-004-0831-x)
33. Pellise F, Vidal X, Hernandez A, Cedraschi C, Bago J, Villanueva C (2005) Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. *Spine* 30:365–368. doi:[10.1097/01.brs.0000152096.48237.7c](https://doi.org/10.1097/01.brs.0000152096.48237.7c)
34. Redelmeier DA, Guyatt GH, Goldstein RS (1996) Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 49:1215–1219. doi:[10.1016/S0895-4356\(96\)00206-5](https://doi.org/10.1016/S0895-4356(96)00206-5)
35. Revicki D, Hays RD, Cella D, Sloan J (2008) Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 61:102–109. doi:[10.1016/j.jclinepi.2007.03.012](https://doi.org/10.1016/j.jclinepi.2007.03.012)
36. Rogosa DR, Willett JB (1983) Demonstrating the reliability of the difference score in the measurement of change. *J Educ Meas* 20:335–343. doi:[10.1111/j.1745-3984.1983.tb00211.x](https://doi.org/10.1111/j.1745-3984.1983.tb00211.x)
37. Roland M, Fairbank J (2000) The Roland-Morris disability questionnaire and the Oswestry disability questionnaire. *Spine* 25:3115–3124. doi:[10.1097/00007632-200012150-00006](https://doi.org/10.1097/00007632-200012150-00006)
38. Sprangers MA, Van Dam FS, Broersen J, Lodder L, Wever L, Visser MR et al (1999) Revealing response shift in longitudinal research on fatigue—the use of the thentest approach. *Acta Oncol* 38:709–718. doi:[10.1080/028418699432860](https://doi.org/10.1080/028418699432860)
39. Streiner DL, Norman GR (2003) Health measurement scales. A practical guide to their development and use. Oxford Medical Publications, Oxford
40. Westaway MD, Stratford PW, Binkley JM (1998) The patient-specific functional scale: validation of its use in persons with neck dysfunction. *J Orthop Sports Phys Ther* 27:331–338
41. Williamson A, Hoggart B (2005) Pain: a review of three commonly used pain rating scales. *J Clin Nurs* 14:798–804. doi:[10.1111/j.1365-2702.2005.01121.x](https://doi.org/10.1111/j.1365-2702.2005.01121.x)
42. Wright JG (1996) The minimal important difference: who's to say what is important? *J Clin Epidemiol* 49:1221–1222. doi:[10.1016/S0895-4356\(96\)00207-7](https://doi.org/10.1016/S0895-4356(96)00207-7)
43. Yelland MJ, Schluter PJ (2006) Defining worthwhile and desired responses to treatment of chronic low back pain. *Pain Med* 7:38–45. doi:[10.1111/j.1526-4637.2006.00087.x](https://doi.org/10.1111/j.1526-4637.2006.00087.x)