Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors
systematic review of empirical studies

Frahm Olsen, Mette; Bjerre, Eik; Hansen, Maria Damkjær; Tendal, Britta; Hilden, Jørgen; Hróbjartsson, Ásbjörn

Published in:
Journal of Clinical Epidemiology

DOI:
10.1016/j.jclinepi.2018.05.007

Publication date:
2018

Document version
Accepted manuscript

Document license
CC BY-NC-ND

Citation for published version (APA):

Terms of use
This work is brought to you by the University of Southern Denmark through the SDU Research Portal. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk
Minimum clinically important differences in chronic pain vary considerable by baseline pain and methodological factors: Systematic review of empirical studies.

Mette Frahm Olsen, PhD, Eik Bjerre, MSc, Maria Damkjær Hansen, MD, Britta Tendal, PhD, Jørgen Hilden, MD, Asbjørn Hróbjartsson, PhD

PII: S0895-4356(18)30124-0
DOI: 10.1016/j.jclinepi.2018.05.007
Reference: JCE 9654

To appear in: Journal of Clinical Epidemiology

Received Date: 9 February 2018
Revised Date: 17 April 2018
Accepted Date: 14 May 2018


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors: Systematic review of empirical studies.

Running title: MCID in chronic pain

Mette Frahm Olsen, PhD
Eik Bjerre, MSc
Maria Damkjær Hansen, MD
Britta Tendal, PhD
Jørgen Hilden, MD
Asbjørn Hróbjartsson, PhD

* Corresponding author
1) Post Doc, Nordic Cochrane Centre, Rigshospitalet, Blegdamsvej 9, Dept. 7811, 2100 Copenhagen Ø, Denmark
2) PhD student, University Hospitals’ Centre for Health Research (UCSF), Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen Ø, Denmark
3) MD, Hvidovre Hospital, Kettegård Alle 30, 2650 Hvidovre, Denmark
4) Senior Advisor, Danish Health Authority, Islands Brygge 67, 2300 Copenhagen S, Denmark
5) Associate Professor (Emeritus), Section of Biostatistics, University of Copenhagen, Østre Farigmagsgade 5, 1353 Copenhagen Ø, Denmark
6) Professor, Center for Evidence-Based Medicine, University of Southern Denmark & Odense University Hospital, Sdr. Boulevard 29, Gate 50 (Videncenteret), 5000 Odense C, Denmark
Abstract

Background: The minimum clinically important difference (MCID) is used to interpret the relevance of treatment effects, e.g. when developing clinical guidelines, evaluating trial results or planning sample sizes. There is currently no agreement on an appropriate MCID in chronic pain and little is known about which contextual factors cause variation.

Methods: Systematic review. We searched PubMed, EMBASE and Cochrane Library. Eligible studies determined MCID for chronic pain based on a one-dimensional pain scale, a patient-reported transition scale of perceived improvement; and either a mean change analysis (mean difference in pain among minimally improved patients) or a threshold analysis (pain reduction associated with best sensitivity and specificity for identifying minimally improved patients). Main results were descriptively summarized due to considerable heterogeneity, which was quantified using meta-analyses and explored using subgroup analyses and meta-regression.

Results: We included 66 studies (31,254 patients). Median absolute MCID was 23 mm on a 0-100 mm scale [IQR 12-39] and median relative MCID was 34% [IQR 22-45] among studies using the mean change approach. In both cases, heterogeneity was very high: absolute MCID $I^2 = 99\%$, relative MCID $I^2 = 96\%$. High variation was also seen among studies using the threshold approach: median absolute MCID was 20 mm [IQR 15-30] and relative MCID was 32% [IQR 15-41]. Absolute MCID was strongly associated with baseline pain, explaining approximately two-thirds of the variation, and to a lesser degree with the operational definition of minimum pain relief and clinical condition. A total of 15 clinical and methodological factors were assessed as possible causes for variation in MCID.

Conclusions: MCID for chronic pain relief vary considerably. Baseline pain is strongly associated with absolute, but not relative, measures. To a much lesser degree, MCID is also influenced by the operational definition of relevant pain relief and possibly by clinical condition. Explicit and conscientious reflections on the choice of an MCID are required when classifying effect sizes as clinically important or trivial.
What is new?

- There is a considerable degree of variation between the results of studies assessing the minimum clinically important difference (MCID) in chronic pain.

- Baseline pain is the main cause of variation in MCID, explaining approximately two-thirds of the variation in absolute measures of MCID in the reviewed studies. Variation in MCID was to a lesser degree also caused by the operational definition of minimal pain relief, while it remains uncertain whether clinical condition influence MCID.

- Individual clinicians, researchers, guideline developers, or consensus building committees will benefit from referring to the overview of studies provided in this systematic review, when deciding on a MCID value for chronic pain in a given clinical setting.

- Explicit and conscientious reflections on the choice of MCID value are required when using it to classify research results as clinically important or trivial.
Introduction

A common challenge for patients, physicians, clinical guideline developers, and health care policy makers is to decide whether a treatment effect is of a magnitude that is clinically important. Such a decision has broad implication for the interpretation of results of clinical studies, such as randomised clinical trials or meta-analyses. In chronic pain, the cut-off for a clinically relevant effect impacts on which interventions are considered clinically useful, for example for arthritis, back pain, cancer-related pain, fibromyalgia, and headache[1–5]. Also, the cut-off will influence directly on the choice of appropriate sample sizes of future clinical pain trials, as a reasonable ideal for a confirmative trial is to be able to detect a clinically relevant effect size.

The concept of minimum clinically important difference (MCID) was defined in 1989 by Jaeschke and colleagues as “the smallest difference in score in the domain of interest which participants perceive as beneficial and which would mandate, in the absence of troublesome side effects and costs, a change in the patient’s management”[6]. The concept defines relevant effect size based on patients’ perception and clinical considerations[6,7], reflecting a clear distinction between clinical relevance and statistical significance.

The concept of minimum clinically important difference was later supplemented by a related notion: the substantial (and not only minimum) clinically important difference [8]. More recently, the concept of MCID has been suggested as the appropriate effect unit in meta-analyses of continuous outcome measures [9].

MCID is sometimes based on objective criteria [10] or expert consensus judgement [11]. However, there is an increasing awareness of the relevance of patient-reported outcomes in general [12], and in pain assessment it seems particularly reasonable to anchor clinical importance to the patients’ subjective experience. A large number of empirical studies have been conducted to estimate MCID in chronic pain, but the studies differ considerably with regard to methodology, clinical conditions, and findings. Baseline pain
may likely influence absolute (as opposed to relative) measures of MCID [13], but it remains unclear whether other clinical or methodological factors cause variation.

Thus, we decided to systematically review empirical studies of MCID in relief of chronic pain and to examine possible causes for variation between study results, with a specific focus on their dependency on baseline pain levels.

**Methods**

*Eligibility criteria*

We included prospective studies of patients with chronic pain, regardless of age, clinical condition, and intervention, in which pain intensity was assessed on a one-dimensional scale, e.g. a 0-100 mm visual analogue scale (VAS) or a 0-10 point numeric rating scale (NRS), and in which MCID was based on a transition scale using patients’ perception of change to determine clinical importance. Pain was considered chronic when duration was more than one month (or if duration not reported, when described as chronic in a study report). Studies were excluded if MCID was derived from objective criteria (e.g. return to work), distribution of data (e.g. the minimum detectable difference), or expert consensus.

A typical eligible study would ask patients to score their pain intensity, e.g. using a VAS, at baseline and follow-up. At follow-up, patients were also asked to categorise their change in pain intensity, or general health, often using response options such as “a little better”, “somewhat better” or a similar expression. MCID was then estimated from the change in scores on the pain scale among patients having categorised their change as indicating a minimum degree of improvement.
We included studies with two types of analytical approaches: 1) the “mean change” analysis, i.e. the mean difference in pain scores among patients with a minimum degree of pain reduction\cite{14}; or 2) the “threshold analysis”, i.e. the threshold for change in pain score which most accurately (best sensitivity and specificity) identified patients experiencing a minimum degree of pain reduction (analogous to a diagnostic test where patients’ perception of change is the gold standard\cite{15}).

**Search strategy**

We searched PubMed, EMBASE and Cochrane Library until end of August 2016 with no language restrictions. The core search string was: (minimal OR minimally OR minimum OR “clinically significant” OR “clinically important” OR “clinically meaningful” OR “clinically relevant”) AND (difference OR change OR relief OR reduction) AND (“pain measurement*” OR “visual analog scale” OR “numeric rating scale”) AND (pain) with variations according to the specific database (Appendix I). The reference lists of all included studies and relevant review papers were read systematically to identify further studies.

Screening of titles and abstracts to determine the eligibility of studies was done by the primary author (MFO), while the selected full-text records were examined by two researchers independently (MFO and EB/BT/MDH). Any disagreement was solved by discussion.

**Data extraction and retrieval**

Data extraction was conducted by two researchers independently (MFO and EB/BT/MDH) using pretested data extraction forms generated in EpiData (EpiData Association, Odense, Denmark). Any disagreements were solved by discussion. For each study, we extracted descriptive data including publication year, study design, setting, type of intervention, sampling method, sample size, and definition of patients with relevant change (see Appendix II for complete list).
For studies using a mean change approach, we extracted the following outcome data: MCID for pain relief (absolute values in mm or points and relative value in percentage change from baseline) and for worsening of pain (absolute and relative values); and the substantial clinically important difference for relief and worsening of pain (absolute and relative values). We extracted MCID as the mean difference in pain score among patients who indicated a one-category improvement (e.g. “a little better”). If unavailable, we extracted the mean difference among patients who were minimally improved by authors’ definition (e.g. some authors defined minimum important change as the mean difference in pain score among all patients with a one- or two-category improvement). Similarly, we extracted substantial clinically important differences as the mean difference among patients with a two-category improvement or used authors’ definition. We extracted the point estimate of MCID with the corresponding standard error or, if unavailable, other measures of variation such as standard deviation or 95% confidence interval.

For studies using a threshold approach, we extracted information about the definition of responders (i.e. patients with relevant change) and non-responders and the cut-off point with its corresponding sensitivity (i.e. percentage of responders correctly classified as such) and specificity (i.e. percentage of non-responders correctly classified as such).

If studies reported pain scores from several concurrent pain assessments (e.g. back pain and leg pain) we extracted the assessment where more data were available or, if no difference found, we randomly selected which to extract. If studies reported pain scores from repeated assessments, we included data from the first assessment conducted. All scales were standardised to a 0-100 mm scale. When studies reported pain assessments based on both VAS and NRS, we used the assessments based on VAS.

For each study, we assessed risk of attrition bias (studies were considered at low risk when attrition <10%) and risk of non-representative study sample (studies were considered at low risk if using consecutive or random sampling).
Data synthesis and analysis

For each study, we noted or calculated absolute and relative MCID for pain relief (e.g. an improvement from 80 to 60 mm on a 100 mm VAS scale corresponds to 20 mm absolute change and 25% relative change). For studies using the mean change approach, we noted, calculated or estimated the standard error. For studies not reporting standard error, this was either calculated based on available information or imputed from the median standard deviation of studies with available data. Also, we noted the result of each study that explored causes of variation, e.g. baseline pain.

We summarised MCID by median and inter-quartile range (IQR) according to analytical strategy (mean change or threshold approach) and type of transition scale (pain or general health). We then meta-analysed MCID based on the mean change approach with inverse-variance analysis and random effects models. Studies using the threshold approach were not meta-analysed as they do not include a straightforward measure for the statistical uncertainty of their results.

We expected, and found, a considerable degree of heterogeneity. We analysed reasons for heterogeneity using four approaches: First, we sub-grouped studies according to predefined criteria: clinical condition, type of pain scale (VAS vs. NRS), concept of transition scale (pain vs. general health), symmetry of transition scale (symmetrical vs. asymmetrical), definitions of relevant pain relief (one category vs. several categories improvement vs. distinction between meaningful and non-meaningful change), follow-up time (< vs. ≥ 3 months), risk of attrition bias (low vs. high or unclear) and risk of non-representative sample (low vs. high or unclear). We noted the median MCID, and meta-analysed MCID, within each subgroup. Second, we summarised results of studies which had analysed within-study associations between MCID and possible reasons for variation, such as individual patients’ baseline pain scores, age, or duration of pain. Third, we meta-regressed mean baseline pain scores and MCID. The analysis of the mean change approach studies incorporated the statistical certainty of each study result while this was ignored for the studies using the
threshold approach. Fourth, we used multivariate meta-regression to investigate the proportion of the between-study variance in MCID that could be explained from clinical or methodological factors. We applied a backward-stepwise selection strategy to identify relevant co-factors for the final model. The initial factors considered for selection was mean baseline pain and the eight clinical and methodological factors predefined for subgrouping (a total of 15 independent factors).

We also analyzed the following supplementary outcomes: Substantial clinically important difference for pain relief (absolute and relative change) and minimum and substantial clinically important differences for worsening of pain (absolute and relative change). These secondary analyses did not involve exploration of heterogeneity. All data analyses were done using Stata/IC version 14.2.

Results

Study selection

We screened 1,553 database records and read 273 full-text publications (Figure 1). We excluded 212 publications, mostly because they included patients with acute pain (n=36) or used multi-dimensional pain scales (n=53). Thus, we included 61 publications[8,16–75] reporting 66 studies (31,254 patients).

Characteristics of included studies

The most frequently studied condition was back pain (26 studies), of which 14 studies included patients with non-specific pain and 12 studies radiculopathy or stenosis (e.g. spondylolisthesis or disc herniation). Other clinical conditions included neck pain (8 studies), arthritis (11 studies), musculoskeletal pain (7 studies), paediatric conditions (3 studies), and other clinical conditions (11 studies, e.g. phantom limb pain or scleroderma) (Table 1).
Thirty-nine studies assessed pain using a 0-100 mm VAS and 28 studies used a 0-10 point NRS (one study used both). Pain was described in general terms such as “pain intensity”, “pain severity”, or just “pain”, while only a few studies specified that patients should report pain on movement or rest[56,69,70] or pain when at worst or best[19,20,38,50].

Two types of transition scales were used. Patients were asked to assess change in pain in 13 studies (e.g. “In what way did your low back pain change?”), and in 53 studies patients were asked to assess change in general health (e.g. “compare your current health to health before treatment”). Transition scales were often symmetrical (36 studies) and including 4-6 response categories (range: 2-15) describing degrees of both improvement and deterioration. Asymmetrical scales (30 studies) were either one-sided, addressing only degrees of improvement, or included more response categories addressing improvement than worsening of pain.

Fifty-four studies used the mean change approach. Of these, MCID estimated as a one-category improvement was available from 31 studies [16,18,21–32,34,35,37,38,40,41,43,44,48,50,51,63,64,66,72,74]. The response categories used similar wordings such as “a little better” or “slightly improved”. In 18 studies [17,19,45,46,49,52,53,55,57,59–61,65,67,73,75], MCID was available as the mean difference in pain score among patients with several degrees of improvement based on authors’ definition, e.g. combining patients answering “minimally improved”, “much improved” and “very much improved”. In five studies, MCID was defined by asking patients to differentiate between non-meaningful and meaningful change, using categories such as “there was some decrease in my pain, but not enough to be meaningful” vs “my pain decreased to a meaningful extent”[36,39,54,68].

MCID was assessed using the threshold approach in 46 studies. Of these, patients were defined as responders if they had at least a one-category improvement in 22 studies [18,23,24,27,28,30,31,34,37,40–42,44,51–53,55,62,75], while they needed at least a two- [8,17,21,29,32,48,49,57,59–61,66,72,73], three-
[19,65], or four-category [20,69,70] improvement in 18 studies. In three studies, patients were asked to differentiate between non-meaningful and meaningful change [33,39,54]. Also, three studies used two-step transition scales, in which patients were first asked if they were improved, and if so, how important the improvement was [45,46,56,71].

**Overview of MCID**

Disregarding analytical approach and type of transition scale, the absolute MCID reported from 66 studies ranged from -1 to 82 mm with a median of 20 mm [IQR: 13 to 31]. The relative MCID was available from 17 studies and ranged from 10 to 56% with a median of 32% [IQR 22 to 41].

**MCID based on mean change analysis**

We included 51 studies (13,591 patients) in our synthesis of MCID based on the mean change analysis, while three studies (4,650 patients) [24,30,31] were excluded from data synthesis as they reported median change rather than mean change estimates of MCID.

Of the 51 studies, 50 (13,561 patients) provided absolute MCID: median 23 mm [IQR 12-39]. Seven studies (1,465 patients) provided relative MCID: median 34% [IQR 22-45]. There was no clear difference in results between MCID depending on type of transition scale (Figures 2a & 2b, Table 2a & 3).

Of the 51 studies, 47 (12,779 patients) were included in meta-analysis of absolute MCID, while five studies (1,200 patients) were included in meta-analysis of relative MCID. Additional studies could not be included as standard errors of MCID were not provided and could not be calculated or imputed based on available data [16,32,37,53,65]. Meta-analyses of both absolute and relative measures revealed very high heterogeneity: $I^2 = 99\%, \ p<0.001$ and $I^2 = 96\%, \ p<0.001$, respectively (Table 2a).

**MCID based on threshold analysis**


We included 44 studies (26,857 patients) of MCID based on the threshold analytical approach, while two studies (759 patients) were excluded since results were reported as a range [76] or were available only for subgroups without data for retrieving an overall estimate [29]. Absolute MCID was available from 43 studies (26,673 patients) (Figure 3a): median 20 mm [IQR 15 to 30]. In 15 studies (9,836 patients), the median relative MCID was 32% [IQR 15 to 41] (Table 2a, Figure 3b).

Analyses of causes for variation

Our exploration of causes for heterogeneity was based on studies using the mean change analytical approach and assessing absolute measures, since the threshold approach provided no measure of statistical precision and there were too few relative measures available for meaningful subgroup analyses.

Subgroup analysis

The median MCID, as well as weighted means unadjusted for baseline pain, for subgroups of studies is presented in Table 3, but should be considered preliminary results as they are potentially influenced by underlying differences in baseline pain and the contextual factors causing variation.

Within-study analyses of potential causes for MCID variation

Six studies assessed and found a positive correlation between patients’ individual baseline pain score and their subsequent assessment of absolute MCID with correlation coefficients ranging from 0.40 to 0.58. Of four studies assessed the correlation between baseline pain score and relative MCID, one study reported an association with a correlation coefficient of 0.37. In addition, 13 studies had stratified their assessment of absolute MCID according to baseline pain scores, but with very variable stratification criteria. The general tendency, however, was for higher baseline pain scores to result in higher absolute MCID (Appendix III).

Other potential causes for variation in MCID had been assessed in 16 of the 61 studies, of which only one adjusted their analysis for baseline pain scores [61]. Of seven studies assessing the impact of
chronicity[29,40,41,61,69,70,72], two studies found that patients with duration of pain <3 months had a lower absolute MCID than patients with longer pain duration[41,72]; and one study found MCID to be higher[61], but the latter finding disappeared when results were stratified by baseline pain. The potential impact of clinical condition was assessed in six studies[30,31,37,46,56,62], of which only one found an association (i.e. MCID was higher in abdominal pain compared to headache and musculoskeletal pain)[37]. The same study found that MCID was more stable (i.e. smaller range) among patients with constant pain compared to intermittent pain. Five studies assessed dependency on sex and age[30,31,68–70], of which one study found that MCID was higher among younger patients[68]. In addition, different types of treatment[30,31,66,73], duration of follow-up[68] and entry point in the health care system (primary vs. secondary sector)[46] were assessed without showing an impact on MCID in chronic pain.

Between-study analysis of the impact of study baseline pain on MCID variation

Meta-regression of 40 studies (10,938 patients) using a mean change analysis and with available mean baseline pain showed a strong association between baseline pain and absolute measures of MCID. For each 10 mm increase in baseline pain, MCID also increased by 10 mm (95% CI: 8 to 12 mm, p<0.001, I²: 96%) (Figure 4a).

Absolute MCID from 37 studies (23,926 patients) using the threshold approach showed a similar dependency on baseline pain (Figure 4b).

In contrast, there was no association between mean baseline pain and relative MCID, neither among the five studies using the mean change approach (p=0.66) nor among 13 studies using the threshold approach (p=0.36).

Between-study analyses of potential causes for MCID variation adjusted for study baseline pain
Meta-regression analyses of MCID adjusted for mean baseline pain within subgroups suggested that one of seven clinical categories (i.e. back pain with radiculopathy/stenosis) as well as definition of relevant pain relief (i.e. a single-category improvement on a transition scale used to operationalize MCID) were associated with lower MCID (Table 3).

A supplementary multivariate meta-regression model resulted in inclusion of three factors: mean baseline pain, definition of relevant pain relief and one of seven clinical categories (back pain with radiculopathy/stenosis). The model explained 77% of the total between-study variance (adjusted R-squared) and pointed at baseline pain as clearly the most important factor as a model including only baseline pain accounted for 67% of the total between-study variation.

**Supplementary outcomes**

Analysis of supplementary outcomes for pain relief and worsening showed a similar high variation between studies (Table 2b).

**Discussion**

We included 66 studies (31,254 patients) of MCID in chronic pain and found considerable variation between studies. The median absolute MCID was 23 mm [IQR 12 to 39] in 50 studies using a mean change approach and 20 mm [IQR 15 to 30] in 43 studies using a threshold approach. The median relative MCID was 34% [IQR 22 to 45] and 32% [IQR 15 to 41] for the two approaches, respectively. Baseline pain score was strongly associated with absolute values, explaining two-thirds of the variation between study results. In addition, weaker associations were observed for the criteria for defining patients with relevant pain relief and one among seven studied clinical conditions.

**Strengths and limitations**
The present study is the first systematic review of MCID in chronic pain without restrictions to a specific clinical condition. We identified 66 studies involving more than 31,000 patients and a broad range of clinical conditions, analytical approaches and transition scales. We expected, found and partly explained considerable variation between study results. Approximately 77% of the inter-study variation for absolute values was accounted for by three study characteristics: baseline pain, definition of relevant pain relief and, to a lesser extent, clinical condition. Our study provides an overview of a multifaceted pattern of results, explains a considerable part of the between-study variation, offers a catalogue of conducted studies, and of factors having been investigated for the inter-study and intra-study variation.

However, there was notable residual heterogeneity. Even within narrowly defined clinical conditions and analyses including only few studies, values of MCID spanned a considerable range and heterogeneity was high, often with I² over 80%. Furthermore, our analyses were restricted by the limitations inherent in aggregated baseline data, which may conceal differences in methods and population (i.e. ecological fallacy [77]). Unfortunately, we did not have sufficient data for meaningful subgroup analyses of relative MCID, but acknowledge that absolute values’ association with baseline pain makes relative values preferable. Lastly, some of the subgroups involved too few studies to ensure that all relevant associations were detected.

We based our heterogeneity analysis on studies using the mean change analysis, as studies using the threshold analysis did not report the statistical uncertainty of their estimate of MCID, thus precluding a detailed analysis of heterogeneity. Furthermore, we included studies that used ‘pain-specific’ transition scales, asking patients about their change in pain, as well as studies with broad transition scales, asking patients about their change in general health status (and not specifically pain). Though strictly speaking, the latter studies confuse the concepts of pain and the concept of general clinical improvement, we included such studies as we expect that for chronic pain patients, pain is often a central aspect of the clinical condition (and also because this type of transition scale is very commonly applied within the field). We found
that results from studies with a narrow transition scale were not clearly different from those from studies with broad transition scale.

Access to individual patient data would have improved the chances of further identifying clinical and methodological causes of heterogeneity. In addition, the available data did not allow us to assess the potential impact of different pain aspects, such as intermittent vs. constant pain or pain at rest vs. on movement, since this was generally not reported by the primary studies. Finally, there may be a risk of recall bias in studies of patients who simultaneously assess their pain status and their pain change [78].

Other studies
Despite the vast number of primary studies that have been done, there are very few systematic reviews of the minimum clinically relevant change in chronic pain. The ones that have been published include reviews among patients with rheumatologic conditions [13,79] and back pain [11,80,81] and mostly include multidimensional scales, such as the Roland-Morris Disability Questionnaire. Similar to the present review, these have reported wide variation in study designs and methods.

Using a similar approach as in the present review, we have previously assessed the minimum clinically important change in acute pain and also found large heterogeneity between study results, partly explained by baseline pain level and study design factors such as definition of relevant pain relief and whether the study involved repeated assessments of perceived pain relief [82].

Mechanisms and perspectives
The variation in results between our included studies may be due either to chance or true differences in MCID related to clinical diagnosis or methodological differences. The large number of studies and patients included, and the very high $I^2$ values observed, implies that random variation played only a minor role in our overall analyses though a more prominent role in subgroup analyses.
A central question is how much of the large variation observed is due to clinical diagnosis. Whereas studies on MCID for acute pain are typically, but not always, done in groups of patients with diverse diagnoses (e.g. everyone presenting with acute pain at an Emergency Department), studies on chronic pain are most often based on patients with a specific clinical diagnosis. Six of the 66 included studies assessed potential differences in MCID between diagnoses, of which only one found a difference (patients with abdominal pain had larger MCID than patients with headache and musculoskeletal pain, but the authors did not report whether this could be explained by differences in baseline pain). Our subgroup and meta-regression analyses suggested that, when adjusted for baseline pain, back pain with radiculopathy/stenosis had somewhat lower MCID than other pain conditions. Still, there is a risk of a spurious finding and the within-study analyses indicate that diagnoses do not have a large impact. We interpret our findings to say that the impact of clinical condition on MCID is possible but that it remains uncertain which clinical conditions are associated with higher or lower values and to what degree.

However, it seems clear that other clinical and methodological factors are important causes for the observed variation. Baseline pain was strongly associated with absolute MCID. Our results predict that for each 10 mm increase in baseline pain, MCID also increases by approximately 10 mm. Compared to absolute values, relative values of MCID are less sensitive to baseline pain scores. Theoretically, a challenge of relative values is that they lack interval scale properties at the scale extremes, for example when baseline values are close to zero and small degrees of pain change result in very large relative changes\[83\]. However, none of the included studies of relative MCID had extremely small or large mean baseline pain scores (i.e. < 20 mm or >80 mm) and, despite this, the heterogeneity of relative MCID was still considerable.

Also, studies applied a wide range of transition scales, with the number of response categories ranging from two to 15, and the operational definition of “minimally improved patients” differing widely. In an attempt to extract comparable data, we standardized MCID as a one-category improvement when available and used
authors’ definition in remaining studies. Not surprisingly, we documented that MCID defined as a single category improvement on a transition scale resulted in lower values than when more categories had been combined. However, even within studies where mean difference in pain score associated with a single-category improvement on a transition scale was available, MCID values still ranged from -1 to 44 mm. It may be difficult to tease out which of the more subtle differences in concepts and definitions influence results, but it seems reasonable that the presentation of scales to patients and the specific wording of questions will have an impact on MCID[84]. It is therefore noteworthy that although these studies used a patient-centred transition scale, the vast majority actually estimated the clinical importance indirectly, as patients were asked to assess their degree of change, while the judgement of whether the change was clinically important was attributed by the researcher [81]. Only six of the 66 studies (9%) asked patients directly to assess whether they perceived their change as important or not. These studies tended to report high values for MCID.

It is obvious that there is no consensus on how the concept of MCID should be operationally defined [14]. Some researchers have even questioned whether change in pain intensity is closely related to perceptions of pain relief[85] and it has been suggested that assessment of pain relief require a much more complex measure reflecting several components of pain, such as specific sensations, unpleasantness, and hope of improvement[86]. There has also been a general critique that the most common approaches do not consider intervention “costs” or any “troublesome side effects” although these were mentioned in the original definition [87]. Alternative approaches have been developed that integrate costs and side effects [86–89], while other approaches suggest focus on the level of symptoms which patients find tolerable i.e. the “patient acceptable symptom state”[70,90]. This approach corresponds well with the dominant aim of clinical patient care to reduce pain to an acceptable level [93] and could be a strong candidate for an alternative to MCID.

**Implications**
The large variability in MCID implies that it is clearly misguided to randomly or selectively pick a single study and use its result as a kind of scale constant. Our results support that an empirically derived measure of MCID for chronic pain is context-dependent, and often quite sensitive to clinical and methodological characteristics [92]. Thus, when MCID informs a distinction between trivial and important effects of an intervention, for example during the development of clinical guidelines, it is important to reflect conscientiously and explicitly on the span of results, the baseline pain of patients, the definition of minimum relief, and the general methods of the studies. A considerable degree of caution is needed when MCID is used as a treatment effect unit in meta-analyses of continuous outcomes [9]. Our overview may be a useful starting point for individual clinicians, researchers, guideline developers, or consensus building committees having to decide on a MCID for chronic pain in a given clinical setting.

A similar degree of caution is warranted when MCID informs calculations of the target sample size for clinical studies, for example randomised clinical trials. The logistical impact and associated cost of a large sample size in a clinical trial is very considerable and if a MCID of 12 mm is chosen, the study will require four times the number of included patients compared to 24 mm. A related challenge occurs when the GRADE approach is used to assess imprecision in meta-analyses [93]. Calculation of the optimal information size suggested for grading imprecision builds on defining an effect size considered clinically important.

It appears likely that the challenges described in pain assessment are not isolated to this specific research field but present a more general challenge with empirical assessment of MCID. Apart from the large variability between potential reference values, a number of additional methodological challenges remain for the current use of MCID. In a systematic review of clinical trials with pain as primary outcome, Ruyssen-Witrand found that the many different terms and definitions of MCID were most often being inappropriately used both for calculating sample size and interpreting results [94]. The main reasons for mistakes were confusion of the various concepts of change, such as failing to differentiate between statistical detectable and clinically important change, and confusion of change at the individual and group level.
Furthermore, the decision on whether an observed change exceeds the minimum clinically relevant difference often fails to consider the number of patients that has not reached the relevant change. It is largely overlooked that MCID estimated as a mean change in pain score will not apply to all individuals in the group, since their change scores are distributed around the mean [7]. In comparison, threshold values are derived with the intention of optimizing sensitivity and specificity, but the risk of false-positive (i.e. patients incorrectly classified as having relevant change) and false-negative results (i.e. patients incorrectly classified as not having relevant change) should still be considered. Consequently, the approach of “responder analysis”, which focuses on the percentage of patients that have reached a relevant change, has been recommended as a readily interpretable measure that is more relevant for both researchers and clinicians and less likely to mislead [14,95,96].

Publication of further studies of absolute MCID without reporting baseline pain and relative changes is hardly informative. For future studies there is a clear need for uniform guidelines for research conduct, analysis and reporting; especially for how transition scales and questions are structured, and how data are analysed and presented. It is noteworthy that only 17 of 66 studies (25%) reported MCID as relative changes and nine studies (14%) did not report the pain level at baseline at all. We encourage standardizing the definition of relevant pain relief, always reporting and adjusting for baseline pain, and reporting both relative and absolute values. In addition, since the influence of clinical and methodological factors is difficult to analyze from aggregated data, improved access to individual patient data would enable further exploration of the causes of heterogeneity between studies.

Conclusions

MCID in chronic pain varies greatly between studies. Baseline pain is strongly associated with absolute, but not relative, measures. To a much lesser degree, MCID is also influenced by the operational definition of relevant pain relief and possibly by clinical condition. MCID in chronic pain is central for clinical guideline
development, interpretation of results of randomised clinical trials or meta-analyses, and for choosing an appropriate sample size for a clinical study, but the measure is potentially misleading if estimated, applied or interpreted inappropriately. Explicit and conscientious reflections on the choice of a MCID value for chronic pain are required, when using it to classify research results as clinically important or trivial.

**Funding**

The study was funded by the Nordic Cochrane Centre.

**Ethics approval and consent to participate**

No ethics committee approval was needed for this systematic review.

**Availability of data and material**

The dataset analysed during the current study is available from the corresponding author upon request. The review protocol has been submitted to Journal of Clinical Epidemiology and can be obtained from the corresponding author upon request.

**Competing interests**

The authors declare that they have no competing interests.

**Author contributions**

AH, BT, and MFO conceptualised the study. MFO, EB, MDH, and BT contributed to the acquisition of data. Data was analysed by MFO with the assistance of JH. MFO, JH, and AH contributed to the interpretation of results. MFO drafted the manuscript, all co-authors contributed to reviewing the manuscript.
References


Figure 1: Flow chart of study identification

PubMed 1,090 records
EMBASE 141 records
Cochrane Library 627 records
Other sources\(^a\) 65 records

No of records after duplicates removed
1,553 records

Screening of titles and abstract
1,280 records excluded

No of full text records assessed
273 records

212 records excluded
- acute pain\(^b\) (36)
- multi-dimensional pain scale (53)
- MCID in pain not assessed (50)
- not empirical study (22)
- outcome not pain (12)
- healthy volunteer study (2)
- study population already included (3)
- physician-rated pain (2)
- distribution-based study (11)
- other method applied\(^c\) (21)

Included
66 studies described in 61 records

---

MCID = Minimum Clinically Important Difference, \(^a\) Additional records identified through “related papers” function in databases and reference lists of screened papers. \(^b\) Two records include both acute and chronic pain studies, \(^c\) Includes studies not based on subjective patient-reported assessment of pain relief e.g. use of individual goals, benefit-harm trade-off, acceptable symptom state, satisfaction with treatment, or objective anchor.
Table 1. Characteristics of studies assessing minimum clinically important difference in chronic pain (66 studies, 31,254 patients)

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Patient characteristics</th>
<th>Methodological characteristics</th>
<th>Minimum clinically important difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Clinical condition</td>
<td>Age</td>
<td>Male sex</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Transition scale based on pain relief**

- Beurskens 1996: 81 Low back pain, 41 y, 54%, 57 mm, 5 w, VAS symmetrical, 7 cat, 36 (29 to 43), -
- Eberle 1999: 190 Knee osteoarthritis, - - - 6 mo. VAS asymmetrical, 5 cat, 13 (7 to 18), -
- Emshoff 2010: 794 Temporomandibular disorder, 39 y, 9%, 50 mm, 3 mo. VAS symmetrical, 5 cat, 11 (10 to 11), 22 (21 to 24), 5 (97, 97), 6 (99, 98)
- Emshoff 2011: 678 Temporomandibular disorder, 40 y, 6%, 40 mm, 3 mo. VAS symmetrical, 5 cat, 11 (10 to 11), -
- Grotle 2004: 50 Low back pain, 40 y, 38%, 34 mm, 3 mo. VAS+ asymmetrical, 6 cat, 12 (-4 to 28), -
- Hanley 2006: 38 Phantom limb pain, 45 y, 91%, 38 mm, 6 w, NRS symmetrical, 5 cat, 16 (2 to 31), 39 (10 to 68), -
- Hanley 2006: 84 Spinal cord injury, 41 y, 79%, 53 mm, 6 w, NRS symmetrical, 5 cat, 19 (12 to 25), 34 (22 to 46), -
- Hagg 2003: 289 Low back pain, 43 y, 49%, 64 mm, 2 y, VAS asymmetrical, 4 cat, 21 (16 to 25), -
- ten Klooster 2006: 200 Arthritis, 60 y, 29%, 59 mm, 2 w, VAS asymmetrical, 5 cat, 37 (31 to 43), 56 (48 to 64), 30 (68, 84), 55 (74, 91)
- Mesrian 2007: 152 Back pain, 44 y, 46%, 51 mm, 6 mo, VAS symmetrical, 3 cat, 25 (20 to 30), -
- Perrot 2013: 3,329 Knee/hip osteoarthritis, 67 y, 50%, 51 mm, 1 w, NRS symmetrical 3 cat + asymmetrical 5 cat, -
- Salaffi 2004: 825 Osteoarthritis, rheumatoid arthritis and ankylosing spondylitis, 61 y (31-84), 45%, 59 mm, 3 mo, VAS asymmetrical, 5 cat, -
- de Vet 2015: 250 Sciatica, 42 y, 65%, 2 mo, VAS symmetrical, 7 cat, 52 (48 to 57), -

**Transition scale based on general health**

- Auffinger 2013: 88 Degeneration of cervical spine, 57 y, 48%, 29 mm, 6 mo, VAS symmetrical, 7 cat, 6^i, -
- Carreon 2010: 682 Degeneration of cervical spine, 53 y, 34%, 72 mm, 12 mo, NRS asymmetrical, 5 cat, 29 (25 to 33), -
- Childs 2005: 131 Low back pain, 34 y, 58%, 58 mm, 1 w, NRS symmetrical, 15 cat, 27 (23 to 31), -
- Cleland 2008: 138 Neck pain, 43 y, 52%, 48 mm, 2.5 d, NRS symmetrical, 15 cat, -
- Coelho 2008: 30 Low back pain, 38 y, 33%, 42 mm, 6 w, VAS asymmetrical, 7 cat, -
- Colangelo 2009: 280 Systemic lupus erythematosus, 50 y, 6%, 42 mm, 7.5 mo, VAS asymmetrical, 5 cat, 16 (10 to 22), -
- Copay 2008: 948 Lumbar spine patients (spinal stenosis etc.), 54 y, 41%, 68 mm, 1 y, NRS asymmetrical, 5 cat, 29 (22 to 35), -
- Crossley 2004: 71 Patellofemoral pain, 28 y, 30%, - 6 w, VAS symmetrical, 5 cat, 20^i [15 to 30], -
- Dhanani 2002: 533 Rheumatology conditions (eg. rheumatoid arthritis), 11 y (1-19), 30%, - 3 mo, VAS symmetrical, 5 cat, 8 (5 to 12), -
<table>
<thead>
<tr>
<th>Year</th>
<th>ID</th>
<th>Title</th>
<th>Age</th>
<th>Sex</th>
<th>Pain Type</th>
<th>Duration</th>
<th>Type</th>
<th>Severity</th>
<th>Pain Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunkl 2000</td>
<td>119</td>
<td>Fibromyalgi</td>
<td>47 y (27-61)</td>
<td>8 %</td>
<td>69 mm</td>
<td>6 mo.</td>
<td>NRS</td>
<td>symmetrical, 3 cat</td>
<td>24 (15 to 32)</td>
</tr>
<tr>
<td>Farrar 2001</td>
<td>2,879</td>
<td>Diabetic peripheral neuropathy, postherpetic neuralgia, osteoarthritis, etc.</td>
<td>58 y</td>
<td>48 %</td>
<td>68 mm</td>
<td>7.5 w</td>
<td>NRS</td>
<td>symmetrical, 7 cat</td>
<td>13° [3 to 24]</td>
</tr>
<tr>
<td>Farrar 2010</td>
<td>1,700</td>
<td>Diabetic peripheral neuropathy or fibromyalgia</td>
<td>56 y</td>
<td>39 %</td>
<td>59 mm</td>
<td>3 mo.</td>
<td>NRS</td>
<td>symmetrical, 7 cat</td>
<td>14° [0 to 30]</td>
</tr>
<tr>
<td>Forouzanfar 2003</td>
<td>61</td>
<td>Complex regional pain syndrome type 1</td>
<td>43 y</td>
<td>31 %</td>
<td>71 mm</td>
<td>6 mo.</td>
<td>VAS</td>
<td>symmetrical, 7 cat</td>
<td>30 °</td>
</tr>
<tr>
<td>Giradeau 2004</td>
<td>85</td>
<td>Sciatica</td>
<td>41 y</td>
<td>60 %</td>
<td>58 mm</td>
<td>20 d</td>
<td>VAS</td>
<td>asymmetrical, 4 cat</td>
<td>-</td>
</tr>
<tr>
<td>Glassman 2008</td>
<td>357</td>
<td>Spinal stenosis, spondylolisthesis, etc. (back pain)</td>
<td>54 y (23-83)</td>
<td>-</td>
<td>76 mm</td>
<td>1 y</td>
<td>NRS</td>
<td>symmetrical, 5 cat</td>
<td>-</td>
</tr>
<tr>
<td>Gum 2013</td>
<td>722</td>
<td>Spinal stenosis, spondylolisthesis, etc (leg pain)</td>
<td>61 y</td>
<td>34 %</td>
<td>74 mm</td>
<td>1 y</td>
<td>NRS</td>
<td>symmetrical, 5 cat</td>
<td>28 (25 to 31)</td>
</tr>
<tr>
<td>Hirschfeld 2014</td>
<td>204</td>
<td>Mixed pain in adolescents</td>
<td>16 y</td>
<td>25 %</td>
<td>-</td>
<td>3 w</td>
<td>NRS</td>
<td>symmetrical, 5 cat</td>
<td>12 (8 to 16)</td>
</tr>
<tr>
<td>Kovacs 2007</td>
<td>658</td>
<td>Low back pain</td>
<td>54 y</td>
<td>32 %</td>
<td>75 mm</td>
<td>3 mo.</td>
<td>NRS</td>
<td>asymmetrical, 4 cat</td>
<td>44 (42 to 46)</td>
</tr>
<tr>
<td>Kovacs 2008</td>
<td>1,349</td>
<td>Neck pain</td>
<td>54 y</td>
<td>23 %</td>
<td>72 mm</td>
<td>3 mo.</td>
<td>NRS</td>
<td>asymmetrical, 4 cat</td>
<td>41 (39 to 43)</td>
</tr>
<tr>
<td>Kvamme 2010</td>
<td>4,036</td>
<td>Rheumatoid arthritis</td>
<td>55 y</td>
<td>27 %</td>
<td>49 mm</td>
<td>3 mo.</td>
<td>VAS</td>
<td>symmetrical, 2 cat</td>
<td>-</td>
</tr>
<tr>
<td>Kwok 2010</td>
<td>223</td>
<td>Psoriatic arthritis</td>
<td>51 y</td>
<td>42 %</td>
<td>41 mm</td>
<td>8 mo</td>
<td>VAS</td>
<td>symmetrical, 5 cat</td>
<td>9 (1 to 18)</td>
</tr>
<tr>
<td>Lauche 2013</td>
<td>200</td>
<td>Neck pain</td>
<td>51 y (18-75)</td>
<td>21 %</td>
<td>44 mm</td>
<td>4/18 d</td>
<td>VAS</td>
<td>symmetrical, 5 cat</td>
<td>19 (15 to 23)</td>
</tr>
<tr>
<td>Lauridsen 2006</td>
<td>147</td>
<td>Low back and/or leg pain</td>
<td>46 y (18-85)</td>
<td>47 %</td>
<td>46 mm</td>
<td>2 mo</td>
<td>NRS</td>
<td>asymmetrical 7 cat + asymmetrical 4 cat</td>
<td>31 (27 to 36)</td>
</tr>
<tr>
<td>Lauridsen 2009</td>
<td>233</td>
<td>Low back and/or leg pain</td>
<td>46 y (19-82)</td>
<td>44 %</td>
<td>62 mm</td>
<td>2 mo</td>
<td>NRS</td>
<td>symmetrical, 7 cat</td>
<td>-</td>
</tr>
<tr>
<td>Maunton 2006</td>
<td>68</td>
<td>Spinal stenosis, spondylolisthesis, etc</td>
<td>53 y</td>
<td>46 %</td>
<td>70 mm</td>
<td>6 mo</td>
<td>VAS</td>
<td>asymmetrical, 5 cat</td>
<td>8 (-1 to 17)</td>
</tr>
<tr>
<td>Maughan 2010</td>
<td>63</td>
<td>Low back pain</td>
<td>52 y (25-78)</td>
<td>33 %</td>
<td>50 mm</td>
<td>5 w</td>
<td>NRS</td>
<td>symmetrical, 7 cat</td>
<td>20 (11 to 29)</td>
</tr>
<tr>
<td>Parker 2011</td>
<td>45</td>
<td>Lumbal spondylolisthesis (leg pain)</td>
<td>51 y</td>
<td>42 %</td>
<td>73 mm</td>
<td>2 y</td>
<td>VAS</td>
<td>asymmetrical, 4 cat</td>
<td>31 (20 to 42)</td>
</tr>
<tr>
<td>Parker 2012a</td>
<td>53</td>
<td>Lumbar stenosis (leg pain)</td>
<td>56 y</td>
<td>34 %</td>
<td>95 mm</td>
<td>2 y</td>
<td>VAS</td>
<td>asymmetrical, 4 cat</td>
<td>75 (67 to 83)</td>
</tr>
<tr>
<td>Parker 2012b</td>
<td>50</td>
<td>Adjacent-segment disease (leg pain)</td>
<td>59 y</td>
<td>42 %</td>
<td>87 mm</td>
<td>2 y</td>
<td>VAS</td>
<td>asymmetrical, 4 cat</td>
<td>43</td>
</tr>
<tr>
<td>Parker 2013a</td>
<td>69</td>
<td>Cervical radiculopathy (arm pain)</td>
<td>49 y</td>
<td>54 %</td>
<td>59 mm</td>
<td>3 mo</td>
<td>VAS</td>
<td>asymmetrical, 4 cat</td>
<td>42 (36 to 48)</td>
</tr>
</tbody>
</table>

*Note: NRS = Numerical Rating Scale, VAS = Visual Analog Scale*
<table>
<thead>
<tr>
<th>Reference</th>
<th>Age</th>
<th>Diagnosis</th>
<th>Male (Female)</th>
<th>Pain Duration (Range)</th>
<th>Pain Assessment</th>
<th>Symmetry</th>
<th>Absolute MCID (95% CI)</th>
<th>Relative MCID (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parker 2013b</td>
<td>50</td>
<td>Chiari malformation I (head pain)</td>
<td>39 y (28%) 81 mm 12 mo NRS symmetrical, 5 cat</td>
<td>54 (46 to 62) 6</td>
<td>44</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pool 2007</td>
<td>183</td>
<td>Neck pain</td>
<td>46 y 39% 60 mm 7 w NRS asymmetrical, 6 cat</td>
<td>16 (10 to 22)</td>
<td>25 (80, 80)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pope 2009</td>
<td>347</td>
<td>Rheumatoid arthritis</td>
<td>61 y 17% 42 mm 7 mo VAS symmetrical, 5 cat</td>
<td>12 (1 to 23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reddy 2013</td>
<td>60</td>
<td>Trigeminal neuralgia</td>
<td>53 y 22% 99 mm 2 y, 2 mo VAS asymmetrical, 5 cat</td>
<td>82 (75 to 88)</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reddy 2014</td>
<td>43</td>
<td>Trigeminal neuralgia</td>
<td>69 y 33% 98 mm 3 y VAS asymmetrical, 5 cat</td>
<td>82 (74 to 90)</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van der Roer 2006a</td>
<td>304</td>
<td>Low back pain &lt; 3 mo</td>
<td>47 y 49% 65 mm 3 mo NRS asymmetrical, 6 cat</td>
<td>49 (46 to 52)</td>
<td>35 (72, 88)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van der Roer 2006b</td>
<td>138</td>
<td>Low back pain &gt; 3 mo</td>
<td>44 y 41% 60 mm 3 mo NRS asymmetrical, 6 cat</td>
<td>41 (36 to 46)</td>
<td>25 (77, 82)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saps 2015</td>
<td>83</td>
<td>Paediatric functional abdominal pain</td>
<td>(8-17 y) 30% - 1 mo VAS symmetrical, 3 cat</td>
<td>20 (15 to 25)</td>
<td>27 (83, 79)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sekhon 2010</td>
<td>148</td>
<td>Scleroderma</td>
<td>57 y 16% 41 mm 7.5 mo VAS symmetrical, 5 cat</td>
<td>8 (15 to 31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheldon 2008</td>
<td>644</td>
<td>Low back pain</td>
<td>52 y 39% 77 mm 3 mo VAS asymmetrical, 5 cat</td>
<td>45</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solberg 2013</td>
<td>894</td>
<td>Lumbar disc herniation (back pain)</td>
<td>46 y 59% 60 mm 1 y NRS symmetrical, 7 cat</td>
<td>10 (0 to 15)</td>
<td>25 (74, 77)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strand 2008</td>
<td>40</td>
<td>Musculoskeletal pain</td>
<td>51 y 22% 60 mm 3.5 w VAS symmetrical, 7 cat</td>
<td>14 (6 to 21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strand 2008</td>
<td>28</td>
<td>Inflammatory rheumatic pain</td>
<td>51 y 42% 57 mm 3.5 w VAS symmetrical, 7 cat</td>
<td>22 (14 to 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strand 2008</td>
<td>69</td>
<td>Knee/hip osteoarthritis</td>
<td>74 y 25% 55 mm 3.5 mo VAS symmetrical, 7 cat</td>
<td>39 (31 to 46)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tashjian 2009</td>
<td>81</td>
<td>Rotator cuff disease</td>
<td>51 y 48% 60 mm 3.6 mo VAS asymmetrical, 4 cat</td>
<td>30 (24 to 37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tubach 2005</td>
<td>211</td>
<td>Knee osteoarthritis</td>
<td>68 y 30% 59 mm 1 mo VAS asymmetrical, 5 cat</td>
<td>- 20 (75, -) 41 (75, -)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tubach 2005</td>
<td>603</td>
<td>Hip osteoarthritis</td>
<td>65 y 37% 57 mm 1 mo VAS asymmetrical, 5 cat</td>
<td>- 15 (75, -) 32 (75, -)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tubach 2012</td>
<td>1,532</td>
<td>Mixed</td>
<td>57 y 30% - 1 mo NRS symmetrical 3 cat + asymmetrical 4 cat</td>
<td>- 15 (75, -) 20 (75, -)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Vet 2007</td>
<td>500</td>
<td>Low back pain</td>
<td>- - 3 mo NRS asymmetrical, 6 cat</td>
<td>18 (13 to 23) 25 (81, 78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheaton 2010</td>
<td>211</td>
<td>Spondyloarthropathies</td>
<td>45 y 69% 46 mm 5.6 mo VAS symmetrical, 5 cat</td>
<td>7 (-2 to 16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young 2010</td>
<td>184</td>
<td>Cervical radiculopathy</td>
<td>49 y 35% 66 mm 1 mo NRS symmetrical, 13/15 cat</td>
<td>28 (20 to 36) 13 (38, 85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Minimum Clinically Important Difference assessed as a mean change or as a threshold, VAS=visual analog scale, NRS=numeric rating scale, -- not reported, "mean (range), "symmetry of transition scale refers to whether there is an equal number of response categories for improvement and worsening of pain, "absolute = mean difference associated with a minimum improvement, "relative = % of baseline pain, "data is median (IQR), "95% CI unavailable, "95% CI estimated from imputed standard variation, "MCID reported as a range, "MCID reported for subgroups with different baseline pain.
Figure 2a. Absolute MCID in chronic pain relief assessed as a mean change* (47 studies, 12,779 patients)

* MCID assessed as the mean change in pain score among patients with minimum improvement of pain. MCID = Minimum Clinically Important Difference (mm reduction on a 100 mm scale). Stratified by concept of transition scale: pain relief or general health.
Figure 2b. Relative MCID in chronic pain relief assessed as a mean change* (5 studies, 1,200 patients)

* MCID assessed as the mean change in pain score among patients with minimum improvement of pain, MCID = Minimum Clinically Important Difference (% reduction from baseline). Stratified by concept of transition scale: pain relief or general health.
Figure 3a. Absolute MCID in chronic pain relief assessed as a threshold* (43 studies, 26,673)

* MCID assessed as the threshold value for pain score change which most accurately (best sensitivity and specificity) identified patients with relevant pain relief, MCID = Minimum Clinically Important Difference (mm reduction on a 100 mm scale), sensitivity = percentage of responders correctly classified as such, specificity = percentage of non-responders correctly classified as such, NA = not available. Stratified by concept of transition scale: pain relief or general health.
Figure 3b. Relative MCID in chronic pain relief assessed as a threshold (15 studies, 9,836 patients)

* MCID assessed as the threshold value for pain score change which most accurately (best sensitivity and specificity) identified patients with relevant pain relief. MCID = Minimum Clinically Important Difference (% reduction from baseline), sensitivity = percentage of responders correctly classified as such, specificity = percentage of non-responders correctly classified as such, NA = not available. Stratified by concept of transition scale: pain relief or general health.
Table 2a. Minimum clinically important difference for pain relief, median and pooled estimates

<table>
<thead>
<tr>
<th>Range</th>
<th>Analysis of medians</th>
<th>Analysis of pooled average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of studies (patients$^b$)</td>
<td>MCID, median [IQR]</td>
</tr>
<tr>
<td>Absolute change, mm on 100 mm scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean change approach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>-1 to 82</td>
<td>50 (13,561)</td>
</tr>
<tr>
<td>Transition scale: Pain relief</td>
<td>11 to 52</td>
<td>11 (2,806)</td>
</tr>
<tr>
<td>Transition scale: General health</td>
<td>-1 to 82</td>
<td>39 (10,755)</td>
</tr>
<tr>
<td>Threshold approach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>2 to 60</td>
<td>43 (26,673)</td>
</tr>
<tr>
<td>Transition scale: Pain relief</td>
<td>2 to 34</td>
<td>7 (5,522)</td>
</tr>
<tr>
<td>Transition scale: General health</td>
<td>8 to 60</td>
<td>36 (21,151)</td>
</tr>
<tr>
<td>Relative change, % of baseline pain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean change approach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>10 to 56</td>
<td>7 (1,465)</td>
</tr>
<tr>
<td>Transition scale: Pain relief</td>
<td>22 to 56</td>
<td>4 (1,000)</td>
</tr>
<tr>
<td>Transition scale: General health</td>
<td>10 to 45</td>
<td>3 (465)</td>
</tr>
<tr>
<td>Threshold approach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>6 to 55</td>
<td>15 (9,836)</td>
</tr>
<tr>
<td>Transition scale: Pain relief</td>
<td>6 to 55</td>
<td>3 (1,703)</td>
</tr>
<tr>
<td>Transition scale: General health</td>
<td>13 to 50</td>
<td>12 (8,133)</td>
</tr>
</tbody>
</table>

MCID = Minimum Clinically Important Difference, IQR = Inter-quartile range, NA = not applicable, $^a$ The median is based on studies included in the pooled average as well as studies with unavailable standard errors, $^b$ Total number of patients in the included studies, $^c$ $I^2$ = percentage of the variability in results that is due to heterogeneity rather than sampling error (chance): $I^2$ of 0% to 40% might not be important, 30% to 60% may represent moderate heterogeneity, 50% to 90% may represent substantial heterogeneity and 75% to 100% represents considerable heterogeneity.
**Table 2b. Supplementary outcomes, median and pooled estimates**

<table>
<thead>
<tr>
<th>Clinically important difference</th>
<th>Range</th>
<th>Analysis of medians a</th>
<th></th>
<th>Analysis of pooled average</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Number of studies (patients^b)</td>
<td>Median [IQR]</td>
<td>Number of studies (patients^b)</td>
<td>Pooled average (95% CI), I^2 c</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Substantial clinically important difference for pain relief</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute change, mm on 100 mm scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean change approach</td>
<td>14 to 68</td>
<td>25 (10,722)</td>
<td>40 [29 to 44]</td>
<td>21 (8,611)</td>
<td>38 (31 to 45), 98%</td>
</tr>
<tr>
<td>Threshold approach</td>
<td>17 to 35</td>
<td>7 (7,148)</td>
<td>22 [20 to 27]</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Relative change, % of baseline pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean change approach</td>
<td>50 to 86</td>
<td>6 (1,700)</td>
<td>70 [62 to 78]</td>
<td>3 (1,078)</td>
<td>78 (67 to 89), 93%</td>
</tr>
<tr>
<td>Threshold approach</td>
<td>28 to 67</td>
<td>5 (6,282)</td>
<td>34 [33 to 38]</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Minimum clinically important difference for pain worsening</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute change, mm on 100 mm scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean change approach</td>
<td>-19 to 7</td>
<td>23 (10,232)</td>
<td>-9 [-12 to -4]</td>
<td>21 (9,442)</td>
<td>-9 (-11 to -7), 73%</td>
</tr>
<tr>
<td>Relative change, % of baseline pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean change approach</td>
<td>-36 to -5</td>
<td>4 (1,143)</td>
<td>-19 [-32 to -8]</td>
<td>2 (878)</td>
<td>-28 (-35 to -21), 0%</td>
</tr>
<tr>
<td><strong>Substantial clinically important difference for pain worsening</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute change, mm on 100 mm scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean change approach</td>
<td>-37 to 9</td>
<td>13 (6,995)</td>
<td>-17 [-23 to -10]</td>
<td>12 (6,273)</td>
<td>-20 (-24 to -16), 84%</td>
</tr>
<tr>
<td>Relative change, % of baseline pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean change approach</td>
<td>-59 to -25</td>
<td>2 (739)</td>
<td>-42 [-39 to -25]</td>
<td>1 (678)</td>
<td>NA</td>
</tr>
</tbody>
</table>

IQR = Inter-quartile range, NA = not applicable, a The median is based on studies included in the pooled average as well as studies with unavailable standard errors, b Total number of patients in the included studies, c I^2 = percentage of the variability in results that is due to heterogeneity rather than sampling error (chance). I^2 of 0% to 40% might not be important, 30% to 60% may represent moderate heterogeneity, 50% to 90% may represent substantial heterogeneity and 75% to 100% represents considerable heterogeneity.
Figure 4a. MCID assessed as mean changes by baseline pain (40 studies, 10,938 patients)

MCID = Minimum Clinically Important Difference (mm reduction on a 100 mm scale). Plot includes studies with available data on mean baseline pain of study population. Bubbles are sized according to the precision of each estimate (the inverse of its within-study variance).

Figure 4b. MCID assessed as thresholds by baseline pain (37 studies, 23,926 patients)

MCID = Minimum Clinically Important Difference (mm reduction on a 100 mm scale). Data on precision of each estimate is not available for threshold results.
### Table 3. Subgroup analysis of absolute MCID for pain relief assessed as mean changes, summary table

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Range mm</th>
<th>Number of studies (patients)</th>
<th>MCID median mm [IQR]</th>
<th>Number of studies (patients)</th>
<th>MCID pooled average mm (95% CI), I²</th>
<th>Meta-regression, adjusted for baseline pain</th>
</tr>
</thead>
<tbody>
<tr>
<td>All studies</td>
<td>-1 to 82</td>
<td>50 (13,561)</td>
<td>23 [12 to 39]</td>
<td>47 (12,779)</td>
<td>27 (23 to 32), 99%</td>
<td>40 (10,938)</td>
</tr>
<tr>
<td>Clinical conditions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-specific back pain</td>
<td>-1 to 49</td>
<td>13 (3,964)</td>
<td>27 [20 to 41]</td>
<td>12 (3,320)</td>
<td>28 (21 to 35), 97%</td>
<td>11 (2,820)</td>
</tr>
<tr>
<td>Back pain w. radiculopathy/stenosis</td>
<td>7 to 75</td>
<td>10 (3,125)</td>
<td>28 [10 to 43]</td>
<td>9 (3,075)</td>
<td>29 (16 to 41), 97%</td>
<td>9 (3,075)</td>
</tr>
<tr>
<td>Neck pain</td>
<td>6 to 42</td>
<td>7 (2,064)</td>
<td>28 [16 to 41]</td>
<td>6 (1,976)</td>
<td>29 (20 to 39), 96%</td>
<td>5 (838)</td>
</tr>
<tr>
<td>Arthritis</td>
<td>9 to 39</td>
<td>6 (1,028)</td>
<td>17 [12 to 37]</td>
<td>6 (1,028)</td>
<td>22 (11 to 33), 93%</td>
<td>6 (1,976)</td>
</tr>
<tr>
<td>Musculoskeletal pain</td>
<td>11 to 30</td>
<td>5 (1,741)</td>
<td>14 [11 to 24]</td>
<td>5 (1,741)</td>
<td>14 (12 to 17), 91%</td>
<td>4 (1,660)</td>
</tr>
<tr>
<td>Paediatric conditions</td>
<td>8 to 20</td>
<td>3 (820)</td>
<td>12 [8 to 20]</td>
<td>3 (820)</td>
<td>13 (7 to 19), 86%</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Other</td>
<td>8 to 20</td>
<td>6 (819)</td>
<td>34 [16 to 82]</td>
<td>6 (819)</td>
<td>43 (19 to 68), 99%</td>
<td>5 (569)</td>
</tr>
<tr>
<td>Concept of transition scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Based on pain relief</td>
<td>11 to 52</td>
<td>11 (2,806)</td>
<td>19 [12 to 36]</td>
<td>11 (2,806)</td>
<td>23 (19 to 28), 98%</td>
<td>9 (2,366)</td>
</tr>
<tr>
<td>Based on general improvement</td>
<td>-1 to 82</td>
<td>39 (10,755)</td>
<td>27 [12 to 41]</td>
<td>36 (9,973)</td>
<td>29 (23 to 34), 98%</td>
<td>31 (8,572)</td>
</tr>
<tr>
<td>Symmetry of transition scale m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symmetrical</td>
<td>6 to 54</td>
<td>25 (5,881)</td>
<td>19 [11 to 27]</td>
<td>24 (5,793)</td>
<td>22 (19 to 26), 97%</td>
<td>21 (5,256)</td>
</tr>
<tr>
<td>Asymmetrical</td>
<td>-1 to 82</td>
<td>25 (7,680)</td>
<td>31 [16 to 43]</td>
<td>23 (6,986)</td>
<td>33 (26 to 40), 98%</td>
<td>19 (5,682)</td>
</tr>
<tr>
<td>Definition of relevant pain relief n</td>
<td>-1 to 44</td>
<td>27 (10,485)</td>
<td>13 [9 to 25]</td>
<td>26 (10,397)</td>
<td>18 (14 to 23), 99%</td>
<td>21 (8,887)</td>
</tr>
<tr>
<td>One category improvement</td>
<td>14 to 82</td>
<td>18 (2,604)</td>
<td>40 [27 to 52]</td>
<td>16 (1,910)</td>
<td>42 (32 to 51), 98%</td>
<td>15 (1,660)</td>
</tr>
<tr>
<td>Several categories improvement</td>
<td>16 to 42</td>
<td>5 (472)</td>
<td>30 [19 to 37]</td>
<td>5 (472)</td>
<td>30 (20 to 39), 88%</td>
<td>4 (391)</td>
</tr>
<tr>
<td>Meaningful change</td>
<td>-1 to 82</td>
<td>31 (6,077)</td>
<td>20 [11 to 39]</td>
<td>28 (5,295)</td>
<td>27 (22 to 32), 99%</td>
<td>23 (4,158)</td>
</tr>
<tr>
<td>Pain scale</td>
<td>10 to 54</td>
<td>19 (7,484)</td>
<td>28 [18 to 41]</td>
<td>19 (7,484)</td>
<td>28 (23 to 34), 97%</td>
<td>17 (6,780)</td>
</tr>
<tr>
<td>Follow-up time</td>
<td>-1 to 52</td>
<td>27 (7,778)</td>
<td>20 [14 to 41]</td>
<td>26 (7,134)</td>
<td>25 (20 to 31), 99%</td>
<td>21 (5,564)</td>
</tr>
<tr>
<td>≤ 3 months</td>
<td>6 to 82</td>
<td>23 (5,783)</td>
<td>25 [10 to 39]</td>
<td>21 (5,645)</td>
<td>30 (22 to 39), 98%</td>
<td>19 (5,374)</td>
</tr>
<tr>
<td>&gt; 3 months</td>
<td>-1 to 54</td>
<td>30 (9,533)</td>
<td>19 [12 to 31]</td>
<td>29 (8,889)</td>
<td>24 (17 to 30), 99%</td>
<td>22 (7,048)</td>
</tr>
<tr>
<td>Risk of non-representative sampling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>6 to 82</td>
<td>20 (4,028)</td>
<td>29 [13 to 40]</td>
<td>18 (3,890)</td>
<td>34 (23 to 44), 99%</td>
<td>18 (3,890)</td>
</tr>
<tr>
<td>High or unclear</td>
<td>-1 to 54</td>
<td>30 (9,533)</td>
<td>19 [12 to 31]</td>
<td>29 (8,889)</td>
<td>24 (17 to 30), 99%</td>
<td>22 (7,048)</td>
</tr>
<tr>
<td>Risk of attrition bias</td>
<td>-1 to 82</td>
<td>21 (5,046)</td>
<td>36 [20 to 44]</td>
<td>18 (4,264)</td>
<td>37 (28 to 45), 98%</td>
<td>15 (3,398)</td>
</tr>
<tr>
<td>Low</td>
<td>7 to 54</td>
<td>29 (8,515)</td>
<td>18 [12 to 29]</td>
<td>29 (8,515)</td>
<td>22 (18 to 26), 98%</td>
<td>25 (7,540)</td>
</tr>
</tbody>
</table>

**MCID** = Minimum Clinically Important Difference (mm reduction on a 100 mm scale), **IQR** = inter-quartile range, **NA** = Not applicable, "The median is based on studies included in the pooled average as well as studies with unavailable standard errors," total number of patients in studies, "I²" = percentage of the variability in results that is due to heterogeneity rather than sampling error (chance). I² of 0% to 40% might not be important, 30% to 60% may represent moderate heterogeneity, 50% to 90% may represent substantial heterogeneity and 75% to 100% represents considerable heterogeneity.

* Includes cervical and lumbar back pain
Includes chiari malformation I, spondylolisthesis, spinal stenosis, spinal cord injury, spondyloarthropathies, adjacent-segment disease, disc herniation, and failed back

Includes cervical radiculopathy, degenerative cervical spine disease, spondylosis, disc herniation, stenosis, and non-specific neck pain.

Includes arthritis, psoriatic arthritis, hip or knee osteoarthritis, inflammatory rheumatic pain, and rheumatoid arthritis

Includes musculoskeletal pain, temporomandibular disorder, fibromyalgia, and rotator cuff syndrome.

Includes paediatric rheumatology patients (mostly rheumatoid arthritis), adolescents with chronic pain (head, back/extremities, abdomen), and paediatric functional abdominal pain

Includes phantom limb pain, scleroderma, systemic lupus erythematosus, trigeminal neuralgia, and sciatica.

Includes musculoskeletal pain, temporomandibular disorder, and fibromyalgia

Includes phantom limb pain, scleroderma, systemic lupus erythematosus, and trigeminal neuralgia.

Symmetry of transition scale refers to whether there is an equal number of response categories for improvement and worsening of pain

Definition of relevant pain relief refers to whether MCID was defined as a one-category improvement, combination of several categories or whether patients were asked to distinguish between meaningful and non-meaningful pain relief.