HHS Public Access

Author manuscript

Pain. Author manuscript; available in PMC 2026 May 09.

Published in final edited form as:

Pain. 2025 October 01; 166(10): 2430–2438. doi:10.1097/j.pain.0000000000003645.

Minimum clinically important differences in acute pain: a patient-level re-analysis of randomized controlled analgesic trials submitted to the US Food and Drug Administration

Christopher J. Miller[a,*], **Warren B. Bilker**[a], **Ian DeLorey**[a], **Charles E. Argoff**[b], **Russell L. Bell**[c], **Andrew Conroy**[a], **Jennifer S. Gewandter**[d], **Ian Gilron**[e,f,g,h], **Jennifer A. Haythornthwaite**[i], **Nathaniel P. Katz**[j], **Tara McWilliams**[a], **Katherine N. Theken**[k], **John T. Farrar**[a]

[a.]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

[b.]Department of Neurology, Albany Medical Center, Albany, NY

[c.]Department of Anesthesiology and Critical Care, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

[d.]Department of Anesthesiology and Perioperative Medicine, University of Rochester School of Medicine and Dentistry, Rochester, NY

[e.]Department of Anesthesiology and Perioperative Medicine, Queen's University, Kingston, Canada

[f.]Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Canada

[g.]Center for Neuroscience Studies, Queen's University, Kingston, Canada

[h.]School of Policy Studies, Queen's University, Kingston, Canada

[i.]Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD

[j.]Ein Sof Innovation, Boston, MA

[k.]Department of Oral and Maxillofacial Surgery / Pharmacology, School of Dental Medicine, University of Pennsylvania, Philadelphia, PA

Abstract

The lack of established minimum clinically important differences (MCIDs) in acute pain has made it challenging to interpret efficacy in analgesic trials. We performed a patient-level re-analysis of double-blind, placebo-controlled trials submitted to the US Food and Drug Administration to estimate MCIDs in acute postoperative pain. Trials were categorized by acute surgical pain model: dental extraction, bunionectomy, orthopedic surgery, and soft tissue surgery. Pain intensity was assessed using the 0–10 numeric rating scale (NRS) or 0–100 visual analog scale (VAS), with

[*]Corresponding author. *Address:* 423 Guardian Drive, Blockley Hall, Room 501A, Philadelphia, PA 19104; chris.miller1@pennmedicine.upenn.edu.

VAS scores converted to NRS for analysis. To avoid misclassification from arbitrary thresholds on global impression of change or pain relief scales, meaningful pain relief was determined using the double stopwatch technique, where patients actively indicated the times they experienced perceptible and meaningful relief. Across 29 trials, 9047 patients with moderate-to-severe baseline pain were included. Patients with severe baseline pain (NRS ≥7) reported meaningful relief at a higher absolute NRS and required larger absolute reductions in pain intensity than those with moderate baseline pain (NRS 4-<7). However, the percent reduction in pain at meaningful relief remained stable across baseline pain levels, suggesting patients assess meaningful relief in relative rather than absolute terms. No appreciable differences in the changes in pain at meaningful relief were observed by age, sex, drug, or route of administration. Receiver operating characteristic curve (ROC) analysis identified a 50% reduction in pain intensity as a consistent and clinically meaningful threshold across surgical pain models, supporting its use as a standardized patient-centric metric for evaluating analgesic efficacy.

1. Introduction

Analgesic trials are typically designed to detect mean differences in pain intensity between treatment groups [3,9]. However, statistically significant differences do not necessarily reflect clinical significance [2,4,13,17], and group averages offer limited insight into the likelihood that patients will experience meaningful pain relief. Despite the urgent need for continued development of effective alternatives to opioids for postsurgical pain management, patient-centric targets for meaningful changes in acute pain have not been clearly established [16,28].

Identifying minimum clinically important differences (MCIDs) in acute postoperative pain is challenging for both clinical and methodologic reasons [11,16,20]. The magnitude of reduction in pain intensity considered meaningful may be modulated by the extent and location of the surgical insult, baseline pain intensity, and other patient-specific factors [16,25]. Most prior studies attempting to identify MCIDs for postsurgical pain pooled data across surgical procedures or did not explore potential moderators [2,15,18,19,23], so it remains unclear whether surgery-specific or subgroup-specific MCIDs are warranted [16,28].

MCIDs are typically determined by either distribution- or anchor-based methods. Distribution-based approaches, which rely on statistical criteria, are generally discouraged because they are not grounded in patient perspectives [16]. Anchor-based methods link changes in pain intensity to a threshold on an external anchor, typically a pain relief scale or a global impression of change [4,11,16]. However, it must be presumed that the threshold selected on the anchor by the investigator accurately classifies meaningful change for all patients [4]. Additionally, when pain intensity and anchor measures are recorded infrequently, the true change in pain intensity associated with achieving the anchor threshold may be missed [5,12,23]. Some studies have instead used the choice to take or abstain from another dose of rescue medication as the anchor for defining an MCID [7,8]. While this approach is based on a deliberate patient action rather than a passive response on another scale, these studies were conducted in breakthrough cancer pain, which may not generalize

to postsurgical pain. Moreover, the decision to use rescue medication may be influenced by factors beyond pain intensity, such as safety or tolerability concerns.

In analgesic trials conducted for regulatory approval, the time to onset of pain relief is measured using the double stopwatch technique, in which patients press the first stopwatch when they experience perceptible relief and the second stopwatch when they experience meaningful relief [26]. This approach avoids the misclassification that comes from imposing arbitrary thresholds on an external anchor, as each patient directly indicates when their pain relief becomes meaningful. Despite the methodology's high precision and intuitive validity, it has not been previously used to determine MCIDs in acute pain.

Our study aimed to determine MCIDs in acute pain for the surgical pain models that have been used to evaluate the efficacy of new analgesics and to explore factors that may influence perceptions of meaningful relief. We analyzed patient-level data from randomized controlled analgesic trials submitted under New Drug Applications (NDAs) to the U.S. Food and Drug Administration (FDA) that collected pain intensity scores and measured meaningful pain relief using the double stopwatch technique.

2. Methods

Under contract from the FDA and with the approval of the University of Pennsylvania Institutional Review Board, we standardized and merged data from all Phase 2 or Phase 3 double-blind, randomized controlled trials of analgesics for the treatment of moderate-to-severe acute pain that were available in electronic format from the FDA Document Archiving, Reporting, and Regulatory Tracking System (DARRTS).

Trials were included if they collected numeric pain intensity scores, ascertained the onset of perceptible and meaningful pain relief using the double stopwatch method [6], and used a fixed dosing schedule (i.e., single-dose studies or multi-dose studies where drugs were administered at set intervals). Trials were excluded if the first pain intensity score taken after randomization occurred more than 15 minutes after dosing because these trials would fail to capture adequate pain intensity data for patients who achieved an early onset of analgesic effect.

Trials were aggregated for analysis according to the four major acute surgical pain models: dental extraction, bunionectomy, orthopedic surgery (e.g., joint replacement), and soft tissue surgery (e.g., abdominoplasty, hysterectomy). Singla and colleagues have previously published an extensive review comparing the clinical and experimental characteristics of these surgical models [21].

Standard approaches to calculating MCIDs could not be applied to this dataset due to the variable timing of pain assessments and the second stopwatch press. Because no method existed for analyzing double stopwatch data to estimate MCIDs in pain intensity, a pre-specified analytic protocol could not be defined. During the conduct of this study, we developed and published an analytic framework specifically to handle the unique aspects of double stopwatch data [14], which was then used to guide our analysis.

2.1. Meaningful Pain Relief

The double stopwatch technique was used in all trials to assess the onset of both perceptible and meaningful pain relief [6]. Before administering the study drug, patients were provided with two stopwatches. They were instructed to press the first stopwatch when they felt perceptible pain relief and the second stopwatch when they experienced pain relief that was meaningful to them.

Some trials permitted the measurement of onset times for pain relief after using rescue pain medication, while others did not. To ensure consistency across studies, we applied a standardized definition of failure to achieve meaningful relief. Failure was defined as the first occurrence of any of the following events before pressing the second stopwatch: taking rescue medication; taking a second dose of the study drug (in multi-dose studies); withdrawing from the study; or completing the study.

2.2. Pain Intensity

Pain intensity scores were ascertained using either the numeric rating scale (NRS; 0–10) or visual analog scale (VAS; 0–100). All trials took frequent scheduled pain intensity assessments, particularly within the first hour after the first administration of study medication, although the specific schedule varied by study (Supplemental Table 1). Unscheduled pain intensity assessments were taken before the use of rescue medication.

To be eligible for the study, each trial required patients to report moderate or severe pain intensity before randomization. These criteria involved reporting a pain intensity score above a specified threshold (i.e., NRS ≥4 or ≥5; VAS ≥40 or ≥50), a response of ‘moderate’ or ‘severe’ on a 4-point categorical pain scale, or both depending on the trials (Supplemental Table 1).

Since pain intensity is more severe and variable when measured with activity than at rest [9,10,24], we sought to confirm that all pain intensity scores included in our study were measured using a similar approach. Six trials explicitly measured pain intensity scores at rest. Due to the absence of protocol-specified procedures for evoking movement, it was assumed that all other trials measured pain intensity at rest.

For analyses of absolute pain intensity scores and absolute changes from baseline in pain intensity at meaningful relief, pain intensity scores reported on the VAS were converted to the NRS by dividing by 10. Subgroup analyses indicated that the distributions of pain intensity scores reported on the VAS and NRS were generally proportional (Supplementary Figure 2 A–D).

2.3. Outcomes

The primary outcome measures were the pain intensity score at the onset of meaningful relief and the corresponding absolute and percent changes in pain intensity from baseline. The onset of meaningful relief was defined as the moment the patient pressed the second stopwatch. Pain intensity scores were collected frequently at pre-specified time points in each study, but the onset of meaningful relief could occur between scheduled assessments.

In such cases, we used linear interpolation between the two nearest recorded pain intensity scores to estimate the pain intensity at the exact time of meaningful relief.

2.4. Statistical Methods

Our data analysis was based on methods developed for determining MCIDs in pain intensity with the double stopwatch technique, which have previously been described in detail [14]. Briefly, this approach conceptualizes the degree of pain relief that a patient would find meaningful after a given surgery as a latent (unobserved) variable that exists prior to randomized treatment assignment. If a patient experiences that level of relief, they press the stopwatch for meaningful relief and their value of meaningful change in pain is directly observed. If a patient does not experience that level of relief, they will take rescue medication or drop out before pressing the stopwatch for meaningful relief. In these cases, the patient's value of meaningful change is unobserved and is censored at their most favorable pain score prior to taking rescue or dropping out.

To estimate the distribution of meaningful changes in pain intensity, we applied an adaptation of the Kaplan-Meier method. Traditionally used for time-to-event data, the Kaplan-Meier method provides unbiased estimates of event-time distributions even when not all patients have experienced the event by the end of follow-up (i.e., at the censoring time). In the context of the double stopwatch, the "event" is the onset of meaningful pain relief and the corresponding change in pain intensity is used in place of "time." Patients who do not achieve meaningful pain relief after their first dose of study drug are censored at the most favorable change in pain intensity observed before their first failure event, as previously defined in Section 2.1.

Just as conventional survival analysis must assume that censored patients would eventually experience the event if follow-up had continued indefinitely, our approach assumes that patients censored for taking rescue medication or dropping out would have pressed the stopwatch if they had experienced sufficient relief (i.e., non-informative censoring). This adapted Kaplan-Meier method enables estimation of the underlying distribution of meaningful pain reductions even when some patients require rescue medication or drop out before achieving meaningful relief, as is common in acute pain studies.

Descriptive statistics were calculated for the distributions of NRS at meaningful pain relief and the absolute and percent change from baseline at meaningful pain relief, stratified by acute surgical pain model overall and in subgroup analyses by study, study drug, active drugs vs placebo, route of administration (oral or IV), age (<55 vs ≥55 years), sex (male vs female), and baseline pain intensity (moderate [<7] vs severe [≥7]). Data were summarized using medians and interquartile range (IQR; 25^{th} and 75^{th} percentiles).

An approach has also been developed for applying binary classification and receiver operating characteristic (ROC) curve methods to meaningful change scores identified using the double stopwatch technique [14]. For a patient who achieves meaningful pain relief, all changes in pain greater than or equal to their meaningful value are classified as meaningful, while all values from zero up to but not including their meaningful value are classified as not meaningful. For a patient who does not achieve meaningful relief, all changes from

zero to their most favorable change in pain observed before their failure event are classified as not meaningful. For a given threshold, each of those values of change in pain intensity can be further classified as a true positive, false positive, true negative, or false negative based on whether it met the threshold and whether the patient found that degree of change meaningful. To determine MCIDs, we considered changes in 1-point increments on the absolute NRS scale and 10 percentage-point increments on the percent scale. For additional methodological details, see the previously published approach [14].

To account for the correlation of multiple pain intensity values analyzed per patient, we used cluster bootstrapping with 2000 resamples to assess the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, and specificity at each threshold along with 95% confidence intervals (CIs). The MCID for each surgery was identified as the threshold with the highest AUC, providing optimal diagnostic accuracy. All analyses were performed using R version 4.3.1 (R Foundation for Statistical Computing).

3. Results

3.1. Analysis Dataset

Forty-four randomized controlled trials of oral or IV analgesic treatments for acute pain with a fixed dosing schedule were identified in DARRTS. Ten trials were not eligible because they did not use the double stopwatch method (n=8 trials) or did not ascertain pain intensity scores on the NRS or VAS (n=2 trials). Five trials were excluded for not measuring pain intensity within the first 15 minutes after dosing (see PRISMA diagram in Supplemental Figure 1). The most common surgical pain models were bunionectomy (n=12 trials) and dental extraction (n=11 trials). Two trials were performed in orthopedic surgery (n=1 in mixed orthopedic procedures and n=1 in knee arthroscopy), and four trials were performed in soft tissue surgery (n=2 in mixed abdominal procedures, n=1 in abdominoplasty, and n=1 in hysterectomy).

A total of 9115 patients were enrolled across the 29 trials. While all patients had to have a qualifying moderate or severe pain intensity score prior to randomization, 68 had a baseline pain intensity score immediately prior to study drug administration of <4 on the NRS scale and were excluded from the analyses, resulting in an analysis set of 9047 patients.

In all but one trial, patients were encouraged to wait for 30 to 90 minutes before requesting rescue medication to allow the study drugs to take effect (Supplemental Table 1), although rescue medication was available at any time. Additional details on study drugs, dosing regimens, enrollment by treatment arm, and the pain intensity assessment methodology for each study are presented in Supplemental Table 1.

As expected, patients who received a placebo were less likely to report meaningful pain relief than patients who received an active drug. However, because the amount of pain relief that would be considered meaningful is a latent, pre-existing characteristic, it was anticipated that randomization would balance its distribution across treatment arms. Consistent with this expectation, the changes in pain intensity scores that were considered meaningful were similar in the active drug and placebo groups (Supplemental Figure 2),

supporting the decision to pool analyses across all patients. The median time between the onset of meaningful pain relief and the closest recorded pain intensity score was 2 minutes (interquartile range [IQR], 1–5).

3.2. Participants

Most patients undergoing bunionectomy or soft tissue surgery were female, while dental extraction and orthopedic surgery had a more even balance of male and female patients (Table 1). The median age was 20 years in dental extraction and 43 to 50 years in the other surgical models. Severe pain (NRS ≥7) at baseline was less common among patients undergoing orthopedic surgery and soft tissue surgery. While orthopedic surgery is generally considered to be among the more painful surgical models, one of the two trials evaluated knee arthroscopy, which is less painful than joint replacement. All patients in the bunionectomy trials were enrolled in the US, while the other surgical pain models had patients from Europe, New Zealand, or both regions, as well as the US.

3.3. Changes in Pain Intensity at the Onset of Meaningful Pain Relief

The overall distributions of observed NRS pain intensity values at the onset of meaningful relief were similar across the four acute surgical pain models (Fig. 1). The overall median values ranged from 2.4 to 3.2, with IQRs spanning approximately 1 to 1.5 points above and below the median. However, the distributions of NRS scores at meaningful relief differed by baseline pain severity (Fig. 2A). Among patients with moderate baseline pain (NRS 4-<7), the median observed NRS at meaningful relief was approximately 2.5, whereas for patients with severe baseline pain (NRS ≥7), it ranged from 3.5 to 4.3 (Fig. 1).

A similar pattern was observed for absolute change in NRS from baseline at meaningful relief. The median absolute change across surgical models ranged from −2.9 to −3.6 (Fig. 1). However, patients with severe baseline pain required greater absolute reductions to achieve meaningful relief (Fig. 2B). Across all four surgical models, the median absolute changes ranged from −2.7 to −3.2 for patients with moderate baseline pain and from −3.8 to −4.6 for those with severe pain (Fig. 1).

In contrast, percent change in pain intensity from baseline at meaningful relief was more stable across baseline pain severities. Median values overall ranged from −48% to −56% across the surgical models (Fig. 1), and the distributions of percent changes in pain were similar for the moderate and severe subgroups (Fig. 2C), suggesting that patients assessed the meaningfulness of pain relief in relative rather than absolute terms.

The observed pain intensity at meaningful relief, as well as the absolute and percent changes from baseline, were generally consistent when stratified by age, sex, pain intensity measure (NRS or VAS), and study drug (active drug or placebo) in dental extraction, bunionectomy, and orthopedic surgery (Supplemental Figure 2). In soft tissue surgery, males reported meaningful pain relief with somewhat less improvement than females. However, this observation may be influenced by the limited sample size, as males only represented 10% of patients in that surgical model. A small difference in the distributions of meaningful pain scores was observed in soft tissue surgery between the NRS and VAS, likely reflecting a study effect, as only one of the four trials used the NRS while the others used the VAS.

3.4. Minimum Clinically Important Differences

The threshold for improvement on the absolute change in the NRS scale that optimized the AUC was 3 for all surgeries (Table 2). These MCID values demonstrated strong diagnostic performance, with AUCs ranging from 0.807 to 0.819. Sensitivity values, representing the probability of correctly identifying meaningful changes in pain intensity, ranged from 79.8% to 89.0%. Specificity values, representing the probability of correctly identifying non-meaningful changes, ranged from 72.3% to 84.0%. Complete operating characteristics and corresponding 95% CIs for all thresholds are provided in Supplemental Table 2.

On the percent scale, the threshold for improvement that optimized the AUC was 50% for dental extraction, orthopedic surgery, and soft tissue surgery, and 40% for bunionectomy (Table 2). These MCID values had excellent classification performance across surgical models, ranging from 0.811 to 0.828. Sensitivity values ranged from 82.7% to 86.0%, while specificity values ranged from 77.8% to 82.7%. For bunionectomy, the AUCs for the 40% and 50% thresholds were similar (0.819 vs 0.812), indicating that both thresholds had very good diagnostic properties. The magnitude of the trade-off between sensitivity and specificity was similar between the two thresholds, with the 40% threshold having higher sensitivity and the 50% threshold having higher specificity. Complete results are provided in Supplemental Table 3.

Across all acute surgical pain models, the optimal AUC was higher for thresholds based on percent change than absolute change. This aligns with our observation that meaningful change scores were more consistent across baseline pain levels on the percent scale, which likely contributed to the improved discriminatory performance.

4. Discussion

In this pooled analysis of 29 randomized controlled trials, we determined how changes in pain intensity scores corresponded to the onset of meaningful pain relief, measured using the double stopwatch technique. We found that the MCIDs in acute postsurgical pain for the four primary acute surgical pain models were 3 points on the NRS scale (30 points on the VAS scale) or 50% on the percent scale. In the bunionectomy model, a 40% threshold showed slightly higher diagnostic performance than 50%, but the difference was insufficient to warrant an alternate cutoff. There was minimal variation by age, sex, or analgesic route of administration. The lack of a difference by sex in all but one surgical model is noteworthy, given prior research showing higher pain sensitivity among female patients [1]. The consistently strong diagnostic performance of a 50% MCID across surgical pain models supports its use as a standardized benchmark for evaluating analgesic efficacy in acute surgical pain trials.

A key finding of this study was that the distributions of meaningful responses varied substantially by baseline pain intensity for both the observed NRS at meaningful pain relief and the absolute change in pain from baseline, but not for percent change. This finding corroborates prior studies which suggest that patients interpret the meaningfulness of changes in acute pain in relative rather than absolute terms [19,23]. If patients had been pressing the stopwatch at a consistent patient-acceptable symptom state (PASS) [16,18],

we would have expected to observe similar distributions of observed NRS pain scores at meaningful relief regardless of baseline pain. Therefore, our findings reinforce percent change as a more stable metric for defining meaningful improvement in acute pain trials, as it inherently adjusts for baseline pain level.

In recognition of potential heterogeneity across surgical procedures, we did not aggregate data to derive an overall MCID for acute pain as has been done in prior studies [2,18,19,23]. However, we found that the optimal thresholds for defining meaningful change in pain intensity fell within a relatively narrow range across all four major surgical pain models. This result was unexpected, given the diverse surgical phenotypes represented. One explanation may be the homogeneity in baseline pain intensities driven by the trial enrollment criteria requiring patients to have moderate or severe pain at baseline.

Another key contribution of this study is the use of the double stopwatch technique to define meaningful pain relief. Prior studies using anchor-based approaches to estimate MCIDs in postoperative pain have relied on investigator-selected thresholds from external scales, such as global rating-of-change questionnaires or pain relief scales [2,5,12,18,19]. Our MCID estimates were consistently higher, demonstrating that an active patient action, like pressing a stopwatch, requires greater reductions in pain intensity than thresholds previously selected by investigators on passive, scale-based ratings. Because the double stopwatch technique directly captures the pain intensity at which each patient experiences meaningful relief, previously reported lower MCIDs are likely due to misclassification arising from externally defined thresholds that do not account for patient-level variation.

Our study is not the first to use a patient-initiated action to classify meaningful pain relief in determining an acute pain MCID. Farrar and colleagues previously calculated the pain reduction necessary for patients to forgo an additional dose of rescue medication among cancer patients experiencing acute breakthrough pain above their daily chronic pain [7,8]. Their analysis identified a 33% improvement in pain intensity as the optimal threshold. While the action of taking rescue medication is a more objective anchor than passive rating scales, the threshold for foregoing rescue medication in the setting of ongoing chronic pain may be lower than in otherwise healthy patients undergoing surgery, given their longer-term experience with pain.

For the development of novel analgesics for acute pain, the FDA recommends that trial sponsors collect complementary measures to fully characterize efficacy [26]. Pain intensity scores, which assess the magnitude of efficacy, are typically the primary outcome assessment, and the sum of pain intensity differences (SPID) over a prespecified duration is typically the summary measure used for primary efficacy analyses. Rescue medication use is another important measure of the overall pain burden that assesses the sufficiency of analgesia and must also be accounted for in the analysis of pain intensity scores.

Although continuous endpoints like SPID provide greater statistical power than binary responder endpoints, they do not provide insight into the clinical relevance of individual patient outcomes. Responder analyses, which classify patients based on clinically meaningful response thresholds (e.g., MCIDs), offer a more patient-centric perspective on

treatment efficacy. Our findings suggest that a 50% improvement in pain intensity from baseline would serve as a reliable benchmark to form responder-based efficacy endpoints in clinical trials. To avoid misinterpretation, we emphasize that our MCIDs apply only to individual patients and cannot be used to determine the clinical meaningfulness of average pain score differences between treatment groups. This is because group means do not capture the distribution of individual patient responses, so two treatments with the same average pain reduction may have very different proportions of patients achieving meaningful relief.

Recent analgesic trials have incorporated responder analyses to assess analgesic efficacy. For example, the two Phase 3 trials of IV oliceridine in bunionectomy and abdominoplasty used a primary endpoint defined as a composite of: (1) a ≥30% improvement in time-weighted SPID from baseline, (2) no use of rescue medication, (3) no early discontinuation of study drug, and (4) not reaching study medication dosing limits [22,27]. If responder-based efficacy endpoints are used in future acute pain trials, as is now common in chronic pain trials, an appropriate threshold must be pre-specified at the trial design phase. Our findings suggest that a 50% improvement in time-weighted SPID may be a more clinically justifiable threshold for a responder analysis in future analgesic trials using one of the acute surgical pain models.

Although the electronic patient-level database of FDA registration trials was large and high-quality due to its regulated nature, several limitations should be acknowledged. First, trials conducted before 1999 were not included because they were not available in electronic format. However, the large number of studies focusing on more recent clinical trials may improve the relevance of MCIDs for future studies. Second, all these trials restricted enrollment to patients with moderate-to-severe pain at baseline, so our MCIDs cannot be assumed to apply to mild pain populations. Third, participants enrolled in registration trials tend to be healthier than the general surgical population, so the results of this study may not be generalizable to surgical patients with more complex comorbidity profiles or preexisting pain conditions. Fourth, the MCIDs for bunionectomy and dental extraction were each derived from more than 10 clinical trials, providing a broader evidence base than for orthopedic (n=2 trials) and soft tissue surgeries (n=4 trials). While the distributions of meaningful change were largely consistent across studies, some between-study variability suggests that additional trials would make estimates for orthopedic and soft tissue surgery more robust. For example, one of the two orthopedic surgery trials studied knee arthroscopy, which had lower baseline pain levels than joint replacement surgeries. This difference suggests that knee arthroscopy may be less well-suited as a surgical pain model, and additional data from joint replacement surgeries using the double stopwatch technique would strengthen the reliability of MCID estimates for orthopedic surgery.

While the double stopwatch technique for capturing the time of meaningful pain relief offers several advantages over conventional anchor-based approaches with external scales, there are also methodological limitations. First, pain intensity scores were not always recorded at the exact moment the patients pressed the stopwatch, requiring interpolation. In most cases, the time difference between observed and interpolated values was small (median, 2 minutes), though interpolation introduces some degree of measurement error. This limitation

could be avoided in future studies by taking an unscheduled pain intensity measurement with each stopwatch press, as is routinely done with each administration of rescue pain medication. Second, some patients may not press the stopwatch at the exact time they experience meaningful relief, either due to forgetfulness or functional impairments. Patients who are older, taking higher doses of sedating medications, or experiencing cognitive impairment may have difficulty responding in real time, potentially leading to delayed or missed stopwatch presses. Third, due to the subjective nature of pain, psychological factors, such as pain catastrophizing, prior pain experiences, and treatment expectations in a placebo-controlled trial may influence stopwatch reporting and introduce variability that cannot be fully measured or accounted for. Finally, the double stopwatch method is not applicable to chronic pain trials because an action-based threshold like pressing a stopwatch does not align with efficacy endpoints for chronic pain, such as reductions in average pain over the past week.

5. Conclusions

This study of 29 randomized controlled analgesic trials submitted to the US FDA found that the MCIDs for moderate-to-severe acute pain intensity across diverse surgical settings were 3 points on a 0-to-10 NRS scale or 50% on the percent scale. Due to its consistency across moderate and severe pain, a 50% improvement in pain intensity emerged as a strong candidate for a standardized responder definition across acute surgical pain models and subgroups. These benchmarks for meaningful pain relief provide a robust framework for interpreting analgesic efficacy and designing patient-centered endpoints. Incorporating MCID-based responder endpoints in future trials would enhance the clinical relevance of statistical assessments and facilitate more meaningful comparisons across analgesic trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was funded by a contract with the US Food and Drug Administration (FDA) awarded to the University of Pennsylvania (FDA-BAA-75F40119C10099). The funder was not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or the decision to submit the manuscript for publication. The FDA provided feedback on the manuscript to ensure no confidential or trade secret information was disclosed. All interpretations presented in this paper reflect the views of the authors and do not reflect the views of the FDA. The data analyzed in this study were submitted to the FDA by industry sponsors as part of regulatory submissions and are not publicly available due to regulatory and confidentiality restrictions.

Conflict of interest statement

C.J.M. was previously a full-time employee and has served as an independent consultant for 3D Communications, LLC, a communications firm with many clients in the pharmaceutical, biotechnology, and medical device industries, including those developing analgesic and anesthetic products. C.E.A. has received consulting fees from AbbVie, XGene Pharmaceutical, Lundbeck, Averitas, and Vertex, as well as research support from AbbVie, Eli Lilly, and Vertex Pharmaceuticals. J.A.H. reports funding from NIH-NIAMS UH3 grant (Co-I). J.T.F. reports over the past three years funding from NIH-NCATS – UL1 Grant (Co-I), NIH-NIDDK – U01 Grants (CoI), NIH-NINDS – U24 Grant (PI), and two FDA-BAA Contracts; and compensation for serving on two NIH DSMBs. J.T.F. has also served on advisory boards as a consultant on clinical trial methods for Vertex, Lilly, and EicOsis. The remaining authors have no conflicts of interest to declare.

References

[1]. Bartley EJ, Fillingim RB. Sex differences in pain: a brief review of clinical and experimental findings. Br J Anaesth 2013;111(1):52–8. [PubMed: 23794645]

[2]. Cepeda MS, Africano JM, Polo R, Alcala R, Carr DB. What decline in pain intensity is meaningful to patients with acute pain? Pain 2003;105(1–2):151–7. [PubMed: 14499431]

[3]. Cooper SA, Desjardins PJ, Turk DC, Dworkin RH, Katz NP, Kehlet H, Ballantyne JC, Burke LB, Carragee E, Cowan P, Croll S, Dionne RA, Farrar JT, Gilron I, Gordon DB, Iyengar S, Jay GW, Kalso EA, Kerns RD, McDermott MP, Raja SN, Rappaport BA, Rauschkolb C, Royal MA, Segerdahl M, Stauffer JW, Todd KH, Vanhove GF, Wallace MS, West C, White RE, Wu C. Research design considerations for single-dose analgesic clinical trials in acute pain: IMMPACT recommendations. Pain 2016;157(2):288–301. [PubMed: 26683233]

[4]. Copay AG, Subach BR, Glassman SD, Polly DW Jr., Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007;7(5):541–6. [PubMed: 17448732]

[5]. Danoff JR, Goel R, Sutton R, Maltenfort MG, Austin MS. How much pain is significant? Defining the minimal clinically important difference for the visual analog scale for pain after total joint arthroplasty. J Arthroplasty 2018;33(7s):S71–S75.e72. [PubMed: 29567002]

[6]. Desjardins PJ, Black PM, Balm TK, Malmstrom HS, Judy K, Davis DD. Onset of analgesia: further validation of a new stopwatch methodology. Clin Pharm Therap 1996;59(2):130.

[7]. Farrar JT, Berlin JA, Strom BL. Clinically important changes in acute pain outcome measures: a validation study. J Pain Symptom Manage 2003;25(5):406–11. [PubMed: 12727037]

[8]. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. Pain 2000;88(3):287–94. [PubMed: 11068116]

[9]. Gilron I, Carr DB, Desjardins PJ, Kehlet H. Current methods and challenges for acute pain clinical trials. Pain Rep 2019;4(3):e647. [PubMed: 31583333]

[10]. Gilron I, Lao N, Carley M, Camire D, Kehlet H, Brennan TJ, Erb J. Movement-evoked pain versus pain at rest in postsurgical clinical trials and in meta-analyses: an updated systematic review. Anesthesiology 2024;140(3):442–9. [PubMed: 38011045]

[11]. Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. J Orthop Surg Res 2015;10:24. [PubMed: 25645576]

[12]. Martin WJ, Ashton-James CE, Skorpil NE, Heymans MW, Forouzanfar T. What constitutes a clinically important pain reduction in patients after third molar surgery? Pain Res Manag 2013;18(6):319–22. [PubMed: 23957018]

[13]. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. JAMA 2014;312(13):1342–3. [PubMed: 25268441]

[14]. Miller CJ, Farrar JT. Methodology for determining clinically important differences in acute pain intensity with the double stopwatch technique. J Pain. 2024 Dec 15:104759.

[15]. Moore A, McQuay H, Gavaghan D. Deriving dichotomous outcome measures from continuous data in randomised controlled trials of analgesics: verification from independent data. Pain 1997;69(1–2):127–30. [PubMed: 9060022]

[16]. Muñoz-Leyva F, El-Boghdadly K, Chan V. Is the minimal clinically important difference (MCID) in acute pain a good measure of analgesic efficacy in regional anesthesia? Reg Anesth Pain Med 2020;45(12):1000–5. [PubMed: 32900985]

[17]. Myles PS, Christelis N. Measuring pain and analgesic response. Eur J Anaesthesiol 2011;28(6):399–400. [PubMed: 21540652]

[18]. Myles PS, Myles DB, Galagher W, Boyd D, Chew C, MacDonald N, Dennis A. Measuring acute postoperative pain using the visual analog scale: the minimal clinically important difference and patient acceptable symptom state. Br J Anaesth 2017;118(3):424–9. [PubMed: 28186223]

[19]. Olsen MF, Bjerre E, Hansen MD, Hilden J, Landler NE, Tendal B, Hróbjartsson A. Pain relief that matters to patients: systematic review of empirical studies assessing the minimum clinically important difference in acute pain. BMC Med 2017;15(1):35. [PubMed: 28215182]

[20]. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61(2):102–9. [PubMed: 18177782]

[21]. Singla NK, Desjardins PJ, Chang PD. A comparison of the clinical and experimental characteristics of four acute surgical pain models: dental extraction, bunionectomy, joint replacement, and soft tissue surgery. Pain 2014;155(3):441–56. [PubMed: 24012952]

[22]. Singla NK, Skobieranda F, Soergel DG, Salamea M, Burt DA, Demitrack MA, Viscusi ER. APOLLO-2: A Randomized, Placebo and Active-Controlled Phase III Study Investigating Oliceridine (TRV130), a G Protein-Biased Ligand at the mu-Opioid Receptor, for Management of Moderate to Severe Acute Pain Following Abdominoplasty. Pain Pract 2019;19(7):715–31. [PubMed: 31162798]

[23]. Sloman R, Wruble AW, Rosen G, Rom M. Determination of clinically meaningful levels of pain reduction in patients experiencing acute postoperative pain. Pain Manag Nurs 2006;7(4):153–8. [PubMed: 17145489]

[24]. Srikandarajah S, Gilron I. Systematic review of movement-evoked pain versus pain at rest in postsurgical clinical trials and meta-analyses: a fundamental distinction requiring standardized measurement. Pain 2011;152(8):1734–9. [PubMed: 21402445]

[25]. Sunshine A, Laska E, Meisner M. Oral aspirin in post-operative pain: a quantitative, systematic review. Edwards et al.,PAIN 81 (1999) 289–297. Pain 2000;88(3):309–10. [PubMed: 11221654]

[26]. U.S. Food and Drug Administration. Development of non-opioid analgesic for acute pain: guidance for industry. Center for Drug Evaluation of Research (CDER); 2022. Available from: https://www.fda.gov/media/156063/download

[27]. Viscusi ER, Skobieranda F, Soergel DG, Cook E, Burt DA, Singla N. APOLLO-1: a randomized placebo and active-controlled phase III study investigating oliceridine (TRV130), a G protein-biased ligand at the micro-opioid receptor, for management of moderate-to-severe acute pain following bunionectomy. J Pain Res 2019;12:927–43. [PubMed: 30881102]

[28]. Wallace MS. Trials for managing acute pain - a clinically meaningful small effect size? N Engl J Med 2023;389(5):464–5. [PubMed: 37530828]

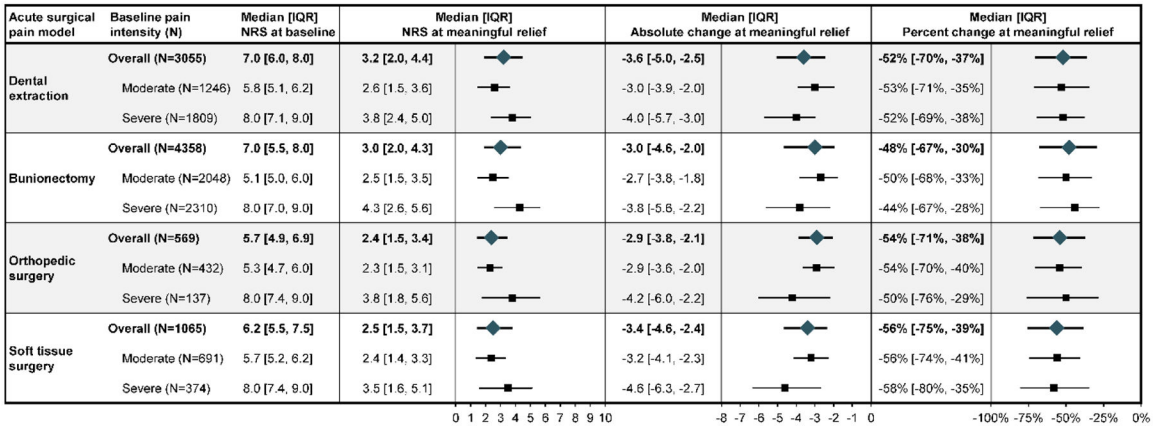

Figure 1.

Summary of observed NRS at meaningful relief, absolute change from baseline at meaningful relief, and percent change from baseline at meaningful relief by acute surgical pain model, overall and stratified by moderate (NRS 4 to <7) and severe (NRS ≥7) baseline pain intensity. Pain intensity scores measured on the 0–100 VAS were converted to the 0–10 NRS by dividing by 10. Estimates were derived using the Kaplan-Meier method for the double-stopwatch technique.

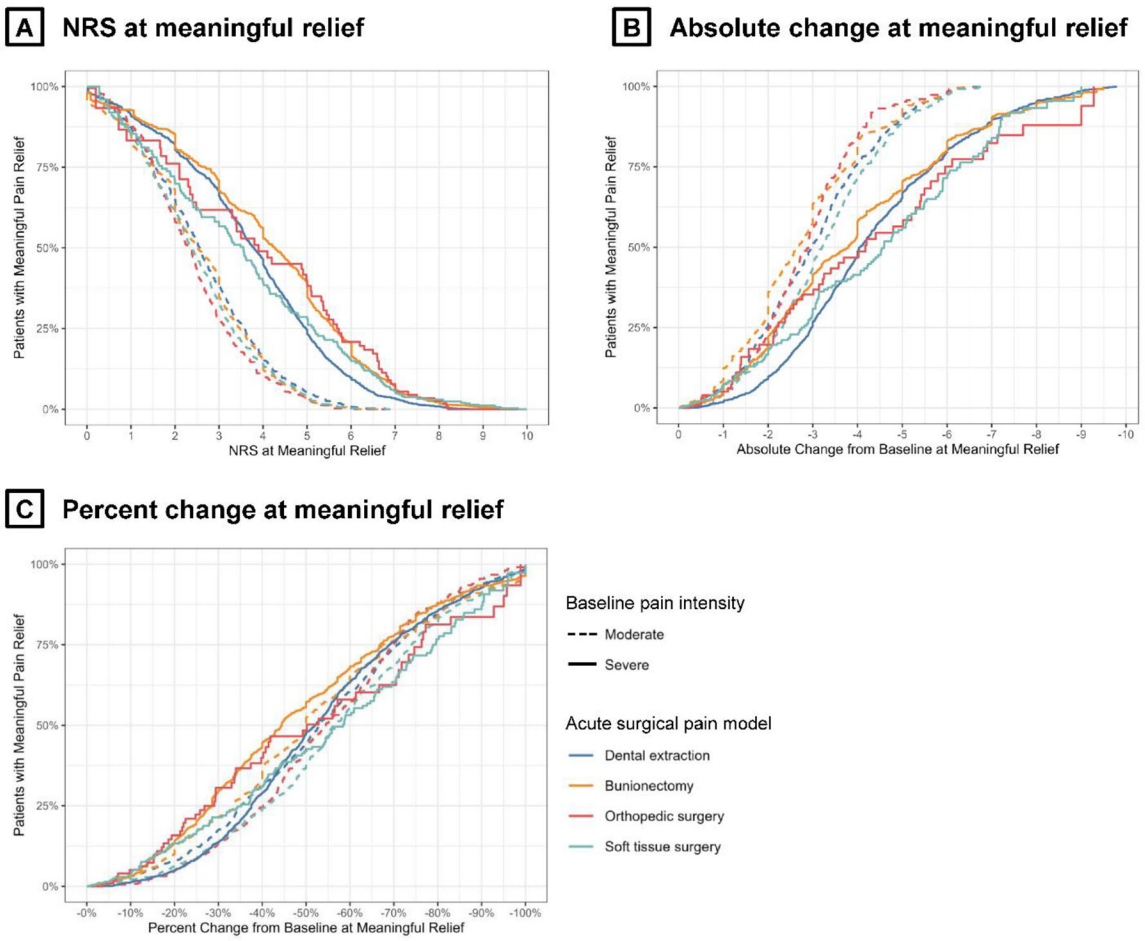

Figure 2.

Distributions of observed NRS at meaningful relief (Panel A), absolute change from baseline at meaningful relief (Panel B), and percent change from baseline at meaningful relief (Panel C) by acute surgical pain model, stratified by moderate (NRS 4 to <7) and severe (NRS ≥7) baseline pain intensity. Pain intensity scores measured on the 0–100 VAS were converted to the 0–10 NRS by dividing by 10. Estimates were derived using the Kaplan-Meier method for the double-stopwatch technique.

Table 1.

Baseline patient characteristics by acute surgical pain model

	Dental extraction (N=3055)	Bunionectomy (N=4358)	Orthopedic surgery (N=569)	Soft tissue surgery (N=1065)
Number of trials	11	12	2	4
Number of patients, active : placebo	2549 : 506	3325 : 1033	422 : 147	807 : 258
Oral study drug	2013 : 388	2951 : 881	225 : 75	0 : 0
IV study drug	536 : 118	374 : 152	197 : 72	807 : 258
Female, n (%)	1827 (59.8)	3716 (85.3)	289 (50.8)	958 (90.0)
Age (years), mean (SD)	22 (4.7)	43 (13.1)	50 (14.8)	44 (9.4)
Median [IQR]	20 [18, 23]	44 [32, 54]	51 [40, 60]	45 [38, 51]
≥ 55, n (%)	1 (<0.1)	1002 (23.0)	236 (41.5)	138 (13.0)
Baseline pain intensity (NRS *), mean (SD)	7.1 (1.47)	6.9 (1.67)	6.1 (1.47)	6.6 (1.42)
Median [IQR]	7.0 [6.0, 8.0]	7.0 [5.5, 8.0]	5.7 [4.9, 6.9]	6.2 [5.5, 7.5]
Severe (≥ 7), n (%)	1809 (59.2)	2310 (53.0)	137 (24.1)	374 (35.1)
Region, n (%)				
United States	2325 (76.1)	4358 (100.0)	269 (47.3)	593 (55.7)
Europe	457 (15.0)	0 (0.0)	0 (0.0)	472 (44.3)
New Zealand	273 (8.9)	0 (0.0)	300 (52.7)	0 (0.0)

IQR, interquartile range; NRS, numeric rating scale.

* Pain intensity scores measured using the 0–100 VAS were converted to the 0–10 NRS scale by dividing by 10.

Table 2.

Operating characteristics of various thresholds of absolute and percent change from baseline in pain intensity for meaningful relief by acute surgical pain model

	Threshold				
Acute surgical pain model	**≥1 NRS / ≥10 VAS**	**≥2 NRS / ≥20 VAS**	**≥3 NRS / ≥30 VAS**	**≥4 NRS / ≥40 VAS**	**≥5 NRS / ≥50 VAS**
Dental extraction					
AUC	0.643	0.751	**0.807**	0.799	0.741
Sensitivity (%)	99.5	96.7	**89.0**	74.7	55.6
Specificity (%)	29.1	53.4	**72.3**	85.0	92.5
Bunionectomy					
AUC	0.692	0.794	**0.812**	0.767	0.694
Sensitivity (%)	99.0	93.1	**80.5**	62.4	42.9
Specificity (%)	39.3	65.7	**81.9**	91.1	95.8
Orthopedic surgery					
AUC	0.681	0.798	**0.819**	0.740	0.631
Sensitivity (%)	99.5	94.3	**79.8**	55.1	29.6
Specificity (%)	36.7	65.3	**84.0**	92.9	96.5
Soft tissue surgery					
AUC	0.656	0.764	**0.807**	0.767	0.673
Sensitivity (%)	99.2	94.7	**83.8**	64.2	38.9
Specificity (%)	32.0	58.2	**77.5**	89.3	95.6
	≥20%	**≥30%**	**≥40%**	**≥50%**	**≥60%**
Dental extraction					
AUC	0.696	0.768	0.814	**0.827**	0.807
Sensitivity (%)	98.9	96.6	91.7	**83.3**	71.6
Specificity (%)	40.2	57.1	71.2	**82.0**	89.7
Bunionectomy					
AUC	0.733	0.793	**0.819**	0.812	0.781
Sensitivity (%)	97.8	92.3	**86.0**	76.1	64.0
Specificity (%)	48.8	65.5	**77.8**	86.4	92.2
Orthopedic surgery					
AUC	0.700	0.770	0.814	**0.828**	0.810
Sensitivity (%)	98.9	96.2	91.0	**83.0**	71.6
Specificity (%)	41.0	57.7	71.8	**82.7**	90.4
Soft tissue surgery					
AUC	0.686	0.751	0.793	**0.811**	0.799
Sensitivity (%)	98.4	95.4	90.3	**82.7**	71.8
Specificity (%)	38.8	54.8	68.4	**79.5**	87.9

AUC, area under the curve; NRS, numeric rating scale; VAS, visual analog scale.

Note: bolded text highlights the threshold value that optimized AUC for each acute surgical pain model.