HHS Public Access

Author manuscript

J Pain. Author manuscript; available in PMC 2026 May 01.

Published in final edited form as:

J Pain. 2025 May ; 30: 104759. doi:10.1016/j.jpain.2024.104759.

Methodology for Determining Minimally Clinically Important Differences in Acute Pain Intensity with the Double Stopwatch Technique

Christopher J. Miller[1], **John T. Farrar**[1]

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

Abstract

Minimum clinically important differences (MCIDs) in acute pain intensity have not been well established. Conventional approaches for estimating MCIDs require an independent reference scale, with a threshold that must be presumed to accurately classify meaningful change in pain for all study participants, to serve as an anchor. The double stopwatch technique is the gold standard for measuring the time to meaningful relief, where participants actively press the second stopwatch when they experience pain relief that is meaningful to them. This technique eliminates the problem of misclassification with arbitrary anchors at a single time point, but the censored nature of the data is not amenable for determining MCIDs using standard methods. We propose a stopwatch-based MCID methodology that employs the double stopwatch technique to identify individualized thresholds for meaningful change in pain. This approach enables direct classification of changes in pain for each participant based on whether they perceived the change as meaningful and whether it exceeded the study cut-off being tested. Pain values of participants who do not achieve meaningful relief are incorporated into the analysis to address censoring and avoid bias. The performance (e.g., sensitivity, specificity) of different thresholds to serve as an MCID can be estimated using standard approaches with variance estimates derived by cluster bootstrapping. The advantages of the stopwatch-based MCID methodology are illustrated relative to a conventional approach using data from a randomized trial in third molar extraction.

Perspective: This article describes a methodology for determining MCIDs using the double stopwatch technique, the gold standard for assessing meaningful changes in acute pain. This methodology can be used to establish MCIDs in different acute pain settings, providing a useful basis to evaluate the meaningfulness of clinical trial results.

Keywords

acute pain; minimum clinically important difference; clinical trials; statistical analysis; pain measurement

Corresponding author: Christopher J. Miller, Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA. chris.miller1@pennmedicine.upenn.edu.

Authorship Contributions

CJM: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review and editing. **JTF:** Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Study pre-registration statement: This research project was not preregistered.

Introduction

Randomized trials of pain therapeutics are generally designed to compare the efficacy of different treatments on an average measure of pain intensity. Interpreting trial results can be challenging since treatment effects that are statistically significant are not necessarily clinically meaningful. Furthermore, average differences in pain intensity between groups may not accurately characterize efficacy when outcomes are not normally distributed, as is often the case with analgesics, where patients tend to experience either substantial improvement or little to none.[1]

Patient-centric efficacy measures, such as those based on a responder status, enhance the validity and interpretability of clinical trial results.[2] Responder-based endpoints for subjective outcomes like pain intensity are usually defined by achieving the minimum clinically important difference (MCID), the smallest level of effect that an individual would identify as important. MCIDs are usually determined with anchor-based methods where the optimal cut-point for the outcome measure is identified by maximizing its classification performance on a reference measure, ideally a gold standard, using receiver operating characteristic (ROC) curve analysis.

While MCIDs in chronic pain have been well established for more than two decades,[3] definitive thresholds for acute pain have remained elusive[2,4] because determining an MCID on a relevant time scale is more challenging for acute pain. The meaningfulness of improvement in chronic pain can be assessed by participants rating their global impression of change at a single time point, considering their pain experience over the past several days. In acute settings, however, pain intensity and the meaningfulness of its change can evolve rapidly over the course of hours or even minutes. Despite the dynamic nature of acute pain and the high inter-patient variability in the onset and magnitude of response to analgesics, prior studies have evaluated as few as one or two time points to estimate MCIDs.[5,6]

The double stopwatch technique is the gold standard for determining the time to meaningful acute pain relief.[7] Participants indicate when they experience perceptible and meaningful pain relief with two stopwatches. The US Food and Drug Administration (FDA) has used results from the second stopwatch for meaningful pain relief to determine the onset of analgesia for drug labeling.[8] Despite the potential advantages of defining meaningfulness at the individual level with a deliberate action, the double stopwatch technique has never been used to determine MCIDs in acute pain intensity.

The failure to leverage this gold standard to identify MCIDs in acute pain may be due to the limitations of existing methodologies. Conventional approaches to identifying cut-points require binary classification of each individual’s status on the reference measure and the corresponding test measurement recorded at the same time. MCIDs are identified based on their performance classification at a given time point using sensitivity and specificity metrics. With the double stopwatch approach, however, meaningful relief can occur at any time point after baseline. Thus, as a time-to-event endpoint with right censoring, double stopwatch data are not directly amenable to binary classification.

Here, we propose a novel approach for estimating MCIDs in acute pain intensity using a stopwatch-based technique. First, the double stopwatch data are used to identify participant-specific thresholds of meaningful pain relief. Then, these personalized thresholds are used to classify each value of change in pain informed by the stopwatch data for each participant as meaningful or not. These values provide the data necessary to use ROC curve analysis to evaluate the full range of potential cut-offs in the study population to identify an optimal cut-point as the MCID. After introducing technical details, we illustrate the conventional and stopwatch-based approaches to determining MCIDs using patient-level data from a randomized trial in third molar extraction.[9]

Data Source for Example Trial

The example trial used in this paper to illustrate the double stopwatch technique and the stopwatch-based method for determining MCIDs was previously submitted to the FDA and the results have been previously published.[9] The example trial was a randomized, double-blind, placebo-controlled trial of 5 doses of intravenous diclofenac and an active comparator of ketorolac 30 mg for third-molar extraction. The median age of the 353 patients in the study was 22 years (range 18 to 42), 62% were female, and 52% reported severe pain (VAS ≥ 70) at baseline. The University of Pennsylvania Institutional Review Board approved the re-analyses of historical trial data submitted to the FDA.

Conventional Approach to Identifying MCIDs for Pain Measures

The conventional approach for identifying an MCID in acute pain entails assessing the change in pain intensity relative to a reference "anchor" measure collected concurrently. Pain intensity is often measured on the NRS (0-10) or the visual analog scale (VAS, 0-100), where 0 represents *'no pain'* and the maximum value represents *'worst possible pain'*. A common choice as an acute pain anchor measure has been the Pain Relief Scale, where the current level of pain relief is rated as *'none', 'a little', 'some', 'a lot',* or *'complete'.* Participants who respond *'a lot'* or *'complete'* are categorized as meaningfully improved based on the face validity, while others are categorized as not meaningfully improved. Participants who drop out of the study or take rescue pain medication before the MCID assessment are typically excluded from the analysis.

Each individual can then be classified as belonging to one cell of a 2×2 table as a true positive, true negative, false positive, or false negative based upon their response according to the anchor and their change in pain intensity relative to the cut-off being evaluated. The MCID is identified using a ROC curve analysis,[10,11] which evaluates a full range of different potential cut-offs to identify the optimal threshold for change in pain intensity that balances sensitivity and specificity according to a prespecified criterion. One popular approach is Youden's index,[12] which maximizes the sum of sensitivity and specificity assuming that each is of equal importance. Alternate weighting schemes could be used depending on the relative importance of false positives and false negatives.[13]

Stopwatch-Based Approach to Identifying MCIDs

Double Stopwatch Technique

With the double stopwatch technique, both stopwatches are started when the study drug is first administered. Participants are instructed to stop their first stopwatch when they experience any perceptible pain relief and their second stopwatch when they experience pain relief that is meaningful to them.[7,8] If a participant decides to withdraw from the study or take a rescue pain medication, they are recorded as having not achieved the corresponding degree of (perceptible or meaningful) pain relief at that time. Due to the right censoring that may occur due to rescue medication or dropout, the time to perceptible pain relief and time to meaningful pain relief are evaluated using the Kaplan-Meier method.

Figure 1A illustrates the Kaplan-Meier analysis of time to meaningful pain relief in the example dental extraction trial by treatment arm. As previously reported in the primary trial publication,[9] all active treatments showed a faster time to meaningful pain relief than placebo. The lowest IV diclofenac dose showed a slower median time to meaningful pain relief than the other active treatments.

Individualized Thresholds for Meaningful Change in Pain Intensity

The double stopwatch technique was designed to determine the times to perceptible pain relief and meaningful pain relief. Our proposed methodology for determining MCIDs takes advantage of the fact that, for any participant i, their change in pain intensity from the time study drug is first administered (i.e., baseline) to the time of meaningful relief (M_i) can also be captured. This change may be measured directly if the participant records their pain intensity at the time they stop the second stopwatch or calculated by interpolation between scheduled pain intensity assessments. Some study participants may not experience a sufficient reduction in pain intensity with their assigned study drug to achieve meaningful relief prior to taking rescue medication, dropping out, or completing the study. Therefore, the distribution of M will be censored in most study populations.

While not necessary for determining an MCID, summarizing the shape and spread of this distribution can be helpful for understanding participant-level variability. This can also be accomplished using the Kaplan-Meier method, where the onset of meaningful relief is the event, and the change in pain intensity at the onset of meaningful relief is used instead of time. Those who do not achieve meaningful relief are censored at their most favorable change in pain intensity during the observation period between baseline and the time of first rescue medication, drop out, or study completion.

Figure 1B illustrates the Kaplan-Meier analysis of change in pain intensity to meaningful relief for the example trial by treatment arm and overall. Unlike the analysis of time to meaningful relief (Figure 1A), there are no differences between the treatment arms in the change in pain intensity required to achieve meaningful relief. This observation aligns with the understanding that the magnitude of improvement perceived as meaningful within a given context (e.g., a dental extraction procedure) is an intrinsic individual-level characteristic independent of the assigned treatment. Differences between treatment arms

may be present in the proportions of participants achieving meaningful relief and in the timing of that relief depending on the characteristics of each study drug. However, because treatment assignment is randomized, the distributions of changes perceived as meaningful would be expected to be similar across arms. Therefore, the stopwatch-based approach can derive an MCID by analyzing the overall study population, without regard to treatment assignment.

Binary Classification of Changes in Pain Intensity

The availability of participant-specific thresholds for meaningful change in pain intensity (M_i) allows for binary classification at the level of individual change scores. To describe how to adapt and utilize these thresholds for analysis, first, let C represent the cut-off being considered as a potential MCID for change in pain intensity. Ultimately, in the ROC curve analysis, the full range of potential values for C will be considered. For each participant i, let Δ_{ij} denote the change from baseline in pain intensity score for the j^{th} pain intensity value, where j indexes the range of unit-level changes in the pain intensity scale informed by the double stopwatch data for each participant. (Note that increasingly negative values of Δ indicate greater reductions in pain intensity.) For a participant i who achieved meaningful pain relief, relevant changes in pain (Δ_{ij}) would range from the inverse of the baseline value up to 0. That is, for someone with a baseline value of 70 on the VAS upon first perceptible pain relief who subsequently achieved meaningful relief, the range of Δ_{ij} would be from −70 to 0. For a participant i who did not achieve meaningful pain relief prior to taking rescue medication, dropping out, or completing the study, Δ_{ij} would range from the most favorable change observed prior to censoring up to 0.

Recall that M_i is directly observed only for participants who stop the second stopwatch. The exact value of M_i cannot be directly identified for censored individuals, but it is known that $M_i < min(\Delta_{ij})$, across all j.

Using the previous notation, Table 1 describes the calculations for computing each cell of a 2×2 table for binary classification at a given cut-off point with the stopwatch-based MCID method. Note that $I()$ is an indicator function that takes on the value 1 if the conditions are true and 0 if they are false. Therefore, each cell in the table represents the total number of changes in pain intensity that meet the classification criterion based on each participant's individualized threshold for meaningful relief, summed across all participants.

For participants who achieve meaningful relief based on the double stopwatch, values of change in pain intensity that meet both the individualized threshold for meaningful pain relief and the cut-off (C) being tested are classified as true positives. Changes in pain intensity that do not meet either the individualized threshold or the cut-off are classified as true negatives. Changes in pain that meet the cut-off but not the individualized threshold are classified as false positives, while changes in pain that meet the individualized threshold for meaningful relief but not the cut-off are classified as false negatives.

For participants who do not achieve meaningful pain relief, the changes in pain intensity informed by the double stopwatch data can only be classified as true negatives or false

positives, depending on whether the change value is above or below the cut-off being considered.

Three case examples of how double stopwatch data are used to perform binary classification with this methodology are provided in Figures 2A, 2B, and 2C. Figures 2A and 2B show how data from one participant would be classified with different values of the cut-off, and Figure 2C shows how data from a censored participant would be classified.

All illustrations in this paper depict absolute changes in pain, though classification using percentage changes would follow a similar approach. For a participant *i* who achieved meaningful pain relief, Δ_{ij} would index all percentage point changes from −100% up to 0%. For a participant *i* who did not achieve meaningful pain relief prior to taking rescue medication, dropping out, or completing the study, Δ_{ij} would range from the most favorable percentage point change observed prior to censoring to up 0%.

Determining MCIDs and Assessing Classification Performance

To determine an MCID, 2×2 tables are created for all possible cut-offs from no change (i.e., 0) up to the maximum possible value of change for the measure (i.e., −100 for the VAS or -10 for the NRS). The sensitivity and specificity for each cut-off are calculated and plotted as an ROC curve. The MCID can then be determined along with the area under the ROC curve using the ROC curve analysis.

Unlike a standard ROC analysis, where each participant is represented once in the dataset and each observation is assumed to be independent, the stopwatch-based method includes multiple observations (i.e., change scores) per participant. Therefore, standard confidence intervals (CIs) for performance characteristics such as sensitivity and specificity will be incorrect due to the dependence of observations within the same individuals. To address this issue, we employ a cluster bootstrapping procedure to estimate CIs that account for intra-individual correlation. In a standard bootstrap, individual observations from the dataset are randomly sampled with replacement to create a bootstrapped sample of the same size as the original sample. The statistic of interest is calculated for the bootstrap sample, and this process is repeated over many iterations to derive the sampling distribution for the statistic across many bootstrap samples.[14] Cluster bootstrapping extends this procedure to correlated data by randomly sampling at the cluster (i.e., participant) level, including all their observations in each bootstrap sample to preserve the within-individual dependence structure. A 95% percentile bootstrap CI is generated by identifying the interval between the 2.5^{th} and 97.5^{th} percentiles of the distribution of the bootstrapped statistic.[15] An example illustration of the methodology with simulated data using R is provided in the Supplementary file.

Data Application

We compared the stopwatch-based method to the conventional method for determining MCIDs. For the conventional method, a response of '*a lot*' or '*complete*' on the Pain Relief Scale at 60 minutes was used as the anchor for categorizing meaningful improvement. Performance was evaluated at 10-point intervals for change in VAS (from −10 to −50) for

both methods, and the optimal cut-point for each approach was identified using Youden's index. Confidence intervals for performance characteristics were generated using the 1000 bootstrap samples for the conventional method and 1000 clustered bootstrap samples for the stopwatch-based method.

The conventional method excluded data from 94 participants (27%) due to the utilization of rescue medication, dropout before the 60-minute assessment, or missing data. At 60 minutes, 139 of the 259 evaluable participants (54%) had meaningful relief according to the Pain Relief Scale. The optimal cut-off using Youden's index with the conventional method was -37 on the VAS (Table 2), with an estimated sensitivity of 0.93 (95% CI: 0.88–0.97) and specificity of 0.82 (95% CI: 0.76–0.89).

The stopwatch-based method evaluated data from all 353 participants with no exclusions. Using this technique, 244 participants (69%) achieved meaningful pain relief prior to taking rescue medication, dropping out, or completing the study. The optimal cut-off using Youden's index for the stopwatch-based method was -34, with an estimated sensitivity of 0.83 (95% CI: 0.81–0.85) and specificity of 0.77 (0.76–0.79).

For each cut-off, the conventional method consistently estimated higher sensitivities, specificities, and accuracies than the stopwatch-based method (Table 2). At face value, this could mistakenly be interpreted as favoring the superiority of the conventional method. However, these higher performance statistics were based on an anchor that is not a gold standard (i.e., pain relief of 'a lot' or 'complete' at 60 minutes). Since the double stopwatch directly measures individual thresholds of meaningful pain relief for each individual (i.e., the gold standard), any deviations in performance characteristics from the stopwatch-based method are actually a manifestation of bias from the conventional approach. The bias of the conventional method may be attributable to three sources. First, the artificial selection of responses that constitute a meaningful improvement on the Pain Relief Scale introduces outcome misclassification since they must be presumed to define meaningfulness accurately for all individuals. Second, restricting assessment to only a single time point leads to measurement error since it ignores the potential for meaningful changes in pain both before and after the assessment. Finally, excluding participants who used rescue medication or dropped out before the evaluation could induce selection bias by restricting the analysis to those more likely to have had a favorable treatment response.

The more extensive utilization of study data also led to greater statistical efficiency with the stopwatch-based method, which had more precise 95% CIs. The conventional method classified change in VAS at 60 minutes as meaningful or not for the 259 subjects with observed data at that time. The stopwatch-based method for determining MCIDs classified all values of change in pain that were informed by the double stopwatch data, which ultimately resulted in 21,519 values among the 353 participants.

Discussion

We have introduced a methodology that allows us to use the double stopwatch technique for determining MCIDs in acute pain clinical trials. By directly identifying individualized

thresholds of meaningful pain relief with the participant-driven, time-independent, action-oriented gold-standard technique, the stopwatch-based method reduces outcome misclassification relative to conventional anchor-based approaches for determining MCIDs. It also increases the precision of estimates since all possible values of change in pain for an individual can be accurately classified and analyzed instead of just a single value at one time point. The methodology also avoids bias associated with missing data and censoring since all relevant pain score values prior to rescue medication or dropout are considered in all participants whether they achieve meaningful relief or not.

The present paper provides an illustrative example of how to determine an MCID using data from a single dental extraction trial.[9] While the trial had an appreciable sample size and provided precise estimates of sensitivity and specificity, establishing a definitive MCID for a particular surgical setting would likely require the synthesis of multiple trials to ensure generalizability. The results within any single trial could be influenced by study-specific effects of the study population, drugs evaluated, or experimental procedures. Thus, MCIDs based on our method would be more robust if results were based on data from multiple clinical trials.

In this paper, we have focused on the second stopwatch corresponding to meaningful pain relief given its preferential use by the FDA to characterize the onset of analgesia in drug labeling. However, other stopwatch-related metrics could also be considered, including the onset of perceptible pain relief from the first stopwatch or the confirmed time to perceptible relief, which considers first stopwatch times only if the second stopwatch is pressed, indicating that meaningful relief was ultimately achieved. The stopwatch-based method could be applied to those metrics, as well, with M_i as previously described being replaced by the pain intensity score recorded at the respective stopwatch time of interest.

A particular strength of the stopwatch-based methodology for determining MCIDs is that it defines the anchor by a deliberate patient action, rather than a passive response to an external questionnaire. To our knowledge, the only other patient-determined indicators of clinical importance in acute pain are found in studies of breakthrough cancer pain, where the patient's decision to take or forgo an additional dose of rescue medication served as a proxy for adequate pain relief.[16,17] While this anchor is also based on a patient action and provides valuable insight into the perceived magnitude of analgesic efficacy, it is unclear whether the threshold of change in pain intensity for opting out of additional rescue medication would consistently align with the degree of improvement that a patient might consider subjectively meaningful. The choice to take an additional dose of rescue medication may be influenced by factors beyond the perceived adequacy of change in pain response, such as a desire to avoid side effects or additional medication use. The stopwatch-based approach, by contrast, directly captures the moment a patient perceives relief that meets their personal standard for meaningful change, providing a patient-centered anchor for assessing clinical significance without potential external interference.

While our approach was developed for acute pain intensity, it can be readily adapted to generate MCIDs for any outcome measure assessed repeatedly over time where participants are able to report a meaningful change in their condition in real time. Similar to the use of

stopwatches in acute pain trials, the onset of meaningful change in any symptom or disease state could be captured from the patient's perspective using electronic diaries, mobile apps, or wearable devices. When the patient reports a meaningful improvement or worsening, they could be prompted to complete the patient-reported outcome (PRO) measure of interest. Alternatively, if the PRO is already ascertained at sufficiently regular intervals, the value of the PRO at meaningful change could be interpolated. Thus, thresholds derived using this methodology could be used to develop more patient-centric endpoints for clinical trials across a wide range of clinical conditions.

Several limitations of this study should be acknowledged. First, we have only compared the performance characteristics of our method relative to the conventional approach to estimating MCIDs using a single anchor for one study. Further research is needed to understand how the conventional and stopwatch-based approaches to determining MCIDs correspond across a wider range of datasets. Second, the double stopwatch technique has generally only been employed in studies for regulatory approval, so it may not be possible to leverage many previously completed studies to estimate MCIDs with this approach. Third, the administration of the double stopwatch technique and the corresponding analyses have not been standardized. Future application of the double stopwatch technique should be consistently applied across studies to enable comparison of MCIDs across acute pain phenotypes. Finally, while pain assessments are taken frequently after baseline, the stopwatch-based method for determining MCIDs may be subject to some measurement error due to the need to interpolate pain intensity scores between scheduled assessments. This measurement error could be reduced if pain intensity scores were taken at each stopwatch press, similar to how protocols currently mandate pain intensity scores be taken before each dose of rescue medication.

In summary, the method outlined in this study enables the determination of MCIDs in acute pain using the gold-standard approach with the double stopwatch technique. This methodology holds promise for addressing the long-standing lack of established MCIDs in acute pain settings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to acknowledge the valuable time and insights provided by Warren Bilker, Ian DeLorey, Charles Argoff, Russell Bell, Andrew Conroy, Jennifer Gewandter, Ian Gilron, Jennifer Haythorthwaite, Nathaniel Katz, Tara McWilliams, Katherine Theken, and Enrique Schisterman, whose feedback was instrumental in refining our approach.

Disclosures

This work was supported by a contract with the US Food and Drug Administration (FDA) [FDA-BAA-75F40119C10099]. CJM 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. JTF reports funding over the past three years 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. He has also served on advisory boards as a consultant on clinical trial methods for Vertex, Lilly, and EicOsis.

Data Availability

The participant-level data included in this study as an illustrative example were submitted to the FDA by an industry sponsor and may not be publicly released. R statistical code is provided for a simulated dataset to illustrate the proposed method as a supplementary appendix.

References

1. Moore A, Moore O, McQuay H, Gavaghan D. Deriving dichotomous outcome measures from continuous data in randomised controlled trials of analgesics: use of pain intensity and visual analogue scales. Pain. 1997;69(3):311–315. doi:10.1016/S0304-3959(96)03306-4 [PubMed: 9085306]
2. 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. Dec 2020;45(12):1000–1005. doi:10.1136/rapm-2020-101670 [PubMed: 32900985]
3. Farrar JT, Young JP Jr., LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain. 2001;94(2):149–158. doi:10.1016/S0304-3959(01)00349-9 [PubMed: 11690728]
4. Wallace MS. Trials for Managing Acute Pain - A Clinically Meaningful Small Effect Size? N Engl J Med. 2023;389(5):464–465. doi:10.1056/NEJMe2305480 [PubMed: 37530828]
5. 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. Dec 2006;7(4):153–158. doi:10.1016/j.pmn.2006.09.001 [PubMed: 17145489]
6. 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):319–322. doi:10.1155/2013/742468 [PubMed: 23957018]
7. Desjardins PJ, Black PM, Balm TK, Malmstrom HS, Judy K, Davis DD. Onset of analgesia: further validation of a new stopwatch methodology. Clin Pharmacol Ther. 1996;59(2):130. doi:10.1038/sj.clpt.1996.21
8. Food and Drug Administration. Development of non-opioid analgesics for acute pain: Guidance for industry - draft guidance. Published February 7, 2022. Accessed November 13, 2024. (https://www.fda.gov/media/156063/download). [Accessed 13 November 2024].
9. Christensen K, Daniels S, Bandy D, et al. A double-blind placebo-controlled comparison of a novel formulation of intravenous diclofenac and ketorolac for postoperative third molar extraction pain. Anesth Prog. 2011;58(2):73–81. doi:10.2344/0003-3006-58.2.73 [PubMed: 21679043]
10. Swets JA, Dawes RM, Monahan J. Better decisions through science. Sci Am. 2000;283(4):82–87. doi:10.1038/scientificamerican1000-82
11. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993;39(4):561–577. [PubMed: 8472349]
12. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32–35. doi:10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3 [PubMed: 15405679]
13. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol. 2006;163(7):670–675. doi:10.1093/aje/kwj063 [PubMed: 16410346]
14. Efron B. Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics. 1979;7(1):1–26, 26. doi:10.1214/aos/1176344552
15. Davison AC, Hinkley DV. Complex Dependence. In: Bootstrap Methods and Their Application. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press; 1997:385–436.
16. 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. doi:10.1016/s0885-3924(03)00162-3 [PubMed: 12727037]

17. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. Pain. 2000;88(3):287–294. doi:10.1016/S0304-3959(00)00339-0 [PubMed: 11068116]

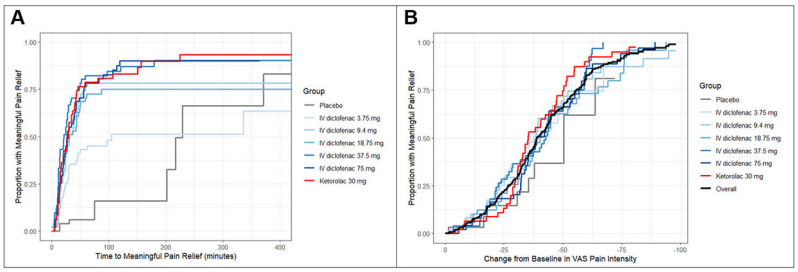

Figure 1.

Kaplan-Meier curves based on double stopwatch data from the example trial. **(A)** Time to meaningful pain relief by treatment arm. **(B)** Change from baseline in VAS pain intensity to meaningful pain relief by treatment arm and overall. Baseline is defined as the time study drug is first administered.

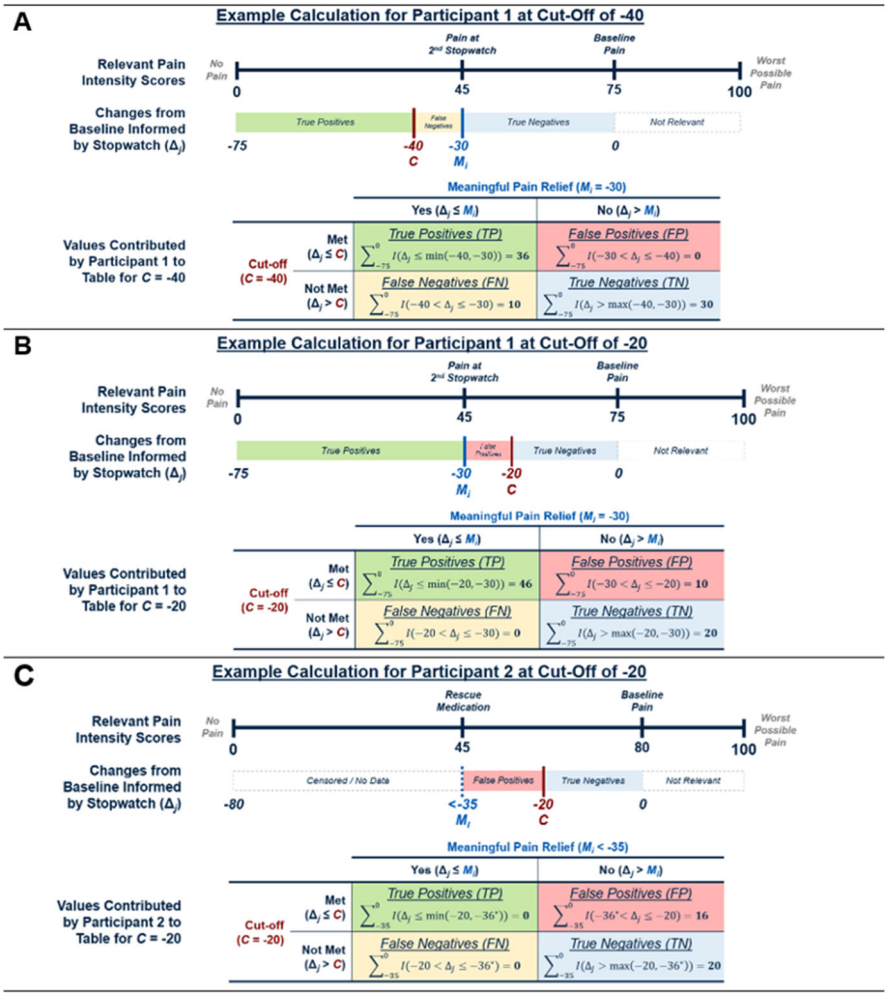

Figure 2.

Example calculations for the stopwatch-based method for individual participants. **(A)** Participant 1 had a baseline VAS of 75 and achieved meaningful relief at a follow-up VAS of 45 ($M_i = 45 - 75 = -30$) and is being evaluated at a cut-off (C) of −40. **(B)** Participant 1 is being evaluated at a cut-off (C) of −20. **(C)** Participant 2 had a baseline VAS of 80 and took rescue at a VAS of 45 without achieving meaningful relief ($M_i < 80 - 45 = -35$) and is being evaluated at a study cut-off (C) of −20. * For this censored case ($M_i < -35$), a value of −36 is used for illustration, though any value less than the pain score at censoring would yield the same result. Baseline is defined as the time study drug is first administered.

Table 1.

Calculations of true positives, false positives, true negatives, and false negatives in a 2×2 Table for a cut-off using the stopwatch-based method

		Meaningful Pain Relief by the Double Stopwatch Method	
		Yes ($\Delta_{ij} \leq M_i$)	**No ($\Delta_{ij} > M_i$)**
Cut-off	**Met ($\Delta_{ij} \leq C$)**	$TP = \sum_i \sum_j I(\Delta_{ij} \leq \min(C, M_i))$	$FP = \sum_i \sum_j I(M_i < \Delta_{ij} \leq C)$
	Not Met$\Delta_{ij} > C$	$FN = \sum_i \sum_j I(C < \Delta_{ij} \leq M_i)$	$TN = \sum_i \sum_j I(\Delta_{ij} > \max(C, M_i))$

Abbreviations: FN, false negative; FP, false positive; MCID, minimum clinically important difference; TN, true negative; TP, true positive.

NOTE. c refers to the cut-off being considered as an MCID for change from baseline in pain intensity. Δ_{ij} refers to the change from baseline in pain intensity score for participant i at pain intensity score j, where j indexes the range of unit-level changes in the pain intensity scale informed by the double stopwatch data for participant M_i refers to the change in pain intensity from baseline at the time of meaningful relief for participant i. Baseline is defined as the time study drug is first administered.

Table 2.

Sensitivity, specificity, and accuracy (95% bootstrap CI) using the conventional method and stopwatch-based method for determining the MCID in the example trial

VAS Cut-off	Sensitivity (95% CI*)	Specificity (95% CI*)	Accuracy (95% CI*)
Conventional Method using Pain Relief Scale at 60 min (N=259)			
−10	0.99 (0.98–1.00)	0.49 (0.40–0.58)	0.76 (0.71–0.81)
−20	0.98 (0.96–1.00)	0.61 (0.52–0.69)	0.81 (0.76–0.86)
−30	0.96 (0.92–0.99)	0.73 (0.65–0.80)	0.85 (0.80–0.89)
−40	0.88 (0.83–0.93)	0.85 (0.78–0.91)	0.87 (0.83–0.91)
−50	0.74 (0.68–0.82)	0.95 (0.91–0.98)	0.84 (0.80–0.88)
Optimal† (−37)	0.93 (0.88–0.97)	0.82 (0.76–0.89)	0.88 (0.84–0.92)
Stopwatch-Based Method (N=352)			
−10	0.99 (0.99–1.00)	0.35 (0.33–0.36)	0.59 (0.57–0.61)
−20	0.96 (0.95–0.97)	0.55 (0.53–0.57)	0.70 (0.69–0.72)
−30	0.88 (0.86–0.89)	0.72 (0.70–0.74)	0.78 (0.76–0.79)
−40	0.74 (0.72–0.76)	0.84 (0.83–0.86)	0.80 (0.79–0.82)
−50	0.53 (0.51–0.56)	0.92 (0.91–0.94)	0.78 (0.76–0.79)
Optimal† (−34)	0.83 (0.81–0.85)	0.77 (0.76–0.79)	0.80 (0.78–0.81)

Abbreviations: CI, confidence interval; MCID, minimum clinically important difference.

*Confidence intervals represent the 2.5th and 97.5th percentiles of 1000 bootstrap replicates for the conventional method and 1000 clustered bootstrap replicates for the stopwatch-based method.

†Optimal cut-off based on Youden's index.