Journal of Clinical Anesthesia 94 (2024) 111366

Contents lists available at ScienceDirect

Journal of Clinical Anesthesia

journal homepage: www.elsevier.com/locate/jclinane

Original Contribution

Minimal clinically important difference: Bridging the gap between statistical significance and clinical meaningfulness

Neel Desai, FRCA [a,b], Eric Albrecht, MD [c,*]

[a] Department of Anaesthesia, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom
[b] King's College London, London, United Kingdom
[c] Department of Anaesthesia, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

HIGHLIGHTS

- NONE as the submitted manuscript is an editorial.

'A difference which makes no difference is no difference at all'
- William James

Clinicians must interpret the findings of trials and systematic reviews and determine if an intervention of interest is of benefit to the patient. In this regard, a patient reported outcome has been defined as one that is reported by the patient without interpretation of their response by a clinician or anyone else, and the patient reported outcome measures (PROM) are instruments, normally self report questionnaires, which are employed to measure a patient reported outcome [1]. Patient reported outcomes are frequently the outcome of most importance to patients and their families. It is erroneous to assume that an intervention is of clinical relevance to the patient simply because first, a difference in a patient reported outcome is present and second, the difference is statistically significant, that is the p value is established to be below a certain threshold. If a difference is statistically significant, then it is not mathematically likely to have occurred by chance should the intervention be ineffective, and it may lead to either a difference that is imperceptible to the patient on the one hand or to a difference which the patient perceives as meaningful on the other hand. In 1989, Jaeschke et al. described the minimum clinically important difference (MCID) as the smallest difference in score in the domain of interest which the patient perceives as beneficial and would mandate, in the absence of troublesome side effects and excessive cost, a change in the management of the patient [2]. The MCID has the potential to bridge the gap between statistical significance and clinical meaningfulness, and is one of the most relevant concepts in contemporary trials and systematic reviews.

For any PROM, such as the pain score, the MCID can be calculated by either the anchor, sensitivity and specificity, distribution or Delphi based approach [3,4]. In the anchor approach, the change in the PROM score is correlated to an external criterion. Importantly, it is debatable if this external criterion or anchor should be an objective clinical end point or another PROM [4–7]. Some argue that the method of anchoring one PROM to another is circular in nature and statistically incorrect [5], but it is our opinion that the PROM should remain anchored in a direct and subjective manner to those who perceive it. In doing so, the anchor is hence likely to have a firm association with the PROM of interest. If the anchor has a very weak or no relationship with the PROM, then the PROM may provide misleading information in determining if a clinically meaningful change has occurred [8]. Studies most commonly compare the PROM scores to the patient's answers to the Global Assessment Rating in which they assess themselves as better, unchanged or worse. The MCID is estimated to be the magnitude of change on the PROM of interest which results in an improvement in the status of the patient on the Global Assessment Rating to a little better or somewhat better (Fig. 1) [4]. One of the drawbacks of the anchor based approach is that the MCID depends on the number of levels present on the scale of the external criterion. Should the scale have a greater number of levels, then the difference between each adjacent level is decreased and the MCID value is reduced. In the sensitivity and specificity based approach, a receiver operating characteristics curve is used to establish the patient reported outcome score with optimal sensitivity and specificity (Fig. 2) [3,4]. The sensitivity is the proportion of patients who report an improvement in the external criterion and whose PROM score exceeds a particular MCID value, and the specificity is the proportion of patients who do not report an improvement in the external criterion and whose PROM score is less than a threshold MCID value.

* Corresponding author.
E-mail address: eric.albrecht@chuv.ch (E. Albrecht).

https://doi.org/10.1016/j.jclinane.2023.111366
Received 11 September 2023; Received in revised form 5 December 2023; Accepted 18 December 2023
Available online 19 January 2024

In the distribution approach, the change in the PROM score is correlated to a measure of statistical variability such as the standard deviation, standard error of measurement, minimum detectable change and the effect size [3,4]. Interestingly, in a systematic review, the MCID in most studies was estimated to be approximately half a standard deviation [9]. The authors explained this relatively consistent association by relating their findings to a study published by Miller in which participants were able to accurately differentiate a number of particular stimuli till the number of categories reached seven, reflecting the uniformity in the limits of human discrimination. They hypothesized that differences in the PROM score which were less than half a standard deviation remained under the discriminatory capacity of patients. The standard error of measurement is the variation in the score due to the unreliability of the PROM, and a difference smaller than the standard error of measurement is therefore likely to represent the presence of measurement error rather than true observed difference [3,4]. If the measurement error, however, has a wide range, then it is possible that a clinically meaningful change may lie within the variation of measurement [8]. One standard error of measurement has been associated with the MCID. The effect size is the change in the PROM score divided by the standard deviation of the baseline score, and the MCID is estimated by multiplying the standard deviation of the baseline score by 0.2, the accepted value for a small effect size [3,4]. One of the shortcomings of the distribution based approach is that it overlooks the fundamental purpose of MCID to relate change to clinical meaningfulness. In the Delphi approach, an iterative and multistep process is followed to achieve a convergence of opinion among experts [10]. They initially receive an open ended or structured survey instrument and, in subsequent rounds of data collection, the responses in the previous round are summarized and experts requested to revise their judgments or specify the reasons for remaining outside the consensus. Some of the weaknesses of the Delphi based approach include the problem of low response rates and investigators influencing and shaping expert opinions through the selection of summarized responses. None of the anchor, sensitivity and specificity, distribution or the Delphi based approaches represent the gold standard and significant variability can be present in their estimation of the MCID. It is thus recommended that multiple methods should be used in a complementary way to triangulate the MCID.

Pain scales such as the numerical rating scale are incomplete and unidimensional constructs of the subjective pain experience for the patient. They are, however, the most widely used metrics of pain following surgery, and are valuable for the assessment and monitoring of postoperative pain. In a systematic review that examined the MCID of the visual analogue and numerical rating scale for acute pain, as determined by the anchor based approach, only eight of the included 37 studies focused on the perioperative period [7]. Of these, three were conducted in the context of dental surgery, one in knee surgery, three in a mixed cohort of surgeries and one in laparotomies. The absolute MCID, on a standardized 100 mm pain scale, ranged from 8 to 40 mm in 30 studies and the relative MCID varied from 13 to 85% in 15 studies. In a prospective observational study which included an unselected cohort of patients recovering from surgery, the MCID of the visual analogue scale for postoperative pain was established with the use of triangulation of the anchor and distribution based methods [11]. The absolute MCID, on a standardized 100 mm pain scale, was 9.9 mm. No studies, to our knowledge, have investigated the MCID for postoperative opioid consumption. The quality of recovery questionnaires provide a global measure of health status following anaesthesia and surgery, and the quality of recovery-15 scale contains items related to emotional state, pain, physical comfort, physical independence and psychological

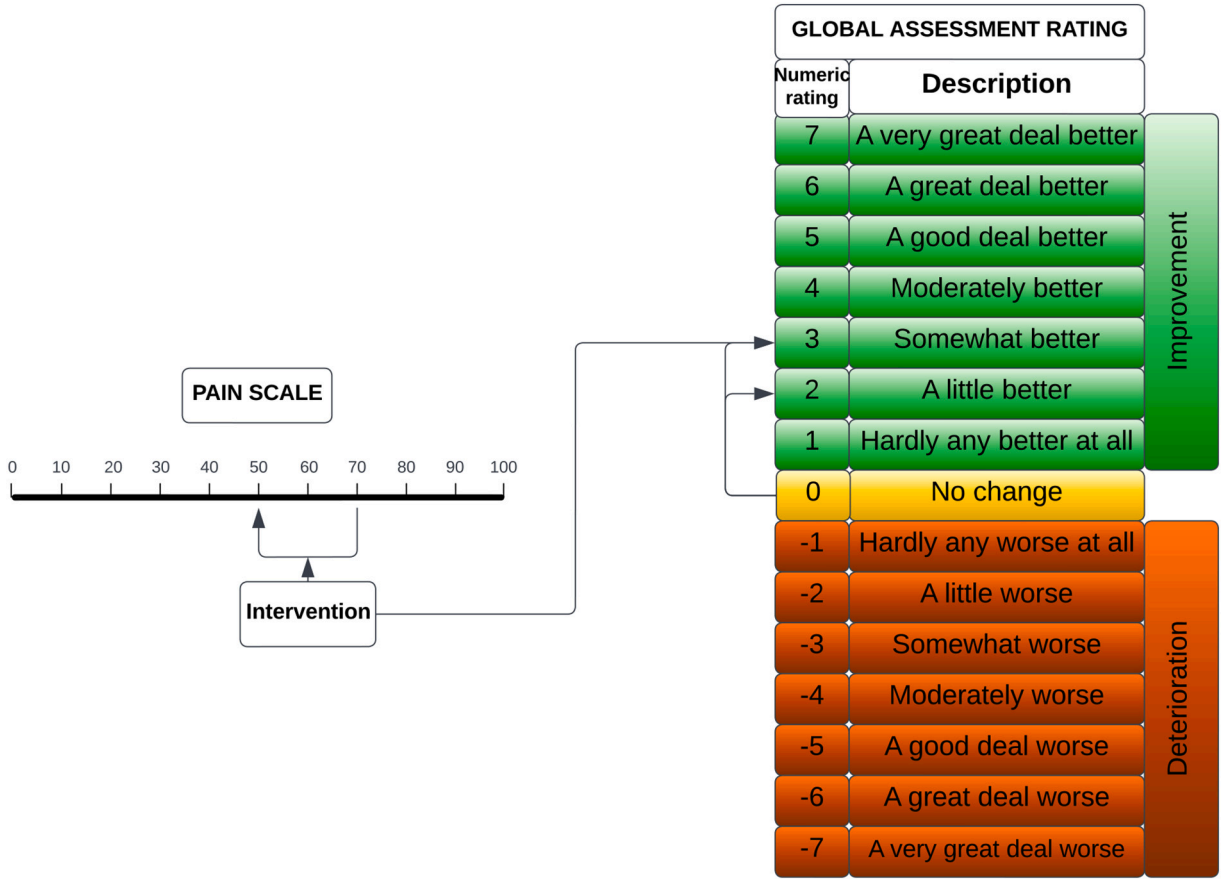

Fig. 1. Illustration of the anchor based approach to calculate the MCID. The MCID is estimated to be 20 mm on a standardized 100 mm pain scale as this magnitude of change leads to an improvement in the status of the patient on the Global Assessment Rating from no change to somewhat better.

support with a maximum score of 150 [12]. In a prospective observational study, performed in a mixed cohort of patients recovering from surgery, and a subsequent reanalysis, the absolute MCID of the quality of recovery-15 score was found with the use of similar triangulation methods to be 6 [11,13]. The MCID has the potential to vary depending on many clinical and methodological factors and is hence population specific. Such influencing factors may include the age and sex of the patient, cultural setting, preoperative disease pathology, duration of disease, presence of preoperative pain, comorbidities, nature of surgery, type of scale used for the PROM, and the postoperative basal analgesic regimen and baseline score on the PROM [14,15]. In the previously mentioned systematic review that explored the MCID for acute pain, one of the studies revealed the MCID to be higher in male compared to female patients [7]. Further, the type of pain scale did not impact on the MCID. They showed that the baseline pain score was associated with the absolute MCID. For each 10 mm increase in the baseline pain score on a standardized 100 mm pain scale, the MCID increased by 3.1 mm and this translated into patients with an initial pain score of less than 40 mm, 40 to 70 mm and more than 70 mm needing a MCID of 6, 13 and 22 mm, respectively.

In the absence of adequate evidence in the literature in regard to population specific MCIDs, clinicians have instead resorted to setting their own clinician perceived minimal important differences (MID) to calculate the sample size of their randomized controlled trials and support the interpretation of the clinical meaningfulness of their results. Importantly, the clinician perceived MID is not the same as the patient perceived MCID. The clinician perceived MID but not the patient perceived MCID is based only on the expectations of the clinicians, previous studies and what is considered to be scientifically or theoretically interesting to the clinicians, and the clinician perceived MID is influenced by pragmatic constraints such as the access to the population of interest, funding and time [16]. In a systematic review, the clinician perceived MID for the pain score at rest and on movement as well as the cumulative intravenous morphine equivalent consumption at 24 h was determined following total hip and knee arthroplasty in 570 randomized controlled trials [17]. For the pain score at rest and on movement, the absolute clinician perceived MID was 15 and 18 mm on a standardized 100 mm pain scale, respectively, and the relative clinician perceived MID for the pain score at rest and on movement was 30%. For the cumulative intravenous morphine equivalent consumption at 24 h, the absolute and relative clinician perceived MIDs were 10 mg and 40%, respectively. In comparison to these clinician perceived MIDs, the patient perceived MCIDs of the visual analogue scale with respect to postoperative pain subsequent to total hip and knee replacement have been established to be greater at 18.6 and 22.6 mm, respectively [18]. Given that these are clinician perceived MIDs rather than patient perceived MCIDs, however, we would suggest clinicians to exercise caution and not set the threshold of the clinician perceived MID unnecessarily high.

In randomized controlled trials, the type one error rate, that is the probability of rejecting the null hypothesis if it is true, power, that is the probability of uncovering a statistically significant result should there be a true difference, standard deviation and the minimum effect size of interest are fundamental in the estimation of the required sample size. The minimum effect size of interest can be the MCID or clinician perceived MID, but it has been reasoned that it does not have to be either of these as long as this minimum effect size remains clinically or scientifically relevant [16]. Importantly, the minimum effect size of interest is equivalent to the minimum detectable difference and trials would not be able to confirm or exclude differences lower than this prespecified value [19]. If the threshold of the clinician MID were to be set inappropriately high, then trials would be unable to identify lesser effects which may be either representative of the determined patient perceived MCID or the as of yet undetermined population specific MCID. In noninferiority and equivalence trials, the MCID establishes the threshold at which noninferiority or equivalence is confirmed. If one intervention does not give rise to a difference in favorable outcomes that is greater than the MCID, then noninferiority is declared. In superiority trials, the MCID establishes the threshold at which superiority is confirmed. If one intervention leads to a difference in favorable outcomes that is greater than the MCID, then superiority is declared. In systematic reviews, the selection of the MCID can result in differences in interpretation of the outcomes. Interestingly, in a meta-analysis that evaluated the erector spinae plane block in surgery for breast cancer, its analgesic efficacy was concluded to be statistically significant yet not clinically meaningful as the differences in outcomes were smaller than

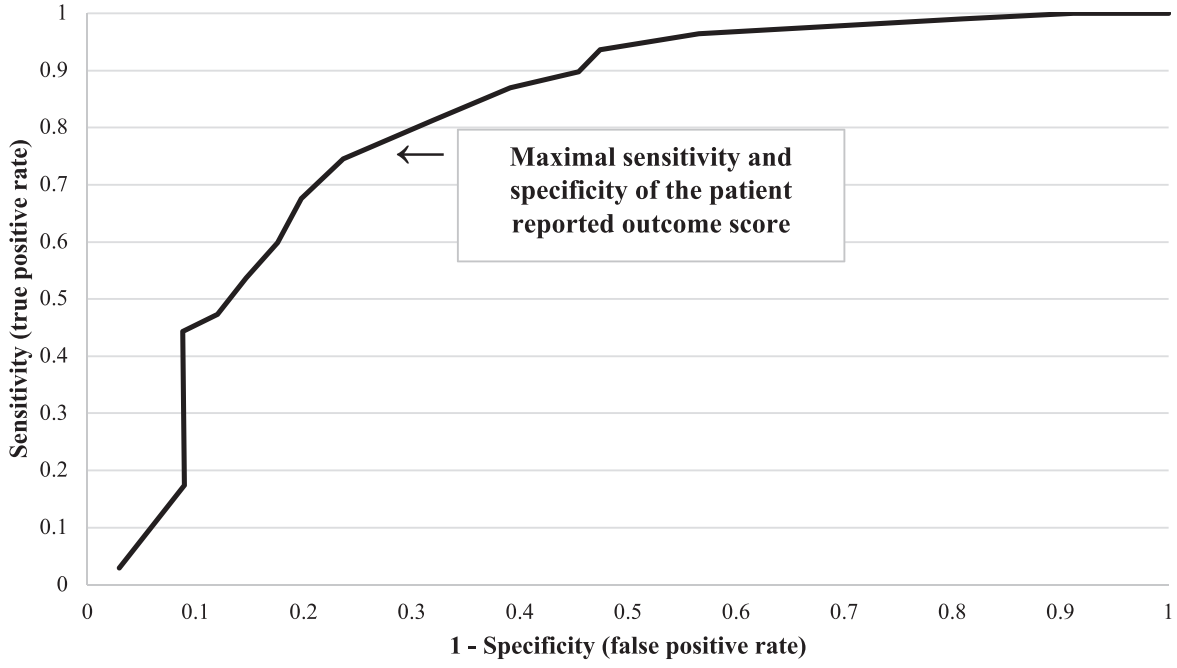

Fig. 2. Illustration of the receiver operating characteristics curve. The MCID is the point on the receiver operator characteristics curve which results in the optimal trade off between sensitivity and specificity.

the prespecified MCID [20]. Subsequent correspondence questioned the value of their MCID and therefore recommended this conclusion to be interpreted with caution [21].

The MCID does have, however, several pitfalls and weaknesses. If the patients were to be at low risk of moderate to severe postoperative pain or the basal analgesic regimen were to reduce the pain score to negligible values, for example, then the effect of a relatively efficacious intervention may not provide a clinically meaningful difference as defined by the absolute MCID. In other words, when viewed through the lens of absolute MCID, patients with a greater baseline pain score are likely to be more responsive to the influence of an intervention of interest. Should the mean difference of a particular outcome in a trial not reach the threshold of the MCID, it is still plausible that many of the patients may have benefitted from the intervention, and this possibility can be characterized by the presentation of responder analysis and number needed to treat [22]. Further, in a placebo controlled trial, the results represent the difference between placebo and intervention and not the difference between no intervention and intervention. The placebo effect is not inconsiderable in magnitude, and thus the effect of an intervention may not be clinically meaningful compared to placebo but may be clinically meaningful relative to no intervention in clinical practice. Moreover, the lack of sufficient population specific MCIDs in the literature has resulted in the application of the MCID from one population to another nonrelated population. In a systematic review of the pectoralis nerve block in surgery for breast cancer, for instance, the MCID for the pain score originated from studies which investigated chronic breast pain [23]. The validity of extrapolating the MCID from chronic or nonsurgical pain to acute perioperative pain remains unknown. Importantly, the MCID is the minimum standard of clinical care we should be striving to achieve for our patients so they feel better. The patient acceptable symptom state (PASS), complementary to the MCID, is the value beyond which patients feel good and well [15]. Even though the original description of MCID did incorporate side effects and cost, its anchor and distribution based derivations do not consider these, but as clinicians we should take account of the side effects and cost as part of a shared decision making process with the patient.

In conclusion, the MCID provides us with a way to interpret the clinical meaningfulness of statistically significant results for our patients. In view of this, there is an urgent requirement for clinicians to establish the MCID for outcomes relevant to specific populations of patients. Until sufficient progress has been made with respect to this, we recommend clinicians, trialists and systematic reviewers to remain cautious and not set the threshold of the MCID unreasonably high to prevent erroneous evaluations of the efficacy of individual anesthetic interventions.

Author statement

Dr. Neel Desai: this author prepared the primary manuscript.
Prof. Eric Albrecht: this author completed the primary manuscript.

Disclosure of funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Neel Desai: Writing – original draft, Writing – review & editing. **Eric Albrecht:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

EA has received grants from the Swiss Academy for Anaesthesia Research (SACAR), Lausanne, Switzerland (no grant numbers attributed), from B. Braun Medical AG, Sempach, Switzerland (no grant numbers attributed) and from the Swiss National Science Foundation to support his clinical research. EA has also received an honorarium from B. Braun Medical AG Switzerland, from Sintetica Ltd. UK and MSD AG Switzerland.

No conflict of interest declared by ND.

References

[1] Johnston B, Patrick D, Devji T, Maxwell L, Bingham III C, Beaton D, et al. Chapter 18: Patient-reported outcomes. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors. Cochrane Handb. Syst. Rev. Interv. Version 6; 2023.
[2] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control. Clin. Trials 1989;10:407–15. https://doi.org/10.1016/0197-2456(89)90005-6.
[3] Bloom DA, Kaplan DJ, Mojica E, Strauss EJ, Gonzalez-Lomas G, Campbell KA, et al. The minimal clinically important difference: a review of clinical significance. Am. J. Sports Med. 2023;51:520–4. https://doi.org/10.1177/03635465211053869.
[4] Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J. 2007;7:541–6. https://doi.org/10.1016/J.SPINEE.2007.01.008.
[5] Gatchel RJ, Mayer TG. Testing minimal clinically important difference: consensus or conundrum? Spine J. 2010;10:321–7. https://doi.org/10.1016/J.SPINEE.2009.10.015.
[6] Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Aaronson N, et al. Methods to explain the clinical significance of health status measures. Mayo Clin. Proc. 2002;77:371–83. https://doi.org/10.4065/77.4.371.
[7] Olsen MF, Bjerre E, Hansen MD, Hilden J, Landler NE, Tendal B, et al. Pain relief that matters to patients: systematic review of empirical studies assessing the minimum clinically important difference in acute pain. BMC Med. 2017;15:35. https://doi.org/10.1186/S12916-016-0775-3.
[8] Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. Health Qual. Life Outcomes 2006;4:70. https://doi.org/10.1186/1477-7525-4-70.
[9] Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med. Care 2003;41:582–92. https://doi.org/10.1097/01.MLR.0000062554.74615.4C.
[10] Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. Pract. Assess. Res. Eval. 2019;12:1–10.
[11] Myles PS, Myles DB, Galagher W, Chew C, Macdonald N, Dennis A. Minimal clinically important difference for three quality of recovery scales. Anesthesiology 2016;125:39–45. https://doi.org/10.1097/ALN.0000000000001158.
[12] Stark PA, Myles PS, Burke JA. Development and psychometric evaluation of a postoperative quality of recovery score: the QoR-15. Anesthesiology 2013;118:1332–40. https://doi.org/10.1097/ALN.0B013E318289B84B.
[13] Myles PS, Myles DB. An updated minimal clinically important difference for the QoR-15 scale. Anesthesiology 2021;135:934–5. https://doi.org/10.1097/ALN.0000000000003977.
[14] Munõz-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;45:1000–5. https://doi.org/10.1136/RAPM-2020-101670.
[15] Tubach F, Wells GA, Ravaud P, Dougados M. Minimal clinically important difference, low disease activity state, and patient acceptable symptom state: methodological issues. J. Rheumatol. 2005;32:2025–9.
[16] Myles PS, Myles DB. Conflating effect size and minimal clinically important difference. Br. J. Anaesth. 2022;128:e1–2. https://doi.org/10.1016/J.BJA.2021.09.018.
[17] Laigaard J, Pedersen C, Rønsbo TN, Mathiesen O, Karlsen APH. Minimal clinically important differences in randomised clinical trials on pain management after total hip and knee arthroplasty: a systematic review. Br. J. Anaesth. 2021;126:1029–37. https://doi.org/10.1016/J.BJA.2021.01.021.
[18] 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. Arthroplast. 2018;33:S71–5.
[19] Gibbs NM, Weightman WM. Beyond effect size: consideration of the minimum effect size of interest in anesthesia trials. Anesth. Analg. 2012;114:471–5.
[20] Hussain N, Brull R, Noble J, Weaver T, Essandoh M, McCartney CJL, et al. Statistically significant but clinically unimportant: a systematic review and meta-analysis of the analgesic benefits of erector spinae plane block following breast cancer surgery. Reg. Anesth. Pain Med. 2021;46:3–12. https://doi.org/10.1136/RAPM-2020-101917.
[21] Versyck B, Pawa A, Chin KJ. Minimal clinically important difference: a context-specific metric. Reg. Anesth. Pain Med. 2021;46:933–4. https://doi.org/10.1136/RAPM-2020-102330.
[22] Myles PS, Myles DB, Galagher W, Boyd D, Chew C, MacDonald N, et al. Measuring acute postoperative pain using the visual analog scale: the minimal clinically important difference and patient acceptable symptom state. Br. J. Anaesth. 2017;118:424–9. https://doi.org/10.1093/BJA/AEW466.
[23] Hussain N, Brull R, McCartney CJL, Wong P, Kumar N, Essandoh M, et al. Pectoralis-II myofascial block and analgesia in breast cancer surgery: a systematic review and meta-analysis. Anesthesiology 2019;131:630–48. https://doi.org/10.1097/ALN.0000000000002822.