BJA

British Journal of Anaesthesia, 126 (5): 1029–1037 (2021)

doi: 10.1016/j.bja.2021.01.021
Advance Access Publication Date: 5 March 2021
Review Article

PAIN

Minimal clinically important differences in randomised clinical trials on pain management after total hip and knee arthroplasty: a systematic review

Jens Laigaard[1,†,*], Casper Pedersen[1,†], Thea Nørgaard Rønsbo[1], Ole Mathiesen[1,2] and Anders Peder Højer Karlsen[1]

[1]Centre for Anaesthesiological Research, Department of Anaesthesia, Zealand University Hospital, Køge, Denmark and [2]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

*Corresponding author. E-mail: Jenslaigaard@hotmail.com

[†]these authors contributed equally

Abstract

Background: Sample size determination is essential for reliable hypothesis testing in clinical trials and should rely on adequate sample size calculations with alpha, beta, variance, and an effect size being the minimal clinically important difference (MCID). This facilitates interpretation of the clinical relevance of statistically significant results. No gold standard for MCIDs exists in postoperative pain research.

Methods: We searched Cochrane Central Register of Controlled Trials, MEDLINE, and Embase for English language articles on randomised trials investigating analgesic interventions after total hip or knee arthroplasty. Primary outcomes were the reported MCIDs for pain score and cumulated rescue opioid consumption. Secondary outcomes included reported sample size calculations and propensity to report statistical significance without reaching MCID. Trend analyses were conducted using statistical process control.

Results: We included 570 trials. Median MCID for 0-24 h opioid consumption was 10 mg i.v. morphine equivalents for absolute reductions (interquartile range [IQR]: 6.8-14.5) and relative 40% (IQR: 30–50%). Median MCIDs for pain scores were absolute 15 mm at rest (IQR: 10–20) and 18 mm during movement (IQR: 10–20) on a 0–100 mm VAS and relative 30% (IQR: 20–30%). No trends were demonstrated for MCIDs. Adequate sample size calculations were reported in 34% of trials. In 46% of trials with statistically significant primary outcomes, the differences did not reach the predetermined MCID.

Conclusions: We provide clinician-perceived MCID estimates for rescue opioid consumption and pain scores that can be used for sample size calculations until reliable evidence-based patient-rated MCIDs emerge. Nearly half of the trials with significant findings did not reach the predetermined MCID.

Keywords: attrition; minimal clinically important difference; numerical analogue scale; patient-controlled analgesia; postoperative morphine consumption; postoperative pain treatment; power calculation; visual analogue scale

Editor's key points

- For key outcomes in trials on postoperative pain, patient-rated thresholds for clinical relevance are unclear.
- This review summarises minimal clinically important differences chosen by authors of 570 RCTs on pain management after total hip and knee arthroplasty.
- These findings may be useful to future trialists until reliable patient-rated evidence emerges.

Received: 28 August 2020 **Accepted:** 10 January 2021

Sample size calculations are essential for the reliability of intervention effects in RCTs, and the designation of minimal clinically important differences (MCIDs) ensures that the effects are clinically meaningful.[1] The MCID for an outcome measure is the smallest difference between groups that a patient would find important.[1,2] Before initiation of RCTs, the MCID should be designated, optimally based on evidence from studies on patient-rated important differences, but often the literature lacks evidence and instead clinician-perceived MCIDs are used.[1–3]

In pain management, MCIDs are controversial. A recent review on acute pain found the patient-rated MCID to be 8–40 mm reductions on a 0–100 mm VAS,[4] and previous studies of unselected cohorts of patients experiencing acute pain have suggested the patient-rated MCID to be 10–13 mm on average.[5,6] For opioid-sparing effects, recent large postoperative trials have used MCIDs of i.v. morphine equivalents 2.5–10 mg 0–24 h postoperatively,[7–9] but the choice is arbitrary, as there is no evidence for a patient-rated MCID. Also, no available guidance from typically used values in previous RCTs exists in the literature, hence MCIDs for opioid-sparing effects are often clinician-perceived. For both opioid and pain MCIDs, the disadvantage of using absolute reductions is that even effectful interventions cannot produce clinically relevant pain relief if the assay sensitivity is low (i.e. if participants are at low risk of moderate-to-severe postoperative pain or co-administered basal analgesic regimens reducing pain to a minimum).[10] In summary, research on relevant MCIDs in postoperative pain trials is needed.

A minimum of four factors must be considered for proper sample size calculations[11–13]: the risk of Type I error (alpha), the risk of Type II error (beta), the within-group variance (for continuous outcomes), and the effect size (the MCID). These factors are often poorly reported in RCTs.[14] Adequate sample size considerations and reporting of sample size calculations are important for the transparency of RCTs and subsequent evaluations of reporting bias in systematic reviews.[14–18] Underpowered RCTs are at high risk of Type II errors,[19] whereas overpowered RCTs raise ethical concerns regarding potential harms, patient inconvenience, unnecessary expenditures, and risk of obtaining statistically significant but clinically irrelevant results.

Total hip arthroplasty (THA) and total knee arthroplasty (TKA) are some of the most frequently performed planned procedures worldwide.[20–22] The procedures are associated with substantial postoperative pain, and numerous RCTs have been published investigating different analgesic treatments.[23,24]

In this systematic review, we aimed to investigate the reported MCIDs in RCTs on postoperative pain management after THA and TKA as markers for clinician-perceived MCIDs. Further, we investigated the adequacy in reporting of sample size calculations, the tendency to report significant but clinically irrelevant differences, and statistical considerations for pain score and opioid consumption outcomes.

Methods

This review was written in coherence with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2009 checklist apart from bias evaluations and meta-analysis, as we aimed to assess statistical strategies regardless of the methodological quality of the trials (Supplementary Appendix 1).[25] The protocol was registered at the PROSPERO database for systematic reviews (CRD42020151317; registered on October 1, 2019). The full protocol and database are available from the corresponding author.

Information sources

Systematic literature searches were performed in Cochrane Central Register of Controlled Trials, MEDLINE, and Embase (Supplementary Appendix 2). Databases were searched up to January 7, 2020 (Supplementary Appendix 2). Identical search strings have been used for other reviews in the same research programme.[24,26]

Eligibility criteria

Eligible trials were published RCTs accessible in English, irrespective of publication year, investigating perioperative medical analgesic interventions for immediate postoperative pain in adults (≥18 yr) undergoing THA or TKA. We excluded conference abstracts, as these rarely report sample size calculations because of limited word count. Also, quasi-randomised trials and trials including fracture surgery, hemiarthroplasties, or bilateral arthroplasties were excluded. Trials were screened for eligibility by three authors independently (CP, JL, and TNR), and any disagreements were resolved by a senior author (APHK or OM).

Data collection process

Data sources were the published articles and attached supplementary materials. Authors were not contacted to obtain further information, as we aimed to assess the actual published material. A pilot data extraction of 20 trials was carried out by four authors (APHK, CP, JL, and TNR) independently to ensure uniformity. Data from the remaining trials were independently extracted into Excel version 16.0 (Microsoft, Redmond, WA, USA) spreadsheets by three authors (CP, JL, and TNR). Discrepancies were detected using conditional formatting, and disagreements were solved by a senior author (APHK or OM).

Study outcomes

The primary outcomes were the chosen MCIDs for pain score assessed at rest, pain score assessed during mobilisation, and cumulated rescue opioid consumption.

Secondary outcomes were (i) reporting of sample size calculations in the included RCTs, including alpha, beta, variance, and MCID; (ii) proportion of trials with a statistically significant, but clinically unimportant, primary outcome; (iii) use of significance levels, *P*-values, and confidence intervals; (iv) use of parametric and non-parametric statistics for outcomes with pain scores and cumulated rescue opioid consumption; and (v) choice of visual presentation (box plot/scatter plot/bar chart/table).

Synthesis of results

Data were presented as percentages or absolute numbers as appropriate. Data were presented as median and inter-quartile range (IQR) for all outcomes, as these are appropriate for non-parametric distributions and correlate well with mean and standard deviation (SD) in parametric distributions. Trend analyses were assessed with statistical process control (SPC) charts using R i386 version 3.6 (https://www.r-project.org/)

and RStudio (https://rstudio.com/), with the package qicharts2 (https://cran.r-project.org/web/packages/qicharts2/). Originally developed for quality control, SPC is specifically developed to monitor changes over time. Changes are detected from the distribution of data points rather than the significance level of a regression line. In SPC, processes are described as random or non-random depending on the progression of data points. A process is non-random when either the length of a 'run' (i.e. the number of consecutive observations on the same side of the mean line) is above a cut-off value, or when the number of times the process crosses the mean line is below a cut-off value. The cut-off values for runs and crossings depend on number of observations in a proportional fashion.[27]

If reported in ≥80 trials, selected predefined outcomes were subgroup analysed for differences over time, continental, procedural (THA/TKA), or interventional differences (systemic analgesics, local infiltration analgesia, neuraxial blockades, and peripheral nerve blocks).

Whenever trials reported effect size in their power calculation, this was interpreted as the chosen MCID. Cumulated opioid consumptions were converted to milligram i.v. morphine equivalents $(24\ h)^{-1}$.[27] All 11-point and 101-point pain scales were converted to millimetres on the 0–100 mm VAS.

Using multivariate logistic regression, we *post hoc* analysed the effect of (i) publication year, (ii) number of participants (per trial arm), (iii) journal impact factor,[29] (iv) multicentricity, and (v) prospective online trial registration on the tendency to report (i) MCID; (ii) an adequate sample size calculation with alpha, beta, variance, and MCID; and (iii) a statistically significant, but clinically unimportant primary outcome. We conducted complete-case analyses using R i386 version 3.6 with *P*-values <0.05 considered statistically significant.

Results

We identified 20 646 records, screened 14 200 abstracts, full-text assessed 1000 eligible articles, and included 570 RCTs (Fig. 1). Trials published between 2011 and 2020 constituted 59% of the included trials. Intervention arms were directly compared without a control group in 45% of the trials. Trial characteristics are shown in Table 1.

Primary outcome

Median MCID for 0–24 h postoperative opioid consumption was i.v. morphine equivalent 10 mg for absolute reductions (IQR: 6.8–14.5; 54 trials) and 40% for relative reductions (IQR: 30–50%; 95 trials).

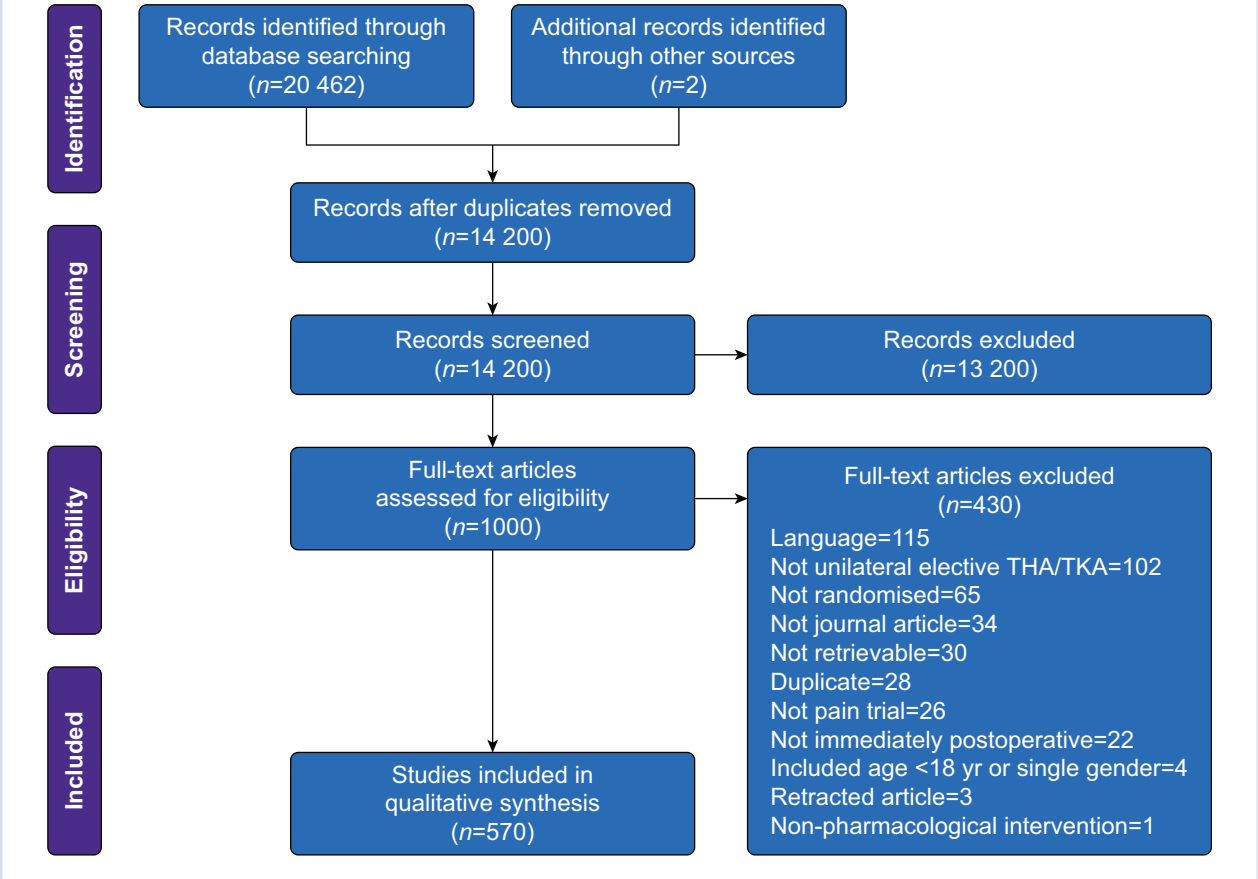

Fig 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram. Minimal clinically important differences in post-arthroplasty pain management research. Latest search performed on January 7, 2020. THA, total hip arthroplasty; TKA, total knee arthroplasty.

Median MCID for absolute reduction in pain scores was 15 mm during rest (IQR: 10–20; 96 trials) and 18 mm during movement (IQR: 10–20; 24 trials). Median MCID for a relative pain reduction was 30% (IQR: 20-40%), but this was rarely reported (36 trials for rest and movement combined).

A total of 98 trials (18%) used MCIDs for other outcomes, including mobility outcomes, duration of analgesia, and length of hospital stay. The propensity of trials using pain score as effect size in the sample size calculation increased over time (Supplementary Appendix 3). The chosen MCIDs for pain scores and rescue opioid consumption were comparable between continents and surgical procedure, whereas the chosen MCID for absolute reduction in pain score was slightly lower for trials investigating systemic analgesics (10 [10–16] mm) compared with other intervention types (Supplementary Appendix 4). In 206 (36%) trials, no MCIDs were mentioned, whereas 37 (6%) trials mentioned more than one MCID as a result of multiple primary outcomes.

Secondary outcomes

An *a priori* sample size calculation was reported in 458 trials (80%) (Table 2). Adequate sample size calculations with reporting of alpha, beta, SD, and MCID were available in 194 trials (34%).

The median sample size was 30 participants per trial arm (IQR: 20–45) with a significant increase over time (Fig. 2). An adjusted larger sample size than the calculated was explicitly chosen to account for attrition and missing data in 205 trials (36%), typically adding 1–20% extra participants to each trial arm. The relative number of added participants did not change over time (Fig. 2).

In 46% of trials with a statistically significant primary outcome, the difference did not reach the predetermined MCID (Table 3).

The proportion of trials reporting a primary outcome with mean change and confidence intervals increased over time (Supplementary Appendix 5). The tendency to use parametric and non-parametric analyses did not change over time (Supplementary Appendix 6). Parametric analyses were more frequently used in trials explicitly testing for data distribution for both pain score (56% *vs* 41%) and rescue opioid consumption (54% *vs* 37%). Scatter plots were the most common graphic presentation of pain scores (31%), whereas opioid consumption was presented predominantly with bar charts (18%) (Supplementary Appendix 7).

In *post hoc* multivariate logistic regression analyses, we found that prospective online trial registration and publication in higher-impact journals were associated with a higher tendency to report MCIDs and adequate sample size calculations (Supplementary Appendix 8). Journal impact factors were unavailable for 82 publications – primarily trials published before 1999.

Discussion

In 570 RCTs on postoperative pain management after THA and TKA, median MCID for opioid consumption was an absolute 10 mg i.v. morphine equivalents and a relative 40%, and for pain scores an absolute 15 mm at rest and 18 mm during movement and a relative 30%. We interpret these as the clinician-perceived MCIDs. *A priori* sample size calculations were used in 80% of trials, but adequate sample size

Table 1 Trial characteristics. Values are no. of trials (proportion of all trials).

	n (%)		*n* (%)
Geographical distribution of studies		*Online trial registration*	197 (35)
Europe	229 (40)	*Type of intervention tested*	
UK	44 (8)	Nerve block	128 (22)
Denmark	32 (6)	Local infiltration analgesia	88 (15)
Germany	23 (4)	Neuroaxial	80 (14)
Italy	20 (4)	NSAID	33 (6)
France	17 (3)	Opioid	30 (5)
Sweden	15 (3)	Gabapentinoid	13 (2)
Finland	12 (2)	Paracetamol	6 (1)
Netherlands	11 (2)	Other drugs	36 (6)
Switzerland	11 (2)	Multiple or mixed	156 (27)
Asia	174 (31)	*Type of comparison*	
China	43 (8)	Placebo	208 (36)
South Korea	22 (4)	No treatment	107 (19)
Turkey	22 (4)	Other intervention	255 (45)
Thailand	20 (4)	*Publication year*	
India	19 (3)	1981–1990	13 (2)
Japan	18 (3)	1991–2000	71 (12)
Taiwan	10 (2)	2001–2010	148 (26)
Singapore	9 (2)	2011–2020	338 (59)
North America	137 (24)	*Type of surgery*	
USA	102 (18)	Total hip arthroplasty	153 (27)
Canada	36 (6)	Total knee arthroplasty	373 (65)
Australia	14 (2)	Both	44 (8)
Australia	13 (2)		

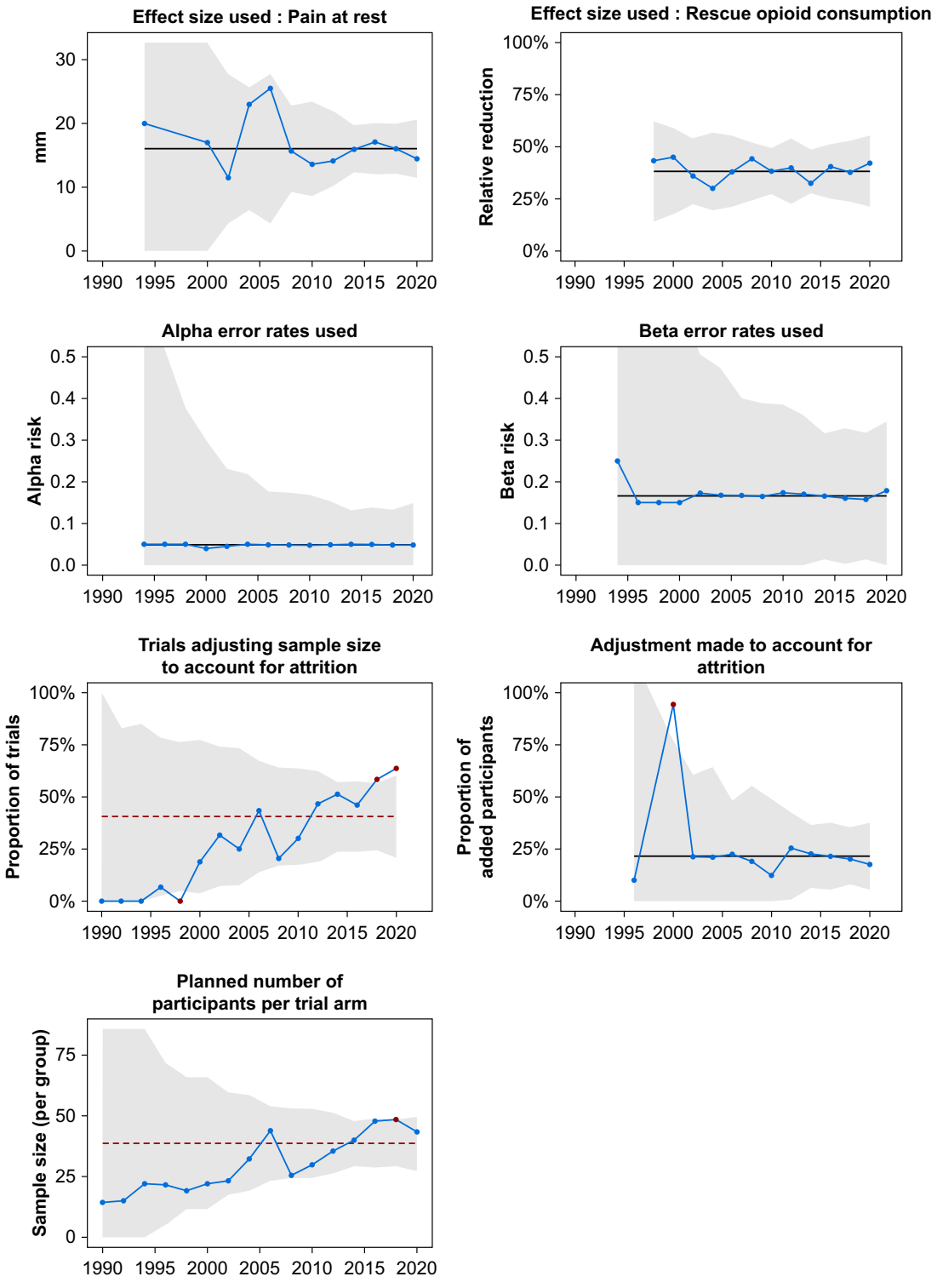

Fig 2. Statistical process control charts for choices of sample size components. Control charts showing trends in chosen effect sizes (absolute for pain score and relative for opioid consumption because those were the only reported in ≥80 trials), alpha and beta error rates, proportion of trials adding a surplus of participants to account for attrition, the average chosen sample size, and the proportion of added participants (blue lines). The point and line graphs are depicted in relation to the mean. Non-random processes (red stippled lines) and unstable processes (red points outside the grey control limits) indicate significant trends. Significant increases over time were observed in average sample size and the proportion of trials adding a surplus of participants to account for attrition. No trends were observed in the remaining charts. The unstable process in the proportion of added participants to account for attrition was not interpreted as a change.

calculations with alpha, beta, variance, and effect size were reported only in 34%.

The clinician-perceived MCIDs for pain score found in our review are comparable with the 17 mm (IQR: 14–23) demonstrated in a recent meta-analysis of 35 acute pain MCID studies, which currently is the highest-quality published evidence on patient-rated pain score MCIDs.[4] The studies included in that review generally used the mean change approach, in which patients were asked at what point they sensed pain relief and where the difference from the anchor score constituted the MCID.[4] In a cohort of 304 patients undergoing THA or TKA, the MCIDs were demonstrated to be 19 and 23 mm, respectively.[30] This is higher than MCIDs used in most of the RCTs included in our review. Both the aforementioned review[4] and cohort study[30] reported pain scores in absolute differences. This resembles our finding of few trials using relative differences for pain scores, which could arise from an inclination amongst triallists to avoid using misleading large relative, but clinically irrelevant, differences in trials with low assay sensitivity. Use of absolute reductions as MCIDs in pain score and opioid consumption outcomes may generate issues with between-trial heterogeneity because of variance in baseline pain risk attributable to differences in populations and basal analgesic regimens. However, methods to account for baseline risk have been proposed for rescue opioid consumption outcomes and could be adapted to pain score outcomes.[10] *Vice versa*, using relative differences may impede the interpretation of clinical relevance in terms of avoiding opioid-related adverse effects, as these seemingly are proportionally related to the absolute amount of administered opioids with a suggested one additional opioid-related adverse event per 3–4 mg of oral morphine consumed.[31] In our opinion, using absolute reductions is the option with fewest downsides.

The need for postoperative opioids can be viewed as an indirect measure of the total pain burden for the investigated period, although other factors, such as reductions in opioid-related adverse effects, may also influence patients' opioid use. We were unable to find studies investigating MCIDs for reductions in postoperative opioid consumption. The patient-deemed relevance of opioid-sparing effects is an important area for future research.

Few reviews have focused on MCID and sample size calculations. A review of sample size calculations in 116 trials of different treatments for hip and knee osteoarthritis found adequate reporting of core components in only 14% of trials.[14] Their core components also included reporting of expected attrition level and use of either one- or two-sided test. However, superiority designs are preferably reported as two-tailed, whereas the convention for non-inferiority designs is one-tailed analyses, therefore we did not include explicit reporting of the number of tails in our core component set. The surplus of participants to account for attrition has been shown to be overestimated in pain research[17,32]; hence, researchers should focus on pragmatic trial designs and inclusion of all participant data in analyses instead.[13] In a review of randomised controlled trials published in top anaesthetic journals in 2013, the reporting of components of the sample size calculation resembled ours.[17]

The importance of distinguishing between significant and relevant results was recently discussed in an editorial by the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials.[33] Discouragingly, in our review, almost half of all trials showing statistically significant effects on primary outcomes failed to reach the predetermined MCID, meaning that these differences *a priori* were deemed clinically unimportant. These statistically, but clinically irrelevant differences were not correlated to small sample sizes, publication in low-impact journals, monocentricity, or lack of prospective trial registration. This finding suggests that this is a widespread problem in pain research, although many trials were not included in the analysis because of absence of designated primary outcomes. Mismatch between significant results and MCIDs is most likely caused by overestimated variance and sample size adjustments that were larger than what was necessary to account for attrition.[17] Conducting sample size calculations without the use of either pilot trials or local data with actual patients receiving treatments similar to that of the control group can cause poor estimation of variance, typically with overestimation, and lead to 'overpowered' trials.[34,35] Further, trials of acute postoperative pain are characterised by low attrition rates because of short follow-up periods.[24] Therefore, we recommend triallists to refrain from large sample size adjustments to minimise significance/MCID mismatch.[14,17]

The clinical relevance of an intervention depends on the effect in clinical practice. In placebo-controlled trials, the result constitutes the difference between the intervention and placebo; still in clinical practice, the alternative to intervention is no intervention.[36] Therefore, seemingly non-relevant intervention effects may reach clinical significance when used in clinical practice via the addition of the placebo effect, which is considerable in pain trials.[36,37] Also, when the mean difference does not reach the MCID, a part of the participants may have still benefitted from the intervention. Therefore, a within-patient MCID analysis, responder analysis, or calculation of number needed to treat may be relevant as a supplement to standard analyses.[33]

The trends in the included RCTs were generally towards clearer reporting of primary outcomes using confidence intervals, larger sample sizes, and better reporting of sample size calculations. Collectively, this reflects a trend towards improved trial methodology in RCTs on analgesic interventions for postoperative pain management after THA and TKA.[13,38]

Strengths and limitations

We chose to include English-written articles only, which may have introduced language bias in the analyses. With a risk of leaving out details, but to ensure focus on the key areas of sample size calculations, we chose not to assess complex parts of the sample size calculation, such as modelling and Bayesian statistics. We did not compare trials and their protocols for methodological discrepancies. Such discrepancies should be assessed during peer review and clearly stated in trials before publication.[39] Data were parallelly extracted to ensure data completeness and minimise typographical errors. Although bias evaluations could have had merits, we know from previous reviews that trials in this field are generally prone to high risk of bias, thus we chose to focus on other methodological issues in the current review.[28,40]

Previous reviews on sample size calculations have included trials from only leading journals or a specific annual range. We included all English-written trials investigating postoperative pain management after THA or TKA regardless of journal or publication year, which makes our review applicable to the entire field of research.[4,17,30] To explore the impact of

Table 2 Reporting of sample size calculations. Values are no. of trials (proportion of all trials).

	n (%)		n (%)
Trials with sample size calculation	458 (80)	*Proportionate adjustment to account for attrition*	
Alpha reported	388 (68)	1–10	53 (9)
0.05	369 (65)	11–20	80 (14)
0.01	7 (1)	21–30	25 (4)
Other	12 (2)	31–40	19 (3)
Beta reported	403 (71)	41–50	11 (2)
0.2	266 (47)	51–100	12 (2)
0.1	93 (16)	101	5 (1)
0.05	16 (3)	*Use of software reported*	380 (67)
0.15	10 (2)	SPSS	219 (38)
Other	18 (3)	SAS	63 (11)
Standard deviation reported	215 (38)	STATA	35 (6)
Effect size reported	364 (64)	GraphPad	16 (3)
Presented as minimal clinically important difference	148 (26)	R	12 (2)
All components reported	194 (34)	*Sample size calculated from*	
		Pilot study	58 (10)
		Data from the department	65 (11)
		Previous studies	134 (24)
		Not mentioned	201 (35)

important characteristics related to trial methodology, we did a *post hoc* regression analysis and found that prospective trial registration and publication in a high-impact journal increased the likelihood of a reported MCID and adequate sample size calculation.

Perspectives

We provide valuable estimates of MCIDs for pain scores and rescue opioid consumption after THA and TKA as provided by authors of RCTs in this field. The MCIDs for pain score demonstrated in this review are supported by the current literature on patient-rated MCIDs. For rescue opioid consumption, no prior evidence was available. Until higher-quality evidence is established on patient-rated MCIDs, we present the current MCID of i.v. morphine 10 mg based on information from 570 trials as an available base for future sample size calculations in RCTs on pain management after THA and TKA surgeries. The relevance of these MCIDs as rated by patients remains to be investigated and should be the focus of future studies. We find it urgent to address mismatches of statistically significant results, which fail to reach the predetermined MCID.

Table 3 Reporting of significance. Values are no. of trials (proportion of all trials).

	n (%)
Significance marker	
Only P-value	479 (84)
Confidence intervals	90 (16)
Number of primary outcomes	
1–2 or mass correction	261 (46)
>2 and no correction	73 (13)
None specified	236 (41)
P-value	
0.05	503 (88)
<0.05	24 (4)
Not specified	43 (8)
Number of trials with significant primary outcome(s)	
Significant	160 (28)
Difference > effect size	86 (15)
Difference < effect size	74 (13)
Not significant	101 (18)

Conclusions

We found that median clinician-perceived MCIDs in postoperative pain management were 10 mg i.v. morphine equivalents or 40% for opioid consumption and 15–18 mm or 30% for pain scores. *A priori* sample size calculations were used in 80% of trials, and adequate sample size calculations with an alpha, beta, variance, and effect size were reported in 34%. Nearly half of the trials with significant findings did not reach their predetermined MCID.

Authors' contributions

Idea conception: JL, CP, OM, APHK
Drafting of protocol and extraction sheet: JL, APHK
Screening of titles and abstracts for eligibility: CP, APHK
Screening of full-text articles for inclusion: JL, CP, APHK
Data extraction: JL, CP, TNR
Data analyses: JL, CP, APHK
Writing of paper: JL, CP, OM, APHK
All authors took part in critically revising the protocol and paper, and take responsibility for the full intellectual content of the paper.

Declarations of interest

None of the authors had conflicts of interest. No external funding was received.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bja.2021.01.021.

References

1. Draak THP, de Greef BTA, Faber CG, Merkies ISJ, PeriNomS Study Group. The minimum clinically important difference: which direction to take. *Eur J Neurol* 2019; **26**: 850–5
2. Sedaghat AR. Understanding the minimal clinically important difference (MCID) of patient-reported outcome measures. *Otolaryngol Head Neck Surg* 2019; **161**: 551–60
3. Munoz-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
4. Olsen MF, Bjerre E, Hansen MD, 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
5. Myles PS, Myles DB, Galagher W, 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
6. Gallagher EJ, Liebman M, Bijur PE. Prospective validation of clinically important changes in pain severity measured on a visual analog scale. *Ann Emerg Med* 2001; **38**: 633–8
7. Gasbjerg KS, Hagi-Pedersen D, Lunn TH, et al. DEX-2-TKA-DEXamethasone twice for pain treatment after total knee arthroplasty: a protocol for a randomized, blinded, three-group multicentre clinical trial. *Acta Anaesthesiol Scand* 2020; **64**: 267–75
8. Thybo KH, Hagi-Pedersen D, Dahl JB, et al. Effect of combination of paracetamol (acetaminophen) and ibuprofen vs either alone on patient-controlled morphine consumption in the first 24 hours after total hip arthroplasty: the PANSAID randomized clinical trial. *JAMA* 2019; **321**: 562–71
9. Hickman SR, Mathieson KM, Bradford LM, Garman CD, Gregg RW, Lukens DW. Randomized trial of oral versus intravenous acetaminophen for postoperative pain control. *Am J Health Syst Pharm* 2018; **75**: 367–75
10. Doleman B, Sutton AJ, Sherwin M, Lund JN, Williams JP. Baseline morphine consumption may explain between-study heterogeneity in meta-analyses of adjuvant analgesics and improve precision and accuracy of effect estimates. *Anesth Analg* 2018; **126**: 648–60
11. Hickey GL, Grant SW, Dunning J, Siepe M. Statistical primer: sample size and power calculations—why, when and how? *Eur J Cardiothorac Surg* 2018; **54**: 4–9
12. Das S, Mitra K, Mandal M. Sample size calculation: basic principles. *Indian J Anaesth* 2016; **60**: 652–6
13. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg* 2012; **10**: 28–55
14. Copsey B, Thompson JY, Vadher K, et al. Sample size calculations are poorly conducted and reported in many randomized trials of hip and knee osteoarthritis: results of a systematic review. *J Clin Epidemiol* 2018; **104**: 52–61
15. McKeown A, Gewandter JS, McDermott MP, et al. Reporting of sample size calculations in analgesic clinical trials: ACTTION systematic review. *J Pain* 2015; **16**: 199–206. e1–7
16. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ* 2009; **338**: b1732
17. Abdulatif M, Mukhtar A, Obayah G. Pitfalls in reporting sample size calculation in randomized controlled trials published in leading anaesthesia journals: a systematic review. *Br J Anaesth* 2015; **115**: 699–707
18. Copsey B, Dutton S, Fitzpatrick R, Lamb SE, Cook JA. Current practice in methodology and reporting of the sample size calculation in randomised trials of hip and knee osteoarthritis: a protocol for a systematic review. *Trials* 2017; **18**: 466
19. Abdullah L, Davis DE, Fabricant PD, Baldwin K, Namdari S. Is there truly "no significant difference"? Underpowered randomized controlled trials in the orthopaedic literature. *J Bone Jt Surg Am* 2015; **97**: 2068–73
20. Inacio MCS, Graves SE, Pratt NL, Roughead EE, Nemes S. Increase in total joint arthroplasty projected from 2014 to 2046 in Australia: a conservative local model with international implications. *Clin Orthop Relat Res* 2017; **475**: 2130–7
21. Pabinger C, Geissler A. Utilization rates of hip arthroplasty in OECD countries. *Osteoarthritis Cartilage* 2014; **22**: 734–41
22. Pabinger C, Lothaller H, Geissler A. Utilization rates of knee-arthroplasty in OECD countries. *Osteoarthritis Cartilage* 2015; **23**: 1664–73
23. Gerbershagen HJ, Aduckathil S, van Wijck AJ, Peelen LM, Kalkman CJ, Meissner W. Pain intensity on the first day after surgery: a prospective cohort study comparing 179 surgical procedures. *Anesthesiology* 2013; **118**: 934–44
24. Koyuncu S, Friis CP, Laigaard J, Anhoj J, Mathiesen O, Karlsen APH. A systematic review of pain outcomes reported by randomised trials of hip and knee arthroplasty. *Anaesthesia* 2021; **76**: 261–9
25. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* 2009; **339**: 332–6
26. Pedersen C, Troensegaard H, Laigaard J, et al. Differences in patient characteristics and external validity of randomized clinical trials on pain management following total hip and knee arthroplasty: a systematic review. *Reg Anesth Pain Med* 2020; **45**: 709–15
27. Anhøj J, Olesen AV. Run charts revisited: a simulation study of run chart rules for detection of non-random variation in health care processes. *PLoS One* 2014; **9**, e113825
28. Karlsen APH, Geisler A, Petersen PL, Mathiesen O, Dahl JB. Postoperative pain treatment after total hip arthroplasty: a systematic review. *Pain* 2015; **156**: 8–30
29. SCImago. *SCImago journal & country rank [portal]* 2020. Available from: http://www.scimagojr.com. accessed November 2020
30. 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**: S71. 5.e2
31. Zhao SZ, Chung F, Hanna DB, Raymundo AL, Cheung RY, Chen C. Dose-response relationship between opioid use and adverse effects after ambulatory surgery. *J Pain Symptom Manage* 2004; **28**: 35–46
32. Flight L, Julious SA. Practical guide to sample size calculations: non-inferiority and equivalence trials. *Pharm Stat* 2016; **15**: 80–9
33. Smith SM, Dworkin RH, Turk DC, et al. Interpretation of chronic pain clinical trial outcomes: IMMPACT recommended considerations. *Pain* 2020; **161**: 2446–61

34. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: basic principles and common pitfalls. *Nephrol Dial Transplant* 2010; **25**: 1388–93
35. Cook J, Julious S, Sones W, et al. Choosing the target difference ('effect size') for a randomised controlled trial—DELTA guidance protocol. *Trials* 2017; **18**: 271
36. Sanderson C, Hardy J, Spruyt O, Currow DC. Placebo and nocebo effects in randomized controlled trials: the implications for research and practice. *J Pain Symptom Manage* 2013; **46**: 722–30
37. Colloca L, Barsky AJ. Placebo and nocebo effects. *N Engl J Med* 2020; **382**: 554–61
38. Karlsen APH, Dahl JB, Mathiesen O. Evolution of bias and sample size in postoperative pain management trials after hip and knee arthroplasty. *Acta Anaesthesiol Scand* 2018; **62**: 666–76
39. Mathieu S, Chan AW, Ravaud P. Use of trial register information during the peer review process. *PLoS One* 2013; **8**, e59910
40. Karlsen AP, Wetterslev M, Hansen SE, Hansen MS, Mathiesen O, Dahl JB. Postoperative pain treatment after total knee arthroplasty: a systematic review. *PLoS One* 2017; **12**, e0173107

Handling editor: Jonathan Hardman