Optimal Strategies for Reporting Pain in Clinical Trials and Systematic Reviews: Recommendations from an OMERACT 12 Workshop

Jason W. Busse, Susan J. Bartlett, Maxime Dougados, Bradley C. Johnston, Gordon H. Guyatt, John R. Kirwan, Kent Kwoh, Lara J. Maxwell, Andrew Moore, Jasvinder A. Singh, Randall Stevens, Vibeke Strand, Maria E. Suarez-Almazor, Peter Tugwell, and George A. Wells

ABSTRACT. *Objective.* Pain is a patient-important outcome, but current reporting in randomized controlled trials and systematic reviews is often suboptimal, impeding clinical interpretation and decision making.

Methods. A working group at the 2014 Outcome Measures in Rheumatology (OMERACT 12) was convened to provide guidance for reporting treatment effects regarding pain for individual studies and systematic reviews.

Results. For individual trials, authors should report, in addition to mean change, the proportion of patients achieving 1 or more thresholds of improvement from baseline pain (e.g., ≥ 20%, ≥ 30%, ≥ 50%), achievement of a desirable pain state (e.g., no worse than mild pain), and/or a combination of change and state. Effects on pain should be accompanied by other patient-important outcomes to facilitate interpretation. When pooling data for metaanalysis, authors should consider converting all continuous measures for pain to a 100 mm visual analog scale (VAS) for pain and use the established, minimally important difference (MID) of 10 mm, and the conventionally used, appreciably important differences of 20 mm, 30 mm, and 50 mm, to facilitate interpretation. Effects ≤ 0.5 units suggest a small or very small effect. To further increase interpretability, the pooled estimate on the VAS should also be transformed to a binary outcome and expressed as a relative risk and risk difference. This transformation can be achieved by calculating the probability of experiencing a treatment effect greater than the MID and the thresholds for appreciably important differences in pain reduction in the control and intervention groups.

Conclusion. Presentation of relative effects regarding pain will facilitate interpretation of treatment effects. (First Release May 15 2015; J Rheumatol 2015;42:1962–70; doi:10.3899/jrheum.141440)

Key Indexing Terms:
OMERACT	OUTCOMES	PAIN	CLINICAL TRIALS
VISUAL ANALOG SCALE			SYSTEMATIC REVIEWS

From the Michael G. DeGroote Institute for Pain Research and Care, McMaster University; Department of Anesthesia, McMaster University; Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario; Department of Medicine, McGill University, Divisions of Rheumatology and Clinical Epidemiology, Royal Victoria Hospital, Montreal, Quebec, Canada; Paris Descartes University, APHP Cochin Hospital, Rheumatology Department, Cochin Hospital, INSERM (U1153) Clinical Epidemiology and Biostatistics, PRES Sorbonne Paris-Cité, Paris, France; The Hospital for Sick Children Research Institute, Toronto; Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto; Department of Anesthesia and Pain Medicine, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; University of Bristol Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol, UK; University of Pittsburgh and Veterans Affairs (VA) Pittsburgh Healthcare System, Pittsburgh, Pennsylvania, USA; Institute of Population Health, University of Ottawa, Ottawa, Ontario, Canada; Pain Research, University of Oxford, Nuffield Division of Anaesthetics, The Churchill, Oxford, UK; Birmingham VA Medical Center and University of Alabama at Birmingham, Birmingham, Alabama; Inflammation and Immunology Clinical Research, Celgene Corporation, Summit, New Jersey; Department of Medicine, Division of Rheumatology, the State University of New Jersey, New Brunswick, New Jersey; Division of Immunology/Rheumatology, Stanford University, Palo Alto, California; University of Texas MD Anderson Cancer Center, Houston, Texas, USA; Department of Medicine, Faculty of Medicine, Institute of Population Health, University of Ottawa; Ottawa Hospital Research Institute, Clinical Epidemiology Program; Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada.

J.W. Busse, DC, PhD, Michael G. DeGroote Institute for Pain Research and Care; Department of Anesthesia, Department of Clinical Epidemiology and Biostatistics, McMaster University; S.J. Bartlett, PhD, Department of Medicine, McGill University, Divisions of Rheumatology and Clinical Epidemiology, Royal Victoria Hospital; M. Dougados, MD, Paris Descartes University, APHP Cochin Hospital; Rheumatology Department, Cochin Hospital, INSERM (U1153) Clinical Epidemiology and Biostatistics, PRES Sorbonne; B.C. Johnston, PhD, Department of Clinical Epidemiology and Biostatistics, McMaster University and Hospital for Sick Children Research Institute, Institute of Health Policy, Management and Evaluation, University of Toronto, and Department of Anesthesia and Pain Medicine, Hospital for Sick Children, University of Toronto; G.H. Guyatt, MD, MSc, Department of Clinical Epidemiology

and Biostatistics, McMaster University; J.R. Kirwan, MD, FRCP, University of Bristol Academic Rheumatology Unit, Bristol Royal Infirmary; K. Kwoh, MD, University of Pittsburgh and VA Pittsburgh Healthcare System; L.J. Maxwell, MSc, Institute of Population Health, University of Ottawa; A. Moore, DSc, Pain Research, University of Oxford, Nuffield Division of Anaesthetics; J.A. Singh, MBBS, MPH, Birmingham Veterans Affairs Medical Center and University of Alabama at Birmingham; R. Stevens, MD, Inflammation and Immunology Clinical Research, Celgene Corporation, and Department of Medicine, Division of Rheumatology, State University of New Jersey; V. Strand, MD, Division of Immunology/Rheumatology, Stanford University; M.E. Suarez-Almazor, MD, PhD, University of Texas, MD Anderson Cancer Center; P. Tugwell, MD, MSC, FRCPC, Department of Medicine, Faculty of Medicine, Institute of Population Health, University of Ottawa, and Ottawa Hospital Research Institute, Clinical Epidemiology Program, Department of Epidemiology and Community Medicine, University of Ottawa; G.A. Wells, PhD, Department of Epidemiology and Community Medicine, University of Ottawa.

Address correspondence to J.W. Busse, Department of Anesthesia, Michael G. DeGroote School of Medicine, McMaster University, Hamilton, Ontario, Canada. E-mail: bussejw@mcmaster.ca

Randomized controlled trials (RCT) and systematic reviews can provide important direction for clinical decision making, but their usefulness may be compromised by failing to report results that provide interpretable estimates of the magnitude of effect. Pain is a common outcome reported among clinical trials. There are, however, many ways to measure this domain — including use of a number of instruments that are unfamiliar to many clinicians and patients. Pain is typically reported as a continuous measure, which further complicates interpretation of treatment effect results. Here, we offer suggestions regarding how best to report treatment effects on pain in both individual studies and systematic reviews (Table 1). Our suggestions are informed by a workshop convened at the 2014 Outcome Measures in Rheumatology (OMERACT) 12 conference.

What Effect Measures Do Clinicians Find Most Useful?

Clinicians generally find dichotomous presentation of

Table 1. Summary of recommendations.

- Pain should be reported directly by patients.
- Global assessments of pain are preferable to assessment of multiple components of pain.
- The effect on pain should be accompanied by presentation of treatment effects on other patient-important outcomes.
- Individual trials should report the proportion of patients achieving a percentage reduction from baseline pain, a desirable pain state, and/or a combination of change and state.
- Metaanalyses should convert all continuous measures for pain to a 10 cm/100 mm visual analog scale (VAS) for pain, report the pooled mean change, and the pooled mean change divided by the minimal (1 cm/10 mm), appreciable (2 cm/20 mm and 3 cm/30 mm), and substantial (5 cm/50 mm) difference in pain improvement.
- To further increase interpretability, the pooled estimate on the VAS for pain should also be transformed to a binary outcome and expressed as a relative risk and risk difference using these same thresholds.

continuous outcomes more useful[1]. A number of studies have documented clinicians' reactions to presentations of binary outcomes as a relative risk, absolute risk, and number needed to treat (NNT). Physicians presented with the relative change in outcome rate are likely to perceive a therapy more effective than if the same data are presented with the absolute change (risk difference) or NNT[2,3,4]. The NNT has been advanced by some as the most helpful measure of association[5,6]; however, some patient and physician surveys have found that lay people[7] and medical doctors[8] have difficulty grasping the concept of NNT. Some evidence suggests that presenting binary outcomes as natural frequencies (a reduction of adverse events is presented as 3 in 100 rather than 3% or the associated NNT, 33) may be the best way to achieve understanding in a variety of audiences[9,10,11]. But other studies suggest that when event rates are sufficiently high (> 1% chance of occurring), the percent change may be more easily grasped than natural units[12]. Our subsequent recommendations are informed by these results.

Clinical Trials

Pain should be reported directly by patients. Pain is a common complaint among patients seeking care. It is a patient-important outcome when it is reported directly by patients, without interpretation by physicians or other proxies. A review of RCT that explored the effect of opioids for chronic non-cancer pain (n = 161) found that while almost all trials (98.8%) reported pain as an outcome measure, 1 trial reported pain data only as observed by clinicians, 6 reported pain data from both patients and clinicians, and the source of pain data was not clear in 26[13]. Although there are rare exceptions involving patients with limited ability to communicate, pain measures should be acquired directly from patients, and trialists should make this explicit when reporting pain data.

Capturing a global assessment of pain is preferable to multiple pain items. Pain may have many different features (e.g., burning, stabbing, aching) and may be associated with both a specific condition under study (e.g., osteoarthritis of the hip) and comorbidity. Trialists are faced with a choice of whether to try to collect all facets of pain among enrolled patients, or to capture a global assessment of pain (e.g., is your overall pain a lot worse, a little worse, the same, a little better, a lot better?). To the extent that patients will be most interested in how interventions will reduce the overall effect of pain, and that exploration of pain characteristics may place a burden on patients that provides little insight into the effect of treatment — both of which are likely to be the case — global assessments will be preferable in most circumstances.

Trialists should facilitate interpretability of pain outcome data. There are outcome domains that patients with painful complaints are interested in, beyond pain relief, particularly for chronic pain. The Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) has recommended that trialists exploring strategies for managing

pain-related complaints consider 8 outcome domains, in addition to pain relief[14,15,16,17]: (1) physical functioning; (2) emotional functioning; (3) participant rating of improvement and satisfaction with treatment; (4) adverse events; (5) participant disposition (for example, adherence to the treatment regime and reasons for premature withdrawal from the trial); (6) role functioning; (7) interpersonal functioning; and (8) sleep and fatigue. From a practical perspective, trialists will have to balance the competing demands of exhaustive outcome data collection with study feasibility.

The effect on pain should be accompanied by presentation of treatment effects on other patient-important outcomes, such as adverse events, function, and sleep, because similar effects cannot be assumed. For example, among trials of opioids versus placebo for chronic non-cancer pain reviewed by Furlan, *et al*, the effect size was twice as large for pain relief [standardized mean difference (SMD) = –0.60; 95% confidence interval (CI) = –0.69 to –0.50) vs improvement in function (0.31; 95% CI = 0.41 to 0.22)][18].

Pain is typically recorded as a continuous outcome measure, and trialists can present the effect of a given intervention on pain in multiple ways. Some may simply indicate whether the effect on pain was statistically significant. It is commonly assumed that a p value ≤ 0.05 is indicative of an important finding; however, the p value does not take into account the size of the observed effect. The clinical implications of a particular study depend on the magnitude of effect and the associated measure of precision (typically 95% CI) and these estimates can have large or small p values, depending on the sample size and number of events.

Many trials report the effect on pain as a mean change with an associated measure of precision, such as an improvement of 10 mm on a 100-mm visual analog scale (VAS) for pain. Such an effect may be statistically significant, but is it important to patients? Even when the mean value represents an effect that is important to patients, many clinicians will extrapolate this effect to all patients[19]; however, treatment response will differ among patients: some will experience benefit greater than the mean difference; some less. Rather than focusing exclusively on the mean difference, examining the difference in the proportion of patients who report an important reduction in their pain, or who have achieved a threshold of acceptable pain, provides complementary information. The differences in these proportions yields a risk difference that one can convert to an NNT — the number of patients that need to receive treatment to achieve an important benefit in 1 patient.

The minimal important difference. One way to dichotomize continuous data is to use the smallest change in an instrument score that patients perceive is important — the minimal important difference (MID). The term minimum clinically important difference (MCID) is also used; however, this terminology focuses on clinicians' perceptions versus those of patients[20]. Establishing the MID requires comparison with an independent standard or anchor that is itself interpretable, and to which the instrument measuring pain is at least moderately correlated. An anchor should focus on measures of improvement informed by a patient's own experiences (e.g., an appreciable improvement in symptoms, return to function, or global response to treatment).

Although it is tempting to conclude that mean differences less than the MID are not worthwhile, and mean differences exceeding the MID suggest that most or all patients will benefit from treatment, this conclusion is misguided[21]. Consider an example where the MID is 0.50 and patients' mean improvement versus control is 0.25. This could mean that 75% had no improvement and 25% experienced a mean change of 1.0, which would result in an NNT of 4, a clearly important benefit.

What to do if the MID is not known. An anchor-based MID has not been established for many continuous outcome measures used to assess pain; however, investigators can still provide estimates of the proportion benefiting and the corresponding NNT. One option is to assume that one-half the baseline SD of the instrument score represents the MID[22]. However, although this represents a moderate effect size, there is evidence that anchor-based and distribution-based MID may differ[23,24]; further, an anchor-based MID directly captures the patients' values and preferences[25,26]. A more satisfactory approach is to convert pain measures to a single instrument for which an anchor-based MID has been established (see below).

Choosing a threshold that is important to patients. The MID may seem like an obvious choice to establish a threshold for meaningful change in pain when measured as a continuous outcome; however, both clinicians and patients may be interested in the ability of a given intervention to provide more than a minimally important difference[27] — to produce improvement that allows patients to feel appreciably, not just minimally, better[28]. Minimal improvements in pain may not be associated with discernable improvements in function, and some evidence suggests that for patients with chronic non-cancer pain, treatment effects on function are only half as large as treatment effects for pain[18]. For this reason, a number of authoritative groups, including many Cochrane groups who are focused on pain, have suggested that trialists and review authors consider not the MID for pain ($\geq 10\%$ reduction from baseline[29,30]), but $\geq 20\%$, $\geq 30\%$, or $\geq 50\%$ reduction from baseline as improvement that is likely to be appreciably important to patients[31].

In the absence of consensus on what constitutes a patient-important threshold in pain relief, it is reasonable to provide a range of options. To provide guidance in this regard, participants of the 2014 OMERACT Workshop advocated for reporting either an appreciable reduction from baseline pain (e.g., 20%, 30%, or 50%), achievement of a desirable pain state (e.g., no worse than mild pain[32], a patient acceptable state[28]), or a combination of change and state:

"My pain has improved and I feel good". Choosing different thresholds for treatment effect may influence the level of statistical significance, and trialists should therefore choose and justify their threshold in advance of their analyses. However, at least in some circumstances[33], and perhaps in most[34], the choice of threshold does not affect the magnitude of relative effect.

Duration of followup. The duration of followup for measuring pain relief is another source of variability among trials. It stands to reason that studies of acute pain should consider shorter time frames and trials of chronic pain should implement longer followup assessments. This should be done in part to inform tolerability of therapy in the longer term, and because outcomes such as quality of life and functional gains require sufficient time to manifest among treatment responders. However, RCT of opioids for chronic non-cancer pain have not measured outcomes for longer than 16 weeks[13], and many chronic pain trials have reported effect estimates at timepoints that most patients would consider unimportant. For example, we are aware of 5 RCT that explored management of chronic post-stroke pain that captured outcome data at ≤ 1 hour after treatment[35,36,37,38,39]. As a general rule, we would suggest that trials enrolling chronic pain patients should capture outcomes at least up to 6 months, and ideally up to 1 year, after treatment. Further, systematic reviews of chronic pain RCT should exclude trials that have followed patients for less than 2 weeks after treatment.

Systematic Reviews

Trials using the same outcome measure. When a common outcome measure for pain is reported among trials, reviewers can preserve the natural units of measure when pooling across trials by calculating the weighted mean difference. Unless the instrument is very familiar, the effect may not be easy to interpret without copresentation of meaningful thresholds such as the MID and appreciable important differences, and even with such context readers may mistakenly attribute the mean effect on pain to all patients[19].

Trials using different outcome measures. This is the more common situation that systematic review authors will face[40]. A recent review of RCT that explored management of fibromyalgia (FM) found that eligible studies that identified pain (n = 241) reported 75 different measures of this outcome (Appendix 1)[41].

There are 5 strategies available to pool different measures that address a common outcome domain:
1. Convert to SMD: As a first approach, different measures of pain can be pooled by converting to SMD, which is the approach recommended by the Cochrane Collaboration[42]; however, this measure of effect is difficult to interpret[1,43] and is affected by differences in baseline heterogeneity among study populations. Greater heterogeneity among pain scores at baseline will result in a smaller SMD versus studies that enroll patients that provide more homogeneous scores,

even when the true underlying effect on pain is the same (Figure 1).
2. Convert to a single instrument: A second approach is to convert different instruments that measure pain into a single, most familiar instrument and the associated estimate of precision[44]. For example, our review of FM found that, among 75 different instruments for reporting pain used, the 10 cm/100 mm VAS was the most commonly reported[41]. There are 2 statistical approaches to convert multiple instruments to a common measure: (A) Multiply SD units × SD of the most familiar instrument. Limitations of this approach include challenges in deciding which SD to use; and the process remains vulnerable to differences in variability of patients' pain scores across studies (Figure 1); or (B) Rescale to units of the most familiar instrument. Both these approaches remain vulnerable to challenges with interpretation and misinterpretation because the mean effect may be mistakenly applied to all patients. For pain, the most familiar instrument is the widely used 10 cm/100 mm VAS. For this instrument the MID has been established as about 1.0 cm/10 mm[29], regardless of pain severity[30]. Although providing readers with the MID facilitates interpretation, review authors should caution readers against dismissing effects less than 1 MID unit, and should provide guidance for interpreting magnitude of effect. If the MID is 1.0 cm/10 mm and the mean difference between treatments is 0.9, clinicians may infer that nobody benefits from the intervention; if the mean difference is 1.1, they may conclude that everyone benefits. Both conclusions are problematic as they ignore the distribution of benefit between individuals. We suggest the following guide for interpretation given a 1.0 MID: if the pooled estimate is ≥ 2.0, and one accepts that the estimate of effect is accurate, this suggests a large effect. If the pooled estimate is between 1.0 and 1.9, many patients may gain important benefits from treatment. If the estimate of effect lies between 0.5 and 1.0, the treatment may benefit an appreciable number of patients. Effects ≤ 0.5 units suggest a small or very small effect.
3. Calculate a ratio of means: A third approach is to calculate a ratio of means, which has the advantage of facilitating pooling continuous outcomes expressed in different units without relying on SD units[45]. This effect estimate is also reasonably straightforward in its interpretation. A ratio of means of 0.75 conveys a relative risk reduction in pain of 25% between those treated and those in the control group. This effect estimate requires a natural 0, which means this method cannot be used when the control group changes for the worse, and the treatment group for the better.
4. Present results in MID units: A fourth approach is, rather than present results in SD units, to present them in MID units. This allows a more direct inference than presenting in natural units and informing readers about the MID. As above, an effect of < 0.5 MID units suggests small or very small effect.
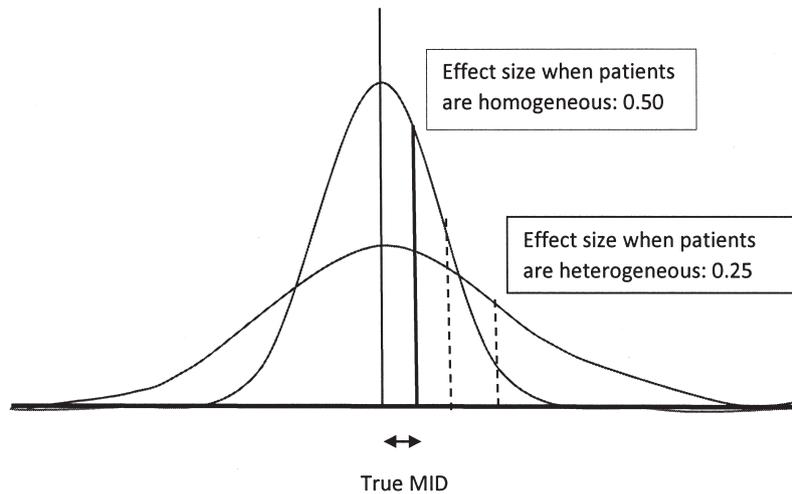5. Apply statistical methods to estimate odds: A fifth approach

Figure 1. Effect of patient heterogeneity on the standardized mean difference. MID: minimally important difference.

is to use statistical methods to provide an estimate of the odds, or probability, of achieving a desirable outcome in the intervention versus the control group. There are 2 fundamental statistical approaches to making this calculation: (A) Convert the SMD into a proportion that confers benefit. Limitations of this approach include the underlying vulnerability of the SMD to population heterogeneity, challenges with interpreting to what the proportion refers (e.g., large or moderate reduction in pain vs minor or no reduction in pain; any reduction in pain vs no reduction in pain), and the requirement for an approximate normal distribution of data and equal variance in intervention and control groups[46]. A final limitation is that the methods demand specification of the success (or failure) rate in the control proportion, and this may not be clear. This is a serious limitation only if the control success or failure rate is likely to be extreme, because the effect estimates differ appreciably only at extremes (< 0.2 or > 0.8; Table 2); or (B) Create a binary outcome, and thus an OR or risk difference (an approach we advocate), avoiding the challenges associated with reliance on SD units. This method uses mean differences and the associated variances in each study to estimate the proportion of patients who achieved an improvement of the MID or greater in that study[44]. To provide insight regarding the proportion of patients who achieve appreciable versus minimally important pain relief, review authors should also present pooled relative and absolute effect estimates using thresholds of 2 cm/20 mm, 3 cm/30 mm, and 5 cm/50 mm[31].

Choosing a strategy to present treatment effect on pain.

Systematic review authors can opt to present the effect of therapy on pain in multiple ways, or select a single measure of effect. Consider a metaanalysis of prophylactic dexamethasone for laparoscopic cholecystectomy that explored the effect on postoperative pain[47]. The effect was informed by 5 RCT that enrolled 539 participants, and certainty in effect estimates was considered "low" according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria because of inconsistency of results across studies, and imprecision associated with pooled estimates of effect. Table 3 presents the effect of dexamethasone on postoperative pain using alternative strategies discussed above, which results in a wide range of effect sizes; from large (SMD of 0.8) to small (0.4 MID units). One reason for this is the likely enrollment of homogeneous patients, resulting in an artificially large SMD (Figure 1).

This example suggests that the presentation of effect estimates for pain reduction using multiple formats has the potential to confuse readers. Accordingly, we believe that if there is strong evidence to inform the anchor-based MID (appreciable and/or substantial thresholds for improvement with a given pain measurement instrument), that systematic review authors should restrict their presentation of effect estimates to approaches relying on these thresholds. Ideally, review authors will convert all continuous measures for pain to a 10 cm/100 mm VAS for pain and inform readers of the established MID of 1 cm/10 mm, and/or the conventionally used appreciably important differences of 2 cm/20 mm, 3 cm/30 mm, and 5 cm/50 mm (See Approach 2 above).

Table 2. The relation between effect size or standardized mean difference and the number needed to treat (NNT) under normality and equal variance assumptions.

Effect size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
NNT		20	13	9	7	5	5	4	4	3

Table 3. Alternate strategies for presenting the effect on pain.

Effect on Postoperative Pain	Estimated Risk or Estimated Score/value with Placebo	Absolute Reduction in Risk or Reduction in Score/value with Dexamethasone	Relative Effect (95% CI)	Comments
Standardized mean difference (SMD)	The pain score in the dexamethasone groups was on average 0.79 SD (1.41 to 0.17) lower than in the placebo groups		—	As a rule to thumb, 0.2 SD represents a small difference, 0.5 a moderate, and 0.8 a large.
Natural units (converted from SD units)	Mean scores with placebo ranged from 43 to 54	Mean scores in the dexamethasone groups was on average 8.1 (1.8 to 14.5) lower	—	The MID on the 0 to 100 pain scale is about 10. Two possible statistical approaches, one relying on the SMD and the other on direct conversion of all instruments to units of the most familiar instrument.
Dichotomizing using "substantial postoperative pain" vs other	20 per 100	Difference in proportion achieving an important improvement is 0.15 (0.19 to 0.04)	RR = 0.25 (0.05 to 0.75)	Two possible statistical approaches, one going directly from SMD to risk difference and the other using the MID to calculate risk differences in individual studies and then pooling across studies.
Ratio of means	28.1	3.7 lower pain score (6.1 to 0.6)	ROM = 0.87 (0.78 to 0.98)	The ROM is the weighted average of the mean pain score in the treatment groups divided by the mean pain score in placebo.
MID units	The pain score in the dexamethasone groups was on average 0.40 (0.74 to 0.07) MID units less than placebo			An effect of less than half the MID difference suggests a small or very small effect.

MID: minimally important difference; RR: risk ratio.

Authors should also use these thresholds to convert the continuous variable to a binary outcome and present the pooled relative and absolute effects (Approach 5 above).

Reporting pain in a GRADE summary of findings table. The GRADE system is an explicit approach to evaluate the certainty of treatment effect estimates[48]. Part of the GRADE process involves presenting the results of systematic reviews in a summary of findings (SoF) table — a succinct presentation of evidence quality and magnitude of effects[49]. GRADE has been adopted by over 70 organizations worldwide, including the World Health Organization, the Cochrane Collaboration, and the American College of Physicians, and now provides detailed guidance on application of GRADE criteria for preparing SoF tables for continuous outcomes. Table 3 demonstrates the presentation of results in SoF tables.

A rule-of-thumb for generating an SoF table is to restrict the number of outcomes presented to a maximum of 7 per table. Attendees of the 2014 OMERACT conference voted (30 to 8) that, for conditions in which pain is the defining feature, 2 SoF rows should be considered for pain-related outcomes. IMMPACT has recommended that 9 outcome measures, including pain, should be reported when assessing treatment effects for clinical conditions defined by pain (Table 4)[14,15,16,17]. This suggests that systematic review authors using the GRADE approach will have to use their

Table 4. Outcome measures recommended by the IMMPACT statement for chronic pain clinical trials[14,15,16,17].

1   Pain
2   Physical functioning (including quality of life)
3   Emotional functioning
4   Participant rating of improvement and satisfaction with treatment
5   Adverse symptoms and adverse events
6   Participant disposition (e.g., adherence to the treatment regime and reasons for premature withdrawal from the trial)
7   Role functioning (i.e., work and educational activities, social and recreational activities, home and family care)
8   Interpersonal functioning (i.e., interpersonal relationships, sexual activities)
9   Sleep and fatigue

IMMPACT: Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials.

judgment to provide no more than 7 outcomes in an SoF table that they believe are of greatest importance to patients.

Research Agenda
Many of the approaches available to convert pain to a binary outcome rely on the continuous data being normally distributed. Future research should explore the distribution of pain outcomes among different clinical conditions to confirm or refute this assumption. We have proposed a number of thresholds to dichotomize continuous pain data, which

reflects the considerable debate in this area, and future research should explore the validity of these thresholds to promote further standardization and consensus. This process should involve input from patients and field clinicians. Other areas for exploration include standardizing the timing of pain data collection (e.g., pain at present, pain in the last 24 h, pain in the last week), and further establishing the relationship between pain reduction and improvement of other patient-important outcomes, such as function and sleep.

REFERENCES

1. Johnston BC, Bandayrel K, Friedrich JO, Akl EA, Da Costa BR, Neumann I, et al. Presentation of continuous outcomes in meta-analysis: a survey of clinicians' understanding and preferences: 21st Cochrane Colloquium. Suppl 1-212. Quebec City, Canada: Cochrane Database Syst Rev 2013;4:097.
2. Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. Am J Med 1992;92:121-4.
3. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? Ann Intern Med 1992;117:916-21.
4. Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. Lancet 1994;343:1209-11.
5. Jaeschke R, Guyatt G, Shannon H, Walter S, Cook D, Heddle N. Basic statistics for clinicians: 3. Assessing the effects of treatment: measures of association. CMAJ 1995;152:351-7.
6. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ 1995;310:452-4.
7. Kristiansen IS, Gyrd-Hansen D, Nexoe J, Nielsen JB. Number needed to treat: easily understood and intuitively meaningful? Theoretical considerations and a randomized trial. J Clin Epidemiol 2002;55:888-92.
8. Nexoe J, Gyrd-Hansen D, Kragstrup J, Kristiansen IS, Nielsen JB. Danish GPs' perception of disease risk and benefit of prevention. Fam Pract 2002;19:3-6.
9. Gigerenzer G. The psychology of good judgment: frequency formats and simple algorithms. Med Decis Making 1996;16:273-80.
10. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. BMJ 2003;327:741-4.
11. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. Acad Med 1998;73:538-40.
12. Woloshin S, Schwartz LM. Communicating data about the benefits and harms of treatment: a randomized trial. Ann Intern Med 2011;155:87-96.
13. Busse JW, Maqbool A, Sivananthan L, Lopes LC, Schandelmaier S, Kamaleldin M, et al. Adherence to IMMPACT outcome reporting recommendations among trials assessing the effect of opioids for chronic non-cancer pain. 22nd Cochrane Colloquium, September 2014, Hyderabad, India.
14. Turk DC, Dworkin RH, Revicki D, Harding G, Burke LB, Cella D, et al. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. Pain 2008;137:276-85.
15. Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J, et al. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. Pain 2006;125:208-15.
16. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9-19.
17. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003;106:337-45.
18. Furlan AD, Sandoval JA, Mailis-Gagnon A, Tunks E. Opioids for chronic noncancer pain: a meta-analysis of effectiveness and side effects. CMAJ 2006;174:1589-94.
19. Chaput de Saintonge DM, Evans S, Crane G, Kirwan JR. Clinical trials and clinical practice: Do doctors use the same criteria to judge outcome? Br J Clin Pharmacol 1997;44:583-6.
20. Schünemann HJ, Guyatt GH. Commentary—goodbye M(C)ID! Hello MID, where do you come from? Health Serv Res 2005;40:593-7.
21. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. BMJ 1998;316:690-3.
22. 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.
23. Kulkarni AV. Distribution-based and anchor-based approaches provided different interpretability estimates for the Hydrocephalus Outcome Questionnaire. J Clin Epidemiol 2006;59:176-184.
24. Johnston BC, Thorlund K, da Costa BR, Furukawa TA, Guyatt GH. New methods can extend the use of minimal important difference units in meta-analyses of continuous outcome measures. J Clin Epidemiol 2012;65:817-26.
25. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting G. Methods to explain the clinical significance of health status measures. Mayo Clin Proc 2002;77:371-83.
26. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res 2011;11:171-84.
27. Kirwan JR. Minimum clinically important difference: the crock of gold at the end of the rainbow? J Rheumatol 2001;28:881-6.
28. Tubach F, Dougados M, Falissard B, Baron G, Logeart I, Ravaud P. Feeling good rather than feeling better matters more to patients. Arthritis Rheum 2006;55:526-30.
29. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105-21.
30. Kelly AM. The minimum clinically significant difference in visual analogue scale pain score does not differ with severity of pain. Emerg Med J 2001;18:205-7.
31. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149-58.
32. Moore RA, Straube S, Aldington D. Pain measures and cut-offs – 'no worse than mild pain' as a simple, universal outcome. Anaesthesia 2013;68:400-12.
33. Furukawa TA, Akechi T, Wagenpfeil S, Leucht S. Relative indices of treatment effect may be constant across different definitions of response in schizophrenia trials. Schizophr Res 2011;126:212-9.
34. Suissa S. Binary methods for continuous outcomes: a parametric alternative. J Clin Epidemiol 1991;44:241-8.
35. Canavero S, Bonicalzi V. Intravenous subhypnotic propofol in central pain: a double-blind, placebo-controlled, crossover study. Clin Neuropharmacol 2004;27:182-6.
36. Lefaucheur JP, Drouot X, Menard-Lefaucheur I, Keravel Y, Nguyen JP. Motor cortex rTMS in chronic neuropathic pain: pain relief is associated with thermal sensory perception improvement. J Neurol Neurosurg Psychiatry 2008;79:1044-9.
37. Lefaucheur JP, Drouot X, Menard-Lefaucheur I, Keravel Y, Nguyen JP. Motor cortex rTMS restores defective intracortical inhibition in chronic neuropathic pain. Neurology 2006;67:1568-74.
38. Lefaucheur JP, Drouot X, Menard-Lefaucheur I, Zerah F, Bendib B, Cesaro P, et al. Neurogenic pain relief by repetitive transcranial

magnetic cortical stimulation depends on the origin and the site of pain. J Neurol Neurosurg Psychiatry 2004;75:612-6.

39. Lefaucheur JP, Drouot X, Nguyen JP. Interventional neurophysiology for pain control: duration of pain relief following repetitive transcranial magnetic stimulation of the motor cortex. Neurophysiol Clin 2001;31:247-52.

40. Macefield RC, Jacobs M, Korfage IJ, Nicklin J, Whistance RN, Brookes ST, et al. Developing core outcomes sets: methods for identifying and including patient-reported outcomes (PROs). Trials 2014;15:49.

41. Busse JW, Ebrahim S, Connell G, Coomes EA, Bruno P, Malik K, et al. Systematic review and network meta-analysis of interventions for fibromyalgia: a protocol. Syst Rev 2013;2:18.

42. Deeks J, Higgins J, Altman D. Analyzing data and undertaking meta-analyses. In: Higgins J, Green S, eds. Cochrane handbook for systematic reviews of interventions version 5.1.0. Chichester, UK: Wiley; 2011.

43. Fern E, Monroe K. Effect-size estimates: issues and problems in their interpretation. J Consum Res 1996;23:89-105.

44. Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH.

Pooling continuous outcomes in meta-analysis: a comprehensive review of methods for enhancing interpretability. Res Synth Methods 2011;2:188-203.

45. Friedrich JO, Adhikari NK, Beyene J. The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: a simulation study. BMC Med Res Methodol 2008;8:32.

46. Furukawa T. From effect size into number needed to treat. Lancet 1999;353:1680.

47. Karanicolas PJ, Smith SE, Kanbur B, Davies E, Guyatt GH. The impact of prophylactic dexamethasone on nausea and vomiting after laparoscopic cholecystectomy: a systematic review and meta-analysis. Ann Surg 2008;248:751-62.

48. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490-7.

49. Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. J Clin Epidemiol 2013;66:173-83.

APPENDIX 1. Outcome measures for reporting pain among trials of therapy for fibromyalgia (n = 241).

1. Pain, 15cm VAS
2. Pain, 100mm/10cm VAS
3. Pain, VAS, 101-point scale
4. Pain, VAS, 11-point scale
5. Pain, VAS, 10-point scale
6. Pain, 10-point scale
7. Pain, 6-point scale
8. Pain; 5-point scale
9. Pain, 4-point scale
10. Pain, VAS, scale not defined
11. Pain, scale not defined
12. Arthritis Self-Efficacy Scale (ASES); Pain subscale
13. Arthritis Impact Measurement Scale-2 (AIMS2); Pain subscale
14. Arthritis Impact Measurement Scale (AIMS); Pain subscale
15. Numeric Rating Scale; 101-point scale (NRS)
16. McGill Pain Questionnaire
17. McGill Pain Questionnaire; Present pain intensity subscale
18. McGill Pain Questionnaire; Sensory pain subscale
19. McGill Pain Questionnaire; Pain rating index
20. McGill Pain Questionnaire-Short Form
21. McGill Pain Questionnaire-Short Form; Present pain intensity subscale
22. McGill Pain Questionnaire-Short Form; Sensory pain subscale
23. Multi-dimension Pain Inventory (MPI)
24. Multi-dimension Pain Inventory; Pain intensity subscale
25. Daily Pain, 21-point scale
26. Daily Pain, 11-point scale
27. Short Form-36; Bodily pain subscale
28. Pain during previous week, 100mm/10cm VAS
29. Pain during previous week, 21-point scale
30. Pain during previous week; 5-point scale
31. Pain during previous week; Global rating, Therapist-rated
32. Overall Pain; 10-point scale
33. Pain at Rest; 100mm/10cm VAS
34. Pain at Rest; 10-point scale
35. Average Pain; 100mm/10cm VAS
36. Highest Pain; 100mm/10cm VAS
37. Pain During Movement; 100mm/10cm VAS
38. Pain During Movement; 10-point scale
39. Regional Pain Scale
40. Comprehensive Psychopathological Rating Scale; Aches and pains
41. Chronic Pain Experience Inventory
42. Present Pain Intensity Rating Scale
43. Health Assessment Questionnaire; Pain Intensity
44. Multidimensional Health Assessment Questionnaire; Pain subscale
45. Pain; Gracely scale, 21-point scale
46. Achiness, VAS (0-100)
47. Muscle Pain, 7-point scale
48. Post Sleep Questionnaire; Pain, 7-point scale
49. Nottingham Health Profile; Pain
50. Muscular Pain
51. Pain Intensity, scale not defined
52. Pain Intensity, composite of 2 items from the McGill Pain Questionnaire
53. Generalized Pain, 10-point scale
54. Specific Pain, 10-point scale
55. CNS Dysfunction Questionnaire; Pain subscale
56. Brief Pain Inventory (BPI); Pain severity, 10-point subscale
57. Lower body pain intensity, 100mm/10cm VAS
58. Upper body pain intensity, 100mm/10cm VAS
59. Pain relief; 6-point scale
60. Pain; Therapist-reported
61. Comprehensive Psychopathological Rating Scale (CPRS); Pain subscale
62. Pain; Composite of 10cm VAS, AIMS and McGill Pain Questionnaire
63. Maastricht Utility Measurement Questionnaire; Pain, 5-point scale
64. Pain intensity in last month, Percentage scale
65. Pain intensity - Morning till breakfast, Percentage scale
66. Pain intensity - Breakfast till lunch, Percentage scale
67. Pain intensity - Lunch till dinner, Percentage scale
68. Pain intensity - Dinner till bedtime, Percentage scale
69. Pain intensity - Before falling asleep, Percentage scale
70. Pain intensity - During everyday activities, Percentage scale
71. Clinician's Pain Score
72. Polyalgia, 4-point scale; Clinician-reported
73. Euroqol-5D (EQ-5D); Pain/discomfort subscale
74. Severity of Pain, Ache and Stiffness, 7-point scale
75. Morning aching (presence of)