
Early version, also known as pre-print

Link to published version (if available): 10.3899/jrheum.141440

Link to publication record in Explore Bristol Research

PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms
Optimal Strategies for Reporting Pain in Clinical Trials and Systematic Reviews: Recommendations from a 2014 OMERACT Workshop

Jason W. Busse 1,2,3
Susan J. Bartlett 4
Maxime Dougados 5
Bradley C. Johnston 3,6,7,8
Gordon H. Guyatt 3
John Kirwan 9
Kent Kwoh 10
Lara J. Maxwell 11
Andrew Moore 12
Jasvinder A. Singh 13
Randall Stevens 14,15
Vibeke Strand 16
Maria E. Suarez-Almazor 17
Peter Tugwell 18-20
George Wells 20

1 The Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Hamilton, Canada
2 Department of Anesthesia, McMaster University, Hamilton, Canada
3 Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada
4 Department of Medicine, McGill University, Divisions of Rheumatology and Clinical Epidemiology, Royal Victoria Hospital, Montreal, Quebec, Canada
5 Paris Descartes University, APHP Cochin Hospital, Rheumatology Department, Cochin Hospital, INSERM (U1153) clinical epidemiology and biostatistics, PRES Sorbonne Paris-Cité, Paris, France
6 The Hospital for Sick Children Research Institute, Toronto, Ontario, Canada
7 Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada
8 Department of Anesthesia and Pain Medicine, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada
9 University of Bristol Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol, UK
10 University of Pittsburgh and VA Pittsburgh Healthcare System, Pittsburgh, Pennsylvania, USA
11 Institute of Population Health, University of Ottawa, Ontario, Canada
12 Pain Research, University of Oxford, Nuffield Division of Anaesthetics, The Churchill, Oxford OX3 7LE
13 Birmingham Veterans Affairs Medical Center and University of Alabama at Birmingham, Birmingham, Alabama, USA
14 Inflammation and Immunology Clinical Research, Celgene Corporation, NJ, USA
15 Department of Medicine, Division of Rheumatology, the State University of New Jersey, NJ, USA
16 Division of Immunology/Rheumatology, Stanford University, Palo Alto California, USA
17 University of Texas MD Anderson Cancer Center, Houston, TX, USA
18 Department of Medicine, Faculty of Medicine, Institute of Population Health, University of Ottawa
19 Ottawa Hospital Research Institute, Clinical Epidemiology Program, Ottawa, Ontario, Canada
20 Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada

**Corresponding Author:** Jason W. Busse, Department of Anesthesia, Michael G. DeGroote School of Medicine, McMaster University, Hamilton, Canada; email: bussejw@mcmaster.ca

**Funding:** This was an unfunded study. JAS is supported by grants from the Agency for Health Quality and Research Center for Education and Research on Therapeutics (AHRQ CERTs) U19 HS021110, National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS) P50 AR060772 and U34 AR062891, National Institute of Aging
(NIA) U01 AG018947, National Cancer Institute (NCI) U10 CA149950, the resources and the use of facilities at the VA Medical Center at Birmingham, Alabama and research contract CE-1304-6631 from the Patient Centered Outcomes Research Institute (PCORI). MSA is the recipient of a K24 award from NIAMS (K24AR053593)

**Competing interests:** Many of the authors have published methodological approaches for interpreting continuous data, and have intellectual conflicts of interest to declare.
Abstract

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. A Working Group at the 2014 Outcome Measures in Rheumatology (OMERACT) was convened to provide guidance regarding how best to report treatment effects regarding pain for individual studies and systematic reviews. We argue that presentation of relative effects will facilitate interpretation of treatment effects and that for individual trials authors should, in addition to mean change, report 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, such as adverse events, function, and sleep, to facilitate interpretation. When pooling data for meta-analysis, authors should consider converting all continuous measures for pain to a 10cm/100mm visual analogue scale (VAS) for pain and use the established minimally important difference (MID) of 1cm/10mm, and the conventionally used appreciably important differences of 2cm/20mm, 3cm/30mm and 5cm/50mm, to facilitate interpretation. Effects ≤0.5 units suggest a small or very small impact. 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.
Background

Randomized controlled trials (RCTs) 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 (Box). Our suggestions are informed by a workshop convened at the 2014 Outcome Measures in Rheumatology (OMERACT) conference.

What effect measures do clinicians find most useful?

Clinicians generally find dichotomous presentation of 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 is presented with the absolute change (risk difference) or NNT.\(^2-4\) 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 as 3 in 100 rather than 3\% or the associated NNT,
33) may be the best way to achieve understanding in a variety of audiences⁹-¹¹ though 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 recent review of RCTs 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 co-morbidity. Trialists are faced with a choice of whether to try and 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 impact of pain, and that exploration of pain characteristics may place a burden on patients that provides little insight into the impact 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**

The effect on pain should be accompanied by presentation of treatment effects on other patient-important outcomes, such as adverse events, function, and sleep, as 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) versus improvement in function (0.31; 95% CI = 0.41 to 0.22).\(^{14}\)

Pain is typically captured 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 or not 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 confidence intervals) 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 10mm on a 100mm visual analogue scale (VAS) for pain.
Such an effect may be statistically significant, but is it important to patients? Even when the mean represents an effect that is important to patients, many clinicians will extrapolate this effect to all patients; however, treatment response will differ between 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 a NNT – the number of patients that need to receive treatment in order 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 minimal important clinical difference (MCID) is also used; however, this terminology focusses on clinician's perceptions versus patient's. 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.

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. Consider an example where the MID is 0.50 and patients mean improvement vs. 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 a 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 \( \frac{1}{2} \) the baseline standard deviation of the instrument score represents the MID.\(^\text{18}\) However, although this represents a moderate effect size there is evidence that anchor-based and distribution-based MIDs may differ.\(^\text{19,20}\) 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\(^\text{21}\) – to produce improvement that allows patients to feel appreciably, not just minimally better\(^\text{22}\). 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.\(^\text{13}\) For this reason, a number of authoritative groups, including many Cochrane groups focussed on pain, have suggested that trialists and review authors consider not the MID for pain (\( \geq 10\% \)
reduction from baseline\textsuperscript{23,24}, but ≥20%, ≥30% or ≥50% reduction from baseline as improvement that is likely to be appreciably important to patients\textsuperscript{25}.

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,\textsuperscript{26} a patient acceptable state\textsuperscript{22}) 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,\textsuperscript{27} and perhaps in most,\textsuperscript{28} the choice of threshold does not affect the magnitude of relative effect.

**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 co-presentation 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\textsuperscript{14}.

*Trials using different outcome measures*
This is the more common scenario that systematic review authors will face. A recent review of RCTs that explored management of fibromyalgia found that eligible studies that captured pain (n=241) reported 75 different measures of this outcome (Appendix 1). Different measures of pain can be pooled by converting to SMDs, which is the approach recommended by the Cochrane Collaboration; however, this measure of effect is difficult to interpret 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).

A second approach is to convert different instruments that measure pain into a single, most familiar instrument and the associated estimate of precision. For example, our review of fibromyalgia found that, among 75 different instruments for reporting pain used, the 10cm/100mm VAS was the most commonly used. There are two statistical approaches to convert multiple instruments to a common measure:

1. multiply standard deviation (SD) units X 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).

2. rescale to units of the most familiar instrument
Both approaches remain vulnerable to challenges with interpretation, and misinterpretation as the mean effect may be mistakenly applied to all patients. For pain, the most familiar instrument is the widely used 10cm/100mm VAS for pain. For this instrument the MID has been established as approximately 1.0cm/10mm,\textsuperscript{23} regardless of pain severity\textsuperscript{24}. Although providing readers with the MID facilitates interpretation, review authors should caution readers against dismissing effects that are less than 1 MID unit and provide guidance for interpreting magnitude of effect. If the MID is 1.0cm/10mm 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 $\geq 2.0$, and one accepts that the estimate of effect is accurate, this suggests a large impact. 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 $\leq 0.5$ units suggest a small or very small impact.

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.\textsuperscript{34} 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 zero, which means this method cannot be used when the control group changes for the worse, and the treatment group for the better.
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 impact.

A fifth and final approach 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 two fundamental statistical approaches to making this calculation. One is to convert the SMD into a proportion that 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. 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 1).

An alternative statistical approach to creating a binary outcome and thus an odds ratio or risk difference – an approach we advocate - avoids 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 that achieved an improvement of the MID or greater in that study. 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 2cm/20mm, 3cm/30mm and 5cm/50mm.25

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 meta-analysis of prophylactic dexamethasone for laparoscopic cholecystectomy that explored the effect on post-operative pain.36 The effect was informed by 5 RCTs that enrolled 539 participants, and certainty in effect estimates was considered 'low' according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria due to inconsistency of results across studies and imprecision associated with pooled estimates of effect. Table 2 presents the effect of dexamethasone on post-operative pain using alternate strategies we have 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 (see 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 10cm/100mm VAS for pain and inform readers of the established MID of 1cm/10mm, and/or the conventionally used appreciably important differences of 2cm/20mm, 3cm/30mm and 5cm/50mm (second
approach above). Authors should also use the thresholds to convert the continuous variable to a binary outcome and present the pooled relative and absolute effects (fifth approach 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. 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. 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 2 demonstrates the presentation of results in SoF tables.

A rule-of-thumb for generating a 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. The Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) has recommended that 9 outcome measures, including pain, should be reported when assessing treatment effects for clinical conditions defined by pain (Table 3). This suggests that systematic review authors using the GRADE approach will have to use their judgement to provide no more than 7 outcomes in a 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. Other areas for exploration include standardizing the timing of pain data collection (e.g. pain at present, pain in the last 24-hours, 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


6. Cook RJ, Sackett DL: The number needed to treat: a clinically useful measure of

understood and intuitively meaningful? Theoretical considerations and a randomized


9. Gigerenzer G. The psychology of good judgment: frequency formats and simple

10. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to


12. Woloshin S, Schwartz LM. Communicating data about the benefits and harms of


16. Schünemann HJ, Guyatt GH. Commentary--goodbye M(C)ID! Hello MID, where do you come from? Health Serv Res. 2005; 40: 593-597.


27. 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(1e3): 212e219.


Box: 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
- Meta-analyses should convert all continuous measures for pain to a 10cm/100mm visual analogue scale (VAS) for pain, report the pooled mean change, and the pooled mean change divided by the minimal (1cm/10mm), appreciable (2cm/20mm and 3cm/30mm), and substantial (5cm/50mm) 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.
Table 1: The relation between effect size or standardized mean difference and the number needed to treat under normality and equal variance assumptions

<table>
<thead>
<tr>
<th>Effect size</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNT</td>
<td>20</td>
<td>13</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
### Table 2: Alternate strategies for presenting the effect on pain

<table>
<thead>
<tr>
<th>Effect on post-operative pain</th>
<th>Estimated risk or estimated score/value with placebo</th>
<th>Absolute reduction in risk or reduction in score/value with Dexamethasone</th>
<th>Relative effect (95% CI)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standardized mean difference (SMD)</td>
<td>The pain score in the dexamethasone groups was on average 0.79 SDs (1.41 to 0.17) lower than in the placebo groups</td>
<td>--</td>
<td>As a rule to thumb, 0.2 SD represents a small difference, 0.5 a moderate, and 0.8 a large</td>
<td></td>
</tr>
<tr>
<td>Natural units (converted from SD units)</td>
<td>Mean scores with placebo ranged from 43 to 54</td>
<td>Mean scores in the dexamethasone groups was on average 8.1 (1.8 to 14.5) lower</td>
<td>--</td>
<td>The MID on the 0 to 100 pain scale is approximately 10</td>
</tr>
<tr>
<td>Dichotomizing using &quot;substantial post-operative pain&quot; vs. other</td>
<td>20 per 100</td>
<td>Difference in proportion achieving an important improvement is 0.15 (0.19 to 0.04)</td>
<td>RR=0.25 (0.05 to 0.75)</td>
<td>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.</td>
</tr>
<tr>
<td>Ratio of means</td>
<td>28.1</td>
<td>3.7 lower pain score (6.1 to 0.6)</td>
<td>ROM=0.87 (0.78 to 0.98)</td>
<td>The ROM is the weighted average of the mean pain score in the treatment groups divided by the mean pain score in placebo</td>
</tr>
<tr>
<td>MID units</td>
<td>The pain score in the dexamethasone groups was on average 0.40 (0.74 to 0.07) MID units less than placebo</td>
<td></td>
<td>An effect of less than half the MID difference suggests a small or very small effect</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Outcome measures recommended by the IMMPACT statement for chronic pain clinical trials

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 & Fatigue
Figure 1: The impact of patient heterogeneity on the standardized mean difference

Effect size when patients are homogeneous: 0.50

Effect size when patients are heterogeneous: 0.25

True MID
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)