Systematic review reveals heterogeneity in definition of a clinically relevant difference in pain.

Adeline Ruyssen-Witrand, Florence Tubach, Philippe Ravaud

Systematic review reveals heterogeneity in definition of a clinically relevant difference in pain

Adeline Ruyssen-Witrand [1] **, Florence Tubach** [1 2] **, Philippe Ravaud** [1 *]

[1] *Modèles et méthodes de l'évaluation thérapeutique des maladies chroniques INSERM : U738 , Université Paris Diderot - Paris 7 , Faculté de médecine Paris 7 16, Rue Henri Huchard 75018 Paris,FR*

[2] *Département d'épidémiologie, biostatistique et recherche clinique Assistance publique - Hôpitaux de Paris (AP-HP) , Hôpital Bichat - Claude Bernard , 46 rue Henri Huchard 75018 Paris,FR*

* Correspondence should be adressed to: Philippe Ravaud <philippe.ravaud@htd.aphp.fr >

Abstract

Objective

To describe the pain decrease considered as clinically relevant when designing a trial and reporting its results.

Methods

A systematic review of the literature in MEDLINE was conducted to select randomised controlled trials (RCTs) with pain as a primary outcome. Data extracted included the definition (terms and values) of a clinically relevant difference in pain, the type of pain studied (acute or chronic), the level of application (group or individual) of the clinically relevant difference, and the reference justifying the choice of value for clinically relevant difference.

Study Design and Setting

74 trials were included and only 16 articles justified the choice of a value for clinically relevant difference with a reference citation. The values chosen for the clinically relevant relative decrease in pain varied from 4 to 40 mm or 15% to 55% at the group level and from 20 to 50 mm at the individual level. In 7 articles, the authors confused the application of the reference value at the individual or group level.

Conclusion

Our review revealed a great heterogeneity in definition, format and values of what is considered a clinically relevant difference in pain in RCTs of analgesics and standardisations are advisable.

MESH Keywords Analgesics ; administration & dosage ; Humans ; Pain ; classification ; drug therapy ; psychology ; Pain Measurement ; standards ; Randomized Controlled Trials as Topic ; Reference Standards

Author Keywords Clinical trials ; clinically relevant ; pain measurement ; data interpretation ; treatment outcome ; group or patient perspective

Pain is a subjective experience, and the wide variation in the experience of pain leads to large variability in ratings of pain by patients undergoing similar interventions [1]. Several scales have been developed to assess pain intensity in daily practice and in randomised controlled trials (RCTs) assessing analgesic therapies. The most widely used scales are the visual analog scale (VAS) and the numeric rating scale (NRS), both of which are sensitive to change [2].

The reporting of a sample size calculation for an RCT has been recommended since the 1990s [3]. At the group level, this calculation relies on demonstrating a clinically relevant difference between groups, if it exists. This clinically relevant difference is used to calculate the sample size and to interpret the RCT results. The choice of the value of this clinically relevant difference is an important part when designing a trial: a large clinically relevant difference will lead to a small sample size but increases the risk of negative results. At the opposite end, a small clinically relevant difference will lead to a very large sample size, is more likely to obtain positive results but this small clinically relevant difference may be not pertinent for a clinician. Much literature addresses what is considered a clinically relevant difference for a patient-reported outcome. The minimal clinically important difference (MCID) is defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in absence of troublesome side effects and excessive cost, a change in patients’ management” [4]. However, definitions of the MCID differ widely. Beaton has proposed a classification of responsiveness in terms of three features or axes [5]. At the individual level, the clinically relevant difference in pain can be assessed by the patient acceptable symptom state (PASS), the highest level of pain beyond which patients consider themselves well [6 – 8].

When analysing the results from trials, comparing continuous data between groups to obtain the absolute mean change in pain intensity offers great statistical power [9 , 10] but is not meaningful for everyone [11]. Presenting results as a percentage of responders after defining a threshold of improvement (MCID) or the threshold of acceptable pain (PASS) at the individual level enhances the relevance of the results [12].

Two international working groups, the IMMPACT [13 –15] and the OMERACT [16] developed guidelines in reporting results of RCT assessing chronic pain and recommended presenting both at the group level with differences in means and at the individual level with proportion of responders.

To investigate these issues in pain assessment, the aim of our study was to review reports of RCTs investigating analgesics for pain to study the difference in pain outcome that trialists have considered clinically relevant when designing such RCTs and reporting results.

Methods

Study design

This study was a systematic review of reports of RCTs assessing analgesics for pain.

Report selection process

A search was conducted using MEDLINE (via PUBMED) to retrieve all reports published from February 19, 2005, to February 19, 2007, of clinical trials with pain as the primary outcome. We used the key word “pain” with limits of “all adults: 19+years”, “Randomized Controlled Trial”, “Humans”, and “Core clinical journals”. The core clinical journal selection limited the search to the 120 journals of immediate interests included in the Abridged Index Medicus (AIM) from MEDLINE.

One of the authors (ARW) read titles and abstracts of retrieved articles. Inclusion criteria were superiority RCTs with 2 parallel arms and pain measured with a VAS or a NRS as the primary outcome. Exclusion criteria were all designs other than 2 parallel arms (such as cross-over design or 3 or more arms), animal studies, paediatric studies, phase I or II trials, or non-therapeutic trials (e.g., metrologic studies and epidemiologic studies).

Definitions

For practical reasons, to avoid confusion between terms, the clinically relevant difference is termed Δ in this paper. In analysed response to treatment, Δ was described according to the three axes of Beaton’s cube [5]: the first axis is the setting, or the individual or group level. Two types of response occur at the individual level: an improvement of more than the Δ or an improvement that leads the patient beyond an acceptable threshold of pain. The second axis is which scores are contrasted: “the difference between”, or absolute difference, is the difference between groups, expressed as the difference in means at the end of the study; “the change within” is the relative change within a group during the study, expressed as a mean (difference between endpoint and baseline within the group) or percentage (percentage of improvement from baseline within the group); “the difference between changes within” is the difference in improvement between groups at the end of the study, expressed as a mean (difference of improvement in means) or percentage (difference of percentage improvement). Finally, the third axis is the type of change (e.g., minimum change actually detectable beyond the error versus minimal clinically important change).

Data extraction

A standardized data collection form was generated on the basis of a review of the literature and finalized by the research team. Before beginning data extraction, one member of the team (AWR) tested the standardized form on a sample of 10 articles from 10 different journals retrieved by the literature search and included in the final sample. Another member of the team reviewed the same articles to discuss difficulties. The same reviewer (AWR) independently completed all the data extraction. The following data were recorded: general information about the trial (i.e. year of publication, pathology, type of pain, aim of the trial…), sample size calculation with value and unit of measurement (millimetre or percentage) of Δ, the term used for the Δ and how Δ was applied (individual or group level), including the axis of the Δ according to Beaton's cube [5]. The expected baseline level of pain in each group was collected from the sample size calculation. If the expected baseline was not available, the final baseline level of pain in the control group or the control group level of pain during the study was collected.

Data analysis

A descriptive analysis of the data extracted from the reports of the RCTs was performed, then assessed whether sample size calculation was reported. The value chosen as the Δ in each trial was described in terms of medical area, type of control group (placebo or other analgesic intervention) and pain type (acute or chronic). The data for values expressed as continuous data (i.e., in millimetres) and as percentages were analysed separately. The values of the Δ were compared to the expected baseline level of pain or if not available, to the

final control group baseline pain or to the control group level of pain during the study. At the group level, all values of the Δ were represented in a graphic plotting the values of the Δ and the (expected) baseline pain, separating values given in means or in percentages.

The use of the Δ in the report was then analysed according to Beaton's cube [5]. The “setting,” meaning individual or group perspective, of the Δ was described. If a reference was cited to justify the choice of the Δ value, the reference was reviewed and the appropriateness of the level for Δ used in the report and in the reference was noted. For RCTs assessing chronic pain, we assessed whether the presentation of results followed the recommendations of IMMPACT [13 –15] and OMERACT [16] in terms of presenting results both at the group level with differences in means and at the individual level with proportion of responders. Whether the scores being contrasted were the same in the sample size calculation and results section was investigated. Finally, the type of change or difference was described: the terms used to define Δ (e.g., minimum Clinically Relevant Difference, Minimum important difference). If a reference was cited to justify the choice of the Δ, the reference was reviewed and the appropriateness of the term and its meaning in the report were noted.

Results

Of the 603 articles retrieved from the MEDLINE search, the selection process lead to include 74 reports of RCTs (figure 1). Characteristics of the trials are summarized in Table 1 . The trials were in the field of anaesthesiology or emergency medicine (n=28), rheumatic diseases (n=19), neurology (n=5), gynaecology or obstetrics (n=9) or surgery (n=13).

A sample size calculation was reported in 59 articles (80%), but in 47, data required for replicating the sample size calculation were incomplete or missing.

Of the 74 reports, 31 (42%) defined the Δ by use of a specific term either in the sample size calculation or the discussion section. Among these, 2 did not report the cut-off value used. Only 12 reports (20%) reported in the sample size calculation the expected baseline pain before starting the experimental treatment.

For reports that described absolute difference at the group level, the values used for defining the Δ varied from 4 to 40 mm for acute pain and from 10 to 20 mm for chronic pain. When addressing differences as percentage change, the Δ varied from 15% to 30% for acute pain and from 30% to 55% for chronic pain. At the individual level, the threshold for improvement ranged from 20 to 50 mm for absolute change and from 20% to 50% for percentage change. The threshold of pain above which a patient considered “painful” ranged from 2 to 40 mm. No influence of the baseline control group expected or reported pain on the value of the Δ was observed in this study (figure 2).

The choice of the value of the delta either expressed on absolute change in mean or difference in percentage change did not depend on the type of pain (acute or chronic, data not shown).. No trend for the choice of the value of the delta was identified depending on the medical area of the RCT (for example, no difference between RCT assessing post-surgery pain and neuropathic pain RCT).

In terms of the first axis of Beaton's cube, the individual or group perspective, 17 reports (23%) described pain results at the individual level and 46 (57%) at the group level; in 15 (20%), results were for both levels. For sample size calculation, the Δ was defined at the individual level more often in rheumatologic trials (42%) than in trials of other medical areas (9%; p=0.01). Only 8 articles investigating chronic pain (30%) gave comparisons both in means and proportion of responders in the results section as recommended by IMMPACT and OMERACT. Only 16 reports (22%) justified the choice of the Δ used in the sample size calculation by citing a reference; in 7 of these, the cut-off value was used inappropriately (table 2) and for inappropriate levels. For example, in one report [17] a 25% to 30% reduction in pain from baseline was expected in the experimental group; to justify such a cut-off, the authors cited a reference [18] in which the threshold of 30% was defined as a clinical improvement at the individual level.

In terms of the second axis of Beaton's cube, which scores were being contrasted, for sample size calculation, the Δ in 30 reports was a “difference between”, or an absolute difference in means between groups; in 12, a “change within”, or a relative change in means or percentage; in 9, a “difference between changes within”; and in 3, was represented by several axes. For the results section, in 62 reports, the Δ was a “difference between”; in 8, a “change within”; in 12, a “difference between changes within”; and in 10, was represented by several axes. In 19 reports (35%), the axis chosen for the Δ in the sample size calculation was not the same as that for the results section. For example, in 6 reports, the authors chose as the Δ a difference of percentage improvement, for the sample size calculation and a difference in means between groups for reporting the results.

In terms of the third axis of Beaton's cube, the type of change or difference, the terms used to define the Δ and the corresponding values are in table 3 . Eighteen different terms were used. Moreover, a given term could correspond to different definitions and to different values. For instance, MCID was defined as a 20mm threshold for improvement of change from baseline at the patient level [19] but also as 17.5 mm mean difference between groups [20]. Confusion of concepts behind the choice of a term was noted. For instance, for sample size calculation, one author [21] used the value of 20 mm as a meaningful difference between groups and justified such a cut-off with a reference [22] that defined the value of 20 mm as the threshold of imprecision at the individual level (table 2).

Discussion

This systematic literature review of reports of RCTs investigating analgesics for pain showed great heterogeneity in the definition of what constitutes a clinically relevant difference in pain and confusion in these definitions that may be misleading in interpreting results. There was a large variation of values used as the Δ in measuring pain, this variation was neither explained by the baseline pain, nor the type of pain, nor the medical area. The systematic review identified 18 different terms to define the Δ. A same term was applied at the group or individual level and the same term could correspond to different values. Consequently, in 7 of 16 reports that justified a value by a reference citation, the trial conclusions were not appropriate because of confusion in application of the value at the group or individual level. In analysing the use of the Δ according to Beaton's cube axes, in 35% of articles, the axis chosen for the Δ in the sample size calculation was not the same as that used to describe the results.

To our knowledge, this is the first study addressing the use and misuse of the Δ in RCTs assessing analgesics for pain. Many reports have tried to define a value for the MCID for pain in specific medical conditions, but consensus is lacking on the value to use. Indeed, the systematic review identified a large range of values used for the Δ both at the individual and group levels.

One important result of this study is that only half of the selected articles reported a sample size calculation while the CONSORT statement [3] had recommended since 1996 to include the sample size calculation in all reports of RCT.

To enhance the relevance of the evaluation of pain in RCTs, the IMMPACT developed guidelines for assessment of chronic pain in RCTs [13 –15] and recommended presenting both absolute changes in pain intensity and proportion of patients showing decrease in pain intensity of at least 30% from baseline. The OMERACT group, an international working group interested in outcome measurements in rheumatology, defined a threshold for reporting the results of change in pain level at the individual level for RCTs [16] As well, the draft document of the US Food and Drug Administration recommended giving results in means and proportions of responders. We found only 8 reports presenting results in terms of both difference in means and proportion of responders, all in the context of chronic pain.

Our study has some limitations. Fewer than half of the reports defined a term for the Δ (n=30) and few reports (n=16) justified the choice of the cut-off value; thus, for most articles, we could not determine whether the conclusions of the study were appropriate. In addition, we studied only the use of the Δ in RCTs with 2 parallel arms. However, what constitutes a variation in clinically relevant pain intensity is not likely to differ across trial designs. We did not find reports in all medical areas. However, we found no difference in the value of the Δ across medical areas, so the values chosen for the Δ in trials from other fields may be close to those identified in our study. Only 12 articles mentioned a hypothetical baseline pain state in both groups when detailing the sample size calculation; thus a sensibility analysis of the variability of the delta depending on the baseline pain was performed using, when data were missing, the observed baseline pain or, if not available, the control group pain during the study. However, the observed pain level could be different to the expected baseline pain that was anticipated by the authors when choosing the value of the Δ.

Nevertheless, our systematic analysis of the literature on RCTs assessing analgesics for pain is a good reflection of the problems in clinical trials assessing pain and in reporting results because we systematically selected all clinical trials for 2 recent years with very large criteria of selection: pain as the primary outcome, pain assessed with a VAS or an NRS, and from all fields and all pathologic conditions whatever the quality of reporting criteria.

This systematic review highlights the need for standardisation in the design and the reporting of results of trials assessing analgesic interventions. We propose some recommendations for defining a clinically relevant difference in clinical trials assessing pain with a VAS or an NRS (table 4). Following Beaton's axes, the authors should first define the perspective of improvement they would like to apply the Δ (group or individual), then they should decide if the Δ is an absolute difference (in mean or percentage) or a change within the same group. Finally, the value of the Δ would be chosen depending on the clinically pertinence of the change the authors would like to demonstrate, differentiating the "smallest detectable change" that is often the standard error of measurement of the pain mean and do not provide good indication of the importance of the observed change and the "clinically important difference" (CID) or "minimum clinically important difference" (MCID) or "Minimum Clinical important improvement" (MCII). A lot of studies proposed values at the individual level for the CID with anchor-based approaches. At the group level, such important CID has not been proposed since it can only be established on the broader context of the disease being treated, the currently available treatment and the overall risk-benefice ratio of the treatment [15].

In conclusion, this study showed that reporting clinical trial results of assessing analgesics for pain is very heterogeneous and highlights the need for standardizing the definition of a clinically relevant difference in pain and cut-off values for such trials.

What is new ?

- Great heterogeneity in the definition of what constitutes a clinically relevant difference in pain and confusion in these definitions may be misleading in interpreting results.

- A large variation of values used as the clinically relevant difference leads to difficulties in interpreting results.
- Standardisations in the choice of the value and the term as the clinically relevant difference are advisable
- The concept of clinically relevant difference in pain used in the sample size calculation should be the same as the primary outcome in the result section.
- In assessment of chronic pain, both absolute difference and proportion of responders should be presented in the results section.

Ackowledgements:

The authors thank Gabriel Baron and Isabelle Boutron for their technical help in this work.

The authors have no financial interest in the subject matter of materials in the manuscript, and the study had no research or project support/funding.

References:

1 . Nielsen CS , Price DD , Vassend O , Stubhaug A , Harris JR . Characterizing individual differences in heat-pain sensitivity . Pain . 2005 ; Dec 15 119 : (1–3) 65 - 74
2 . Holdgate A , Asha S , Craig J , Thompson J . Comparison of a verbal numeric rating scale with the visual analogue scale for the measurement of acute pain . Emerg Med (Fremantle) . 2003 ; Oct-Dec 15 : (5–6) 441 - 6
3 . Begg C , Cho M , Eastwood S , Horton R , Moher D , Olkin I . Improving the quality of reporting of randomized controlled trials. The CONSORT statement . Jama . 1996 ; Aug 28 276 : (8) 637 - 9
4 . Jaeschke R , Singer J , Guyatt GH . Measurement of health status. Ascertaining the minimal clinically important difference . Control Clin Trials . 1989 ; Dec 10 : (4) 407 - 15
5 . Beaton DE , Bombardier C , Katz JN , Wright JG , Wells G , Boers M . Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal Clinically Important Difference . J Rheumatol . 2001 ; Feb 28 : (2) 400 - 5
6 . Kvien TK , Heiberg T , Hagen KB . Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? . Ann Rheum Dis . 2007 ; Nov 66 : (Suppl 3) iii40 - 1
7 . 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 ; Aug 15 55 : (4) 526 - 30
8 . Tubach F , Ravaud P , Baron G , Falissard B , Logeart I , Bellamy N . Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement . Ann Rheum Dis . 2005 ; Jan 64 : (1) 29 - 33
9 . Altman DG , Royston P . The cost of dichotomising continuous variables . Bmj . 2006 ; May 6 332 : (7549) 1080 -
10 . Anderson JJ . Mean changes versus dichotomous definitions of improvement . Stat Methods Med Res . 2007 ; Feb 16 : (1) 7 - 12
11 . Estellat C , Faisy C , Colombet I , Chatellier G , Burnand B , Durieux P . French academic physicians had a poor knowledge of terms used in clinical epidemiology . J Clin Epidemiol . 2006 ; Sep 59 : (9) 1009 - 14
12 . Farrar JT , Dworkin RH , Max MB . Use of the cumulative proportion of responders analysis graph to present pain data over a range of cut-off points: making clinical trial data more understandable . J Pain Symptom Manage . 2006 ; Apr 31 : (4) 369 - 77
13 . Dworkin RH , Turk DC , Farrar JT , Haythornthwaite JA , Jensen MP , Katz NP . Core outcome measures for chronic pain clinical trials: IMMPACT recommendations . Pain . 2005 ; Jan 113 : (1–2) 9 - 19
14 . Turk DC , Dworkin RH , Allen RR , Bellamy N , Brandenburg N , Carr DB . Core outcome domains for chronic pain clinical trials: IMMPACT recommendations . Pain . 2003 ; Dec 106 : (3) 337 - 45
15 . Dworkin RH , Turk DC , Wyrwich KW , Beaton D , Cleeland CS , Farrar JT . Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations . J Pain . 2008 ; Feb 9 : (2) 105 - 21
16 . Tubach F , Ravaud P , Beaton D , Boers M , Bombardier C , Felson DT . Minimal clinically important improvement and patient acceptable symptom state for subjective outcome measures in rheumatic disorders . J Rheumatol . 2007 ; May 34 : (5) 1188 - 93
17 . Qerama E , Fuglsang-Frederiksen A , Kasch H , Bach FW , Jensen TS . A double-blind, controlled study of botulinum toxin A in chronic myofascial pain . Neurology . 2006 ; Jul 25 67 : (2) 241 - 5
18 . 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 ; Nov 94 : (2) 149 - 58
19 . Chae J , Yu DT , Walker ME , Kirsteins A , Elovic EP , Flanagan SR . Intramuscular electrical stimulation for hemiplegic shoulder pain: a 12-month follow-up of a multiple-center, randomized clinical trial . Am J Phys Med Rehabil . 2005 ; Nov 84 : (11) 832 - 42
20 . Hinman RS , Heywood SE , Day AR . Aquatic physical therapy for hip and knee osteoarthritis: results of a single-blind randomized controlled trial . Phys Ther . 2007 ; Jan 87 : (1) 32 - 43 Epub 2006 Dec 1
21 . Rattanachaiyanont M , Leerasiri P , Indhavivadhana S . Effectiveness of intrauterine anesthesia for pain relief during fractional curettage . Obstet Gynecol . 2005 ; Sep 106 : (3) 533 - 9
22 . DeLoach LJ , Higgins MS , Caplan AB , Stiff JL . The visual analog scale in the immediate postoperative period: intrasubject variability and correlation with a numeric scale . Anesth Analg . 1998 ; Jan 86 : (1) 102 - 6
23 . Collins SD , Chessell IP . Emerging therapies for neuropathic pain . Expert Opin Emerg Drugs . 2005 ; Feb 10 : (1) 95 - 108
24 . Lipscomb GH , Roberts KA , Givens VM , Robbins D . A trial that compares Monsel's paste with ball electrode for hemostasis after loop electrosurgical excision procedure . Am J Obstet Gynecol . 2006 ; Jun 194 : (6) 1591 - 4 discussion 5. Epub 2006 Mar 30
25 . Berry JD , Petersen KL . A single dose of gabapentin reduces acute pain and allodynia in patients with herpes zoster . Neurology . 2005 ; Aug 9 65 : (3) 444 - 7
26 . Farrar JT , Berlin JA , Strom BL . Clinically important changes in acute pain outcome measures: a validation study . J Pain Symptom Manage . 2003 ; May 25 : (5) 406 - 11
27 . Suzuki M , Haraguti S , Sugimoto K , Kikutani T , Shimada Y , Sakamoto A . Low-dose intravenous ketamine potentiates epidural analgesia after thoracotomy . Anesthesiology . 2006 ; Jul 105 : (1) 111 - 9
28 . Felson DT , Anderson JJ , Boers M , Bombardier C , Furst D , Goldsmith C . American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis . Arthritis Rheum . 1995 ; Jun 38 : (6) 727 - 35
29 . Wong SM , Hui AC , Tong PY , Poon DW , Yu E , Wong LK . Treatment of lateral epicondylitis with botulinum toxin: a randomized, double-blind, placebo-controlled trial . Ann Intern Med . 2005 ; Dec 6 143 : (11) 793 - 7
30 . McQuay HJ , Barden J , Moore RA . Clinically important changes-what's important and whose change is it anyway? . J Pain Symptom Manage . 2003 ; May 25 : (5) 395 - 6

- 31 . Langford R , McKenna F , Ratcliffe S , Vojtassak J , Richarz U . Transdermal fentanyl for improvement of pain and functioning in osteoarthritis: a randomized, placebo-controlled trial . Arthritis Rheum . 2006 ; Jun 54 : (6) 1829 - 37
- 32 . Edelman A , Nichols MD , Leclair C , Jensen JT . Four percent intrauterine lidocaine infusion for pain management in first-trimester abortions . Obstet Gynecol . 2006 ; Feb 107 : (2 Pt 1) 269 - 75
- 33 . Hui SK , Lee L , Ong C , Yu V , Ho LC . Intrauterine lignocaine as an anaesthetic during endometrial sampling: a randomised double-blind controlled trial . Bjog . 2006 ; Jan 113 : (1) 53 - 7
- 34 . Assis MR , Silva LE , Alves AM , Pessanha AP , Valim V , Feldman D . A randomized controlled trial of deep water running: clinical effectiveness of aquatic exercise to treat fibromyalgia . Arthritis Rheum . 2006 ; Feb 15 55 : (1) 57 - 65
- 35 . Morgan SJ , Jeray KJ , Saliman LH , Miller HJ , Williams AE , Tanner SL . Continuous infusion of local anesthetic at iliac crest bone-graft sites for postoperative pain relief. A randomized, double-blind study . J Bone Joint Surg Am . 2006 ; Dec 88 : (12) 2606 - 12
- 36 . Pendleton J , Costa J , Wludyka P , Carvin DM , Rosser CJ . Combination of oral tramadol, acetaminophen and 1% lidocaine induced periprostatic nerve block for pain control during transrectal ultrasound guided biopsy of the prostate: a prospective, randomized, controlled trial . J Urol . 2006 ; Oct 176 : (4 Pt 1) 1372 - 5
- 37 . Cepeda MS , Carr DB , Miranda N , Diaz A , Silva C , Morales O . Comparison of morphine, ketorolac, and their combination for postoperative pain: results from a large, randomized, double-blind trial . Anesthesiology . 2005 ; Dec 103 : (6) 1225 - 32
- 38 . Ilfeld BM , Morey TE , Thannikary LJ , Wright TW , Enneking FK . Clonidine added to a continuous interscalene ropivacaine perineural infusion to improve postoperative analgesia: a randomized, double-blind, controlled study . Anesth Analg . 2005 ; Apr 100 : (4) 1172 - 8
- 39 . Nikolajsen L , Finnerup NB , Kramp S , Vimtrup AS , Keller J , Jensen TS . A randomized study of the effects of gabapentin on postamputation pain . Anesthesiology . 2006 ; Nov 105 : (5) 1008 - 15
- 40 . Finnerup NB , Biering-Sorensen F , Johannesen IL , Terkelsen AJ , Juhl GI , Kristensen AD . Intravenous lidocaine relieves spinal cord injury pain: a randomized controlled trial . Anesthesiology . 2005 ; May 102 : (5) 1023 - 30
- 41 . van Wijck AJ , Opstelten W , Moons KG , van Essen GA , Stolker RJ , Kalkman CJ . The PINE study of epidural steroids and local anaesthetics to prevent postherpetic neuralgia: a randomised controlled trial . Lancet . 2006 ; Jan 21 367 : (9506) 219 - 24
- 42 . Turan A , Memis D , Karamanlioglu B , Guler T , Pamukcu Z . Intravenous regional anesthesia using lidocaine and magnesium . Anesth Analg . 2005 ; Apr 100 : (4) 1189 - 92
- 43 . Rog DJ , Nurmikko TJ , Friede T , Young CA . Randomized, controlled trial of cannabis-based medicine in central pain in multiple sclerosis . Neurology . 2005 ; Sep 27 65 : (6) 812 - 9
- 44 . Carvalho B , Chu L , Fuller A , Cohen SE , Riley ET . Valdecoxib for postoperative pain management after cesarean delivery: a randomized, double-blind, placebo-controlled study . Anesth Analg . 2006 ; Sep 103 : (3) 664 - 70
- 45 . Ong CK , Lirk P , Tan JM , Sow BW . The analgesic efficacy of intravenous versus oral tramadol for preventing postoperative pain after third molar surgery . J Oral Maxillofac Surg . 2005 ; Aug 63 : (8) 1162 - 8
- 46 . Chang AK , Bijur PE , Meyer RH , Kenny MK , Solorzano C , Gallagher EJ . Safety and efficacy of hydromorphone as an analgesic alternative to morphine in acute pain: a randomized clinical trial . Ann Emerg Med . 2006 ; Aug 48 : (2) 164 - 72 Epub 2006 Apr 27
- 47 . Casey G , Nortcliffe SA , Sharpe P , Buggy DJ . Perioperative nimodipine and postoperative analgesia . Anesth Analg . 2006 ; Feb 102 : (2) 504 - 8
- 48 . Schenk MR , Putzier M , Kugler B , Tohtz S , Voigt K , Schink T . Postoperative analgesia after major spine surgery: patient-controlled epidural analgesia versus patient-controlled intravenous analgesia . Anesth Analg . 2006 ; Nov 103 : (5) 1311 - 7
- 49 . Babcock MS , Foster L , Pasquina P , Jabbari B . Treatment of pain attributed to plantar fasciitis with botulinum toxin a: a short-term, randomized, placebo-controlled, double-blind study . Am J Phys Med Rehabil . 2005 ; Sep 84 : (9) 649 - 54
- 50 . Turan A , White PF , Karamanlioglu B , Pamukcu Z . Premedication with gabapentin: the effect on tourniquet pain and quality of intravenous regional anesthesia . Anesth Analg . 2007 ; Jan 104 : (1) 97 - 101
- 51 . Abbott JA , Jarvis SK , Lyons SD , Thomson A , Vancaille TG . Botulinum toxin type A for chronic pain and pelvic floor spasm in women: a randomized controlled trial . Obstet Gynecol . 2006 ; Oct 108 : (4) 915 - 23

Figure 1

Flow-chart of the selection of articles of randomized controlled trials assessing analgesics for pain

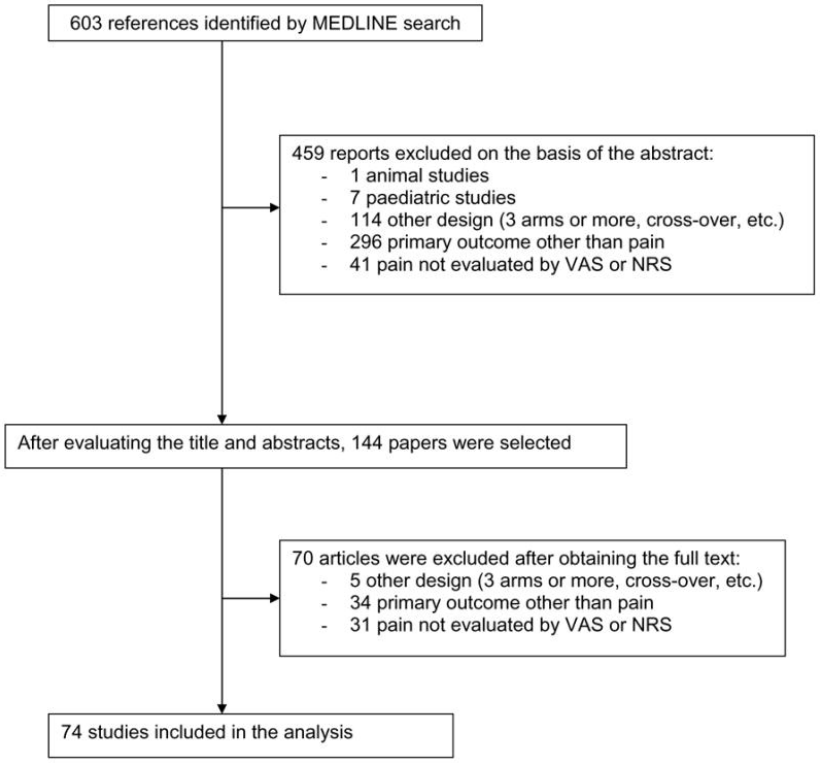
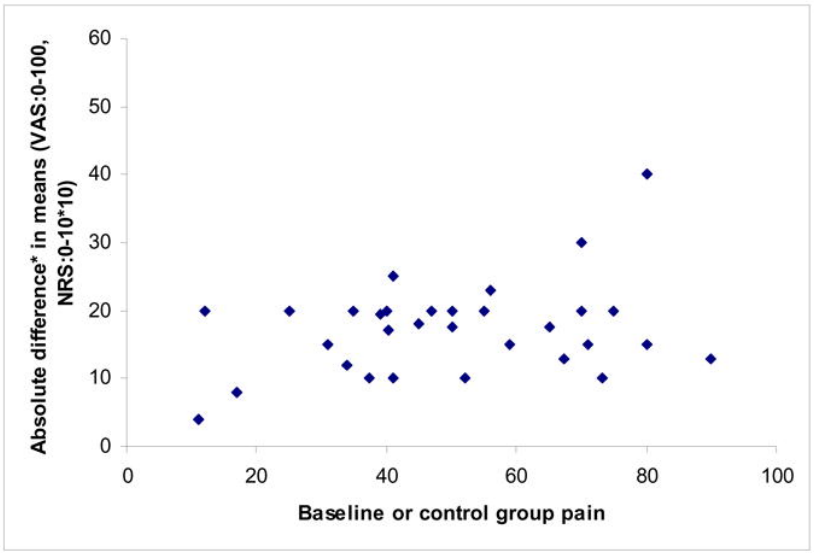


Figure 2

Plots representing the values of the clinically relevant difference in pain used in randomized controlled trials investigating analgesics depending of the expected or observed baseline pain: absolute difference in means (mm) and difference in percentage change (%)

* Absolute difference: $|\mu_{fc} - \mu_{fe}|$ ** Difference in percentage change: $|(\mu_{be} - \mu_{fe})/(\mu_{be}) - (\mu_{bc} - \mu_{fc})/(\mu_{bc})|$ μ_{bc} : Mean baseline score (at final visit) in control group μ_{be} : Mean baseline score (at final visit) in experimental group μ_{fc} : Mean final score (at final visit) in control group μ_{fe} : Mean final score (at final visit) in experimental group VAS=visual analog scale NRS=numeric rating scale



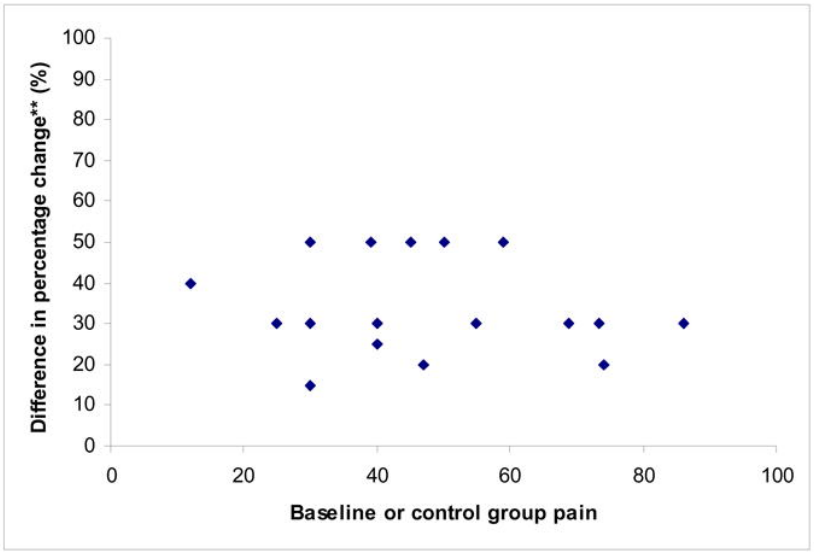


Table 1

Characteristics of randomized controlled trials assessing analgesics for pain

Description	Number	Percentage
Type of trial		
- Pharmacological	47	63.5
- Nonpharmacological	27	36.5
Primary outcome assessment		
- VAS [1]	56	75.7
- NRS [2]	17	23.0
- Both	1	1.35
Type of pain		
- Acute	44	59.5
- Chronic	30	40.5
Control group		
- Placebo	40	54.1
- Other analgesic procedure	29	39.2
- No care	5	6.8
Sample size calculation reported	59	79.7
Sample size calculation data		
- Alpha error not reported	8	13.6
- Power not reported	2	3.4
- Δ[3] not reported	7	11.9
- Control or treatment group assumptions incomplete or not reported[4]	40	67.8
- At least one of the previous	47	79.7

1 Visual Analog Scale

2 Numeric Rating Scale

3 Definition of the clinically relevant difference used in the study

4 Expected mean value in the control or experimental groups or expected success rate in the control or experimental groups

Table 2

Randomized controlled trials assessing analgesics for pain and inappropriate use of clinically relevant difference from references cited to justify values

Reference	Trial
Collins and Chessell [23]: The threshold for acceptable pain was 30 mm at the individual level.	Lipscomb et al. [24]: Difference between groups was considered not clinically relevant because both group means were < 30 mm.
Farrar et al. [18]: The clinical improvement was 20 mm or 30% of improvement at the individual level.	Berry and Petersen [25]: The value of 30% was used at the group level in the sample size calculation.
Farrar et al. [18]: The clinical improvement was 20 mm or 30% of improvement at the individual level.	Qerama et al. [17]: A 25% to 30% reduction from baseline was expected in the experimental group.
Farrar et al. [26]: The clinical improvement was 20 mm or 33% of improvement at the individual level.	Suzuki et al. [27]: The mean improvement in the experimental group was considered too small (< 20 mm) to be relevant.
Felson et al. [28]: An improvement of 20% significant at the individual level.	Wong et al. [29]: The authors concluded that the results for the experimental and placebo groups might be the same because the lower boundary of the confidence interval for the between-groups differences in means was < 20% of improvement.
McQuay et al. [30]: An significant improvement was a 30 % decrease from baseline at the individual level.	Langford et al. [31]: For the sample size calculation, the expected difference between groups at endpoint was an absolute difference in means of 10 mm or a relative change of 38% at the group level. The authors did not specify which value they used in the sample size calculation.
DeLoach et al. [22]: Defined a threshold of imprecision at 20 mm at the individual level	Rattanachaiyanont et al. [21]: Expected difference in means was 20 mm between groups.

Table 3

Terms and values used to define the clinically relevant difference (absolute values are presented) in the 28 randomized controlled trials assessing analgesics for pain.

Study	Terms used for the clinically relevant difference	Value			
		Group level		Individual level	
		Absolute difference[1] (mm)	Difference of percentage of improvement[2] (%)	Threshold of improvement[3] (mm or %)	Threshold of acceptable pain[4] (mm)
(Edelman et al. 2006)	Clinical relevance	15			
(Hui et al. 2006)	Clinical significance		20		
(Assis et al. 2006)	Clinical significance change	20		<10%, 11–20%, 21–30%, >30%	
(Suzuki et al. 2006)	Clinically beneficial effect				10
(Morgan et al. 2006)	Clinically important difference	30			
(Pendleton et al. 2006)	Clinically or substantively significant	10			
(Cepeda et al. 2005)	Clinically meaningful decrease in pain			50%	
(Langford et al. 2006)	Clinically meaningful difference	10	38		
(Ilfeld et al. 2005)	Clinically relevant	40			
(Nikolajsen et al. 2006)	Clinically relevant	20	40		
(Qerama et al. 2006)	Clinically relevant	20	30	30%	
(Finnerup et al. 2005)	Clinically relevant difference	15		33%	
	Clinically relevant pain				30

(van Wijck et al. 2006)					
(Lipscomb et al. 2006)	Clinically significant	4			
(Turan et al. 2005)	Clinically significant		15		
(Rog et al. 2005)	Clinically significant	17.5		50%	
(Carvalho et al. 2006)	Clinically significant difference		30		
(Ong et al. 2005)	Clinically significant difference	20			
(Berry and Petersen 2005)	Effect-size. Clinically significant		30		
(Chang et al. 2006)	Minimal clinically significant difference	13			
(Wong et al. 2005)	Minimum clinically important			20%	
(Hinman et al. 2007)	Minimum clinically important difference	175			
(Chae et al. 2005)	Minimum clinically significant difference			20mm	
(Casey et al. 2006)	Minimum reduction in pain	20			
(Rattanachaiyanont et al. 2005)	Meaningful difference	20			40
(Schenk et al. 2006)	Relevant value for better pain			20 mm	30
(Babcock et al. 2005)	Significant improvement			30%	
(Turan et al. 2007)	Significant reduction		25		

μ_{bc} : Mean baseline score (at the final visit of the trial) in control group
μ_{be} : Mean baseline score (at the final visit of the trial) in experimental group
μ_{fc} : Mean final score (at the final visit of the trial) in control group
μ_{fe} : Mean final score (at the final visit of the trial) in experimental group
1 Absolute difference: $|\mu_{fc} - \mu_{fe}|$
2 Difference in percentage change: $|(\mu_{be} - \mu_{fe})/(\mu_{be}) - (\mu_{bc} - \mu_{fc})/(\mu_{bc})|$
3 Threshold of improvement: definition of a threshold in change above which the patient is considered as improved and is classified as a responder. The threshold can be for an absolute change in millimeters or a relative change in percentage.
4 Threshold of pain: threshold defining an acceptable level of pain above which a patient is considered painful (mm)

Table 4
Recommendations for reporting results of pain from randomised controlled trials

The relevant difference in pain used to calculate the sample size of a trial should be clinically relevant and realistic, and, if possible, should be supported by a reference addressing the same type of pain (acute or chronic) and applied at the same level (group or individual).
The concept of clinically relevant difference in pain (i.e., axis according to Beaton) used for the sample size calculation should be the same as that used to report the main results in the results section.
In the assessment of chronic pain, both absolute difference and proportion of responders (IMMPACT and OMERACT recommendations) should be reported in the results section.