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

Ruyssen-Witrand A., Tubach F, Ravaud P.

Ruyssen-Witrand¹ A., MD, Tubach¹ F, MD, PhD, Ravaud¹ P., MD, PhD

¹ Université Paris 7 Denis Diderot, UFR de Médecine ; INSERM U738 ; AP-HP, Hôpital Bichat, Département d’Épidémiologie, Biostatistique et Recherche Clinique, Paris, France

Number of text pages: 17
Number of tables: 4
Number of figures: 2

Corresponding Author: Pr Ravaud P., Département d’Épidémiologie, de Biostatistiques et de Recherche Clinique, INSERM U738, Hôpital Bichat, 46 rue Henri Huchard, 75018 Paris, France ; tel +33 140257931 ; fax : +33 140256773 ; philippe.ravaud@bch.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.

Key words: Clinical trials, clinically relevant, pain measurement, data interpretation, treatment outcome, group or patient perspective.

Running title: definition of clinically relevant difference in pain assessment

Word count: 197
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.
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 $\Delta$ in this paper. In analysed response to treatment, $\Delta$ 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 $\Delta$ 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 $\Delta$ 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 $\Delta$ 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 $\Delta$ were represented in a graphic plotting the values of the $\Delta$ and the (expected) baseline pain, separating values given in means or in percentages.

The use of the $\Delta$ in the report was then analysed according to Beaton’s cube [5]. The “setting,” meaning individual or group perspective, of the $\Delta$ was described. If a reference was cited to justify the choice of the $\Delta$ value, the reference was reviewed and the appropriateness of the level for $\Delta$ 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 $\Delta$ (e.g., minimum Clinically Relevant Difference, Minimum important difference). If a reference was cited to justify the choice of the $\Delta$, 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 $\Delta$ 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 $\Delta$. 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 $\Delta$ according to Beaton’s cube axes, in 35% of articles, the axis chosen for the $\Delta$ 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 $\Delta$ 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 $\Delta$ 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.

**Acknowledgements**

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


[38] Ilfeld BM, Morey TE, Thanikkary 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.


