Checklist for the preparation and review of pain clinical trial publications: a pain-specific supplement to CONSORT

Jennifer S. Gewandter[a], **James C. Eisenach**[b], **Robert A. Gross**[c], **Mark P. Jensen**[d], **Francis J. Keefe**[e], **David A. Lee**[f], **Dennis C. Turk**[g]

[a]Department of Anesthesiology, University of Rochester, Rochester, NY

[b]Departments of Anesthesiology and Physiology & Pharmacology, Wake Forest, Winston-Salem, NC

[c]Departments of Neurology and Pharmacology & Physiology, University of Rochester, Rochester, NY

[d]Department of Rehabilitation Medicine, University of Washington, Seattle, WA

[e]Department of Psychiatry and Behavioral Science and Department Psychology and Neuroscience, Duke University, Durham NC

[f]Seattle, WA

[g]Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA

Abstract

Introduction—Randomized clinical trials (RCTs) are considered the gold standard when assessing the efficacy of interventions because randomization of treatment assignment minimizes bias in treatment effect estimates. However, if RCTs are not performed with methodological rigor, many opportunities for bias in treatment effect estimates remain. Clear and transparent reporting of RCTs is essential to allow the reader to consider the opportunities for bias when critically evaluating the results. To promote such transparent reporting, the Consolidated Standards of Reporting Trials (CONSORT) group has published a series of recommendations starting in 1996. However, a decade after the publication of the first CONSORT guidelines, systematic reviews of clinical trials in the pain field identified a number of common deficiencies in reporting (e.g., failure to identify primary outcome measures and analyses, indicate clearly the numbers of participants who completed the trial and were included in the analyses, or report harms adequately).

*Corresponding Author: Jennifer S. Gewandter, Ph.D., Tel: +1 585 276 5661, Fax: +1 585 244 7271, jennifer_gewandter@urmc.rochester.edu.

Methods—Qualitative review of a diverse set of published recommendations and systematic reviews that addressed the reporting of clinical trials, including those related to all therapeutic indications (e.g., CONSORT) and those specific to pain clinical trials.

Results—A checklist designed to supplement the content covered in the CONSORT checklist with added details relating to challenges specific to pain trials or found to be poorly reported in recent pain trials was developed.

Conclusions—Authors and reviewers of analgesic RCTs should consult the CONSORT guidelines and this checklist to ensure that the issues most pertinent to pain RCTs are reported with transparency.

1. Introduction

Randomized clinical trials (RCTs) are considered the gold standard when assessing the efficacy of interventions because randomization of treatment assignment minimizes bias in treatment effect estimates. However, depending on the methodological rigor, many opportunities for bias in RCTs remain [26, 37]. The opportunities for such bias should be considered when evaluating and interpreting results of RCTs. This critical evaluation depends on the transparent reporting of clinical trial methods and results in the peer-reviewed literature. To promote such transparent reporting, the Consolidated Standards of Reporting Trials (CONSORT) group has published a series of recommendations starting in 1996. These recommendations cover a wide range of factors including randomization and blinding methods, statistical details, and participant flow, as well as guidance on which details should be covered in various sections of the manuscript [31]. Table 1 outlines the categories covered in CONSORT.

A decade after the publication of the first CONSORT guidelines, systematic reviews of clinical trials in the pain field identified a number of common deficiencies in reporting of clinical trials, including failure to identify primary outcome measures and analyses, indicate clearly the numbers of participants who completed the trial and were included in the analyses, or report harms adequately [10, 17, 18, 20, 21, 23, 41–43]. In this article we describe a checklist (Table 2) designed to supplement the content covered in the CONSORT checklist with added details relating to challenges specific to pain trials or found to be poorly reported in recent pain trials. We have not included areas for which reporting has been found to be poor in pain trials when further expansion of the CONSORT checklist seems unlikely improve reporting (e.g., harms reporting [23, 41]). Although some discussion of various trial design issues as they relate to reporting is inevitable, the purpose of this checklist and accompanying manuscript is not to inform pain trial design. For recommendations regarding study design and outcome measures for various types of pain trials, please see the other articles in this series. We believe that the use of this checklist by authors and reviewers in conjunction with the CONSORT statement [31] will improve the reporting and enhance the interpretability of RCTs of pain treatments.

2. Methods

In preparation for developing this checklist the authors reviewed a diverse set of published recommendations and systematic reviews that addressed the reporting of clinical trials, including those related to all therapeutic indications [e.g., 1,14,15,24,32,36,37,46,47] and those specific to pain clinical trials [e.g., 10,12,13,16–21,23,40–44]. The checklist was modified multiple times based on input from all authors as well as the editors of this series. Examples were developed based on hypothetical protocols and although they may include elements of existing studies are not based largely on any particular example from the literature.

3. Checklist items

In this section, we provide an explanation for our reasoning as to why each item is particularly relevant for RCTs of analgesic treatments. Examples are provided for different types of interventions (i.e., pharmacologic, behavioral, or interventional) or study designs where warranted. The examples are not meant to be inclusive of all possible design features, but rather an example of the types of details that are necessary when reporting.

1. Participants

Clear definitions of eligibility criteria are imperative understanding the study design and for evaluating the generalizability of the study results. A clear definition of eligible participants is important in all trials; however, in pain trials it is particularly important when patients with comorbid pain conditions or mental health conditions are excluded. These patients will certainly be seen clinically for the pain conditions being studied and may not respond as well to treatment or be at higher risk for adverse events in response to treatment with certain pharmacologic classes often used for pain. Additionally, results from baseline screening periods are often used as part of eligibility criteria in pain trials (e.g., requiring response to open-label treatment with the experimental drug in Enriched Enrollment Randomized Withdrawal (EERW) trials or excluding participants who fail to respond to 2 treatments with known efficacy in the condition of interest) [28,30,34] and clear description of such eligibility criteria is imperative.

Example: Eligible patients were at least 18 years of age, had confirmed diabetes (i.e., HbA1C 7) and a diagnosis of diabetic peripheral neuropathy (DPN), confirmed by a neurologist study investigator. They reported having pain associated with diabetes in their lower limbs for at least 3 months prior to the enrollment visit, which occurred 1 week prior to randomization. Patients were ineligible if they had a documented history of major depressive disorder or suicidal thoughts in their medical record or were unwilling to abstain from starting any new pain medications or altering dosages of any current pain medications for the duration of the trial. Patients were not excluded for co-morbid pain conditions unless they caused pain in the lower limbs that may have been difficult to discern from DPN pain in self-report ratings. At the screening visit, participants were given a week-long daily diary that asked them to rate their average pain intensity and their worst pain intensity on 0 – 10 numeric rating scales [0 = *No pain*, 10 = *Worst pain imaginable*]. The following criteria

were required for trial eligibility after the baseline week: (1) a minimum of 4 of 7 diary entries were completed, (2) the mean average pain score for the week was 4 out of 10, and (3) *average* pain ratings were less than or equal to *worst* pain ratings for each day of ratings that were provided.

Example (Same patient eligibility as above example, with the exception of this alternative screening requirement substituted for the one above): All enrolled participants were treated with treatment A for 4 weeks in a run-in period. Treatment A was started at 15mg BID with food for week 1 and increased to 30mg BID for week 2 and 45mg BID for weeks 3 and 4. If participants experienced adverse events while taking 45mg BID that would otherwise cause them to discontinue treatment, they were allowed to revert to 30mg BID. Participants who (1) could tolerate a minimum of 30mg BID for the final 3 weeks of the run-in period, (2) experienced at least a 30% decrease in pain from baseline to the end of the 4 week run-in period and (3) had a mean pain score of 4 out of 10 during the last week of the run-in period were randomized to continue on their maximum tolerated dosage or to placebo treatment.

2. Interventions

2i.* Treatment definition—C*areful reporting of details is necessary when describing behavioral interventions, interventions involving invasive procedures, and pharmacologic interventions involving complicated titration protocols, all of which are commonly evaluated in pain trials. Full understanding of the intervention(s) is required not only for future research that will replicate and extend the RCT findings, but also for translation into clinical practice. Treatment descriptions should include what the intervention consists of, at what schedule and for how long they will receive it, and what modifications to the treatment are allowed within the protocol (if applicable).

Pharmacologic example: Participants were randomly assigned to receive either Treatment or matching placebo for 12 weeks. Participants were given 1 capsule of treatment A (30mg/day) or matching placebo for the first week and 2 capsules (30 mg/day) taken once in the morning for the remaining 11 weeks of the trial. Participants were instructed to take all treatments with food. Participants experiencing adverse events that would otherwise lead to discontinuation were allowed to revert back to 1 capsule per day with approval from the investigator. If participants could not tolerate 1 capsule per day they were withdrawn from the study. Acetaminophen (up to 3g/day) was allowed as rescue therapy if the participants felt it was necessary. Use of all other rescue medications was not permitted. Whether to discontinue the assigned treatment due to adverse events was at the discretion of the participant. Investigators could discontinue treatment for a participant if they felt it was medically necessary in response to an adverse event.

Behavioral example: Participants were randomized to receive the strength training intervention alone or as part of a dyad including a relative or friend. The participants (including partners in the dyad treatment group) attended weekly 45-minute training sessions with a physical therapist at the study site for 12 weeks. In these sessions they performed strength training exercises for the lower extremities including walking, air squats,

stair climbing, and lunges (an Appendix can be provided where a complete description of exercises could be presented). In addition, participants were instructed to perform the same exercises at home 2 times per week either alone or with their partner depending on their group assignment. Participants were required to have the ability to perform a minimum amount of each of these exercises for entry in the study (see manual and eligibility criteria); however, if on a certain day during the study the participant did not feel that they could perform some of the exercises safely, they were allowed to skip the exercise and this was documented.

Procedural example: Participants were randomized to use a temporary spinal cord stimulator for 12 weeks or to the wait-list control group. The spinal cord stimulator (authors should provide manufacturer and model number) was implanted by a neurologist. Participants were positioned prone on the procedure table. The interlaminar space was identified in the midline under fluoroscopy. Landmarks for a paramedian approach were identified under fluoroscopy to address pain in the affected lower limb for each participant. Using a 14-gauge Tuohy needle, the bevel was advanced from the medial aspect of the pedicle after local anesthesia was administered using standardized needles and local anesthetic medication with modifications based on body habitus. The Tuohy needle was advanced in the midline until there was a clear loss of resistance to saline. The leads were placed along the span of the thoracic segments corresponding to the affected segments and lateralization of target symptoms. The stimulator positioning was tested after initiation of the stimulus by discussion with the participant regarding the stimulus coverage. If necessary, the leads were repositioned to optimize coverage. The programming process used a standardized algorithm outlined in the device programming guide and training materials provided by the manufacturer (the authors should provide the detailed algorithm in a supplemental appendix). Once lead positioning was confirmed, the Tuohy needle and style were removed. The lead was secured to the participants' lumbar regions using sterile dressing with Steri-Strips and Tegaderm for the duration of the 12-week trial. A research coordinator reviewed the available pre-set stimulation program options with the participants in the recovery room in different positions (i.e., reclined, sitting, and standing). At this time the participants identified the programs that provided the best coverage and pain relief. The participants were instructed to use the programs that worked best for them and adjust the stimulators to an intensity that they could easily feel but was not uncomfortable. Participants were instructed to turn on the system for at least 1 hour in the morning, 1 hour in the middle of the day, and 1 hour in the evening; however, they were also free to use it as frequently as they would like during the day or night.

2ii. Investigator training

a. Participant interaction to minimize non-specific trial effects on outcomes: Detecting differences in treatment effects using subjective patient-reported symptoms like pain can be complicated by multiple factors. Pain ratings are susceptible to expectations and non-specific effects such as attention received during clinical trial visits. These factors, among others, likely contribute to the large placebo responses that often occur in modern chronic pain trials [45]. In order to demonstrate a difference in the effects between treatments being studied, it is helpful to minimize the non-specific responses in all treatment groups. Training

trial investigators to minimize participant expectations for the experimental treatment by explaining that the efficacy of the treatment is unknown may decrease the placebo response in both treatment groups. In addition, explaining to participants that being as accurate as possible in their pain ratings is important [13].

Example: Investigators and research staff were trained in strategies to minimize the placebo effect when interacting with patients (e.g., managing expectations about treatment efficacy and minimizing excess social interaction). A video training module was used to teach research staff how to deliver instructions to patients in a standardized fashion (*an Appendix can be provided where authors provide the training video*).

b. Treatment integrity: In order to ensure that the treatment was administered in a manner consistent with the treatment manual, investigators should be adequately trained to deliver the intervention and treatment integrity [4] should be assessed and reported. This is particularly important for behavioral interventions which are commonly used for pain.

Example: Ten physical therapists were trained by 4 highly experienced clinical psychologists to deliver the mindfulness meditation treatment portion of the intervention at a 2-day workshop facilitated by the principal investigator (author(s) can provide an Appendix where they give a more detailed description of the course). In brief, the course included didactic presentations to describe the theory underlying the intervention followed by role play demonstration for treatment delivery and then practice sessions in which pairs of physical therapists practiced delivering the treatment to one another with observation and feedback from the workshop facilitators. Additionally, each physical therapist delivered the intervention to study participants at least 2 times in the presence of an experienced clinical psychologist who monitored the delivery for fidelity to the treatment manual and provided feedback. Further rounds of observation were utilized if deemed necessary by the clinical psychologist. In addition, all of the interventions were audio recorded and ongoing supervision was provided by a clinical psychologist based on these audio recordings. Finally, a random selection of 20% of the audio tapes were reviewed by two clinical psychologists not involved in the study and coded to assess both treatment integrity and therapist competence. Half of the recordings were coded by both psychologists; these were compared to assess reliability of coding (Kappa coefficient = .82, indicating a high degree of reliability). Coding items included those for inclusion of essential content (i.e., treatment integrity, e.g., teaching and encouraging incorporating mindfulness practices in everyday situations for the mindfulness condition) and therapist competence (i.e., delivering treatment components in a skillful and responsive way, e.g., using appropriate language and examples with a patient with low health literacy).

3. Outcomes

Outcome measures in pain trials are often self-report measures of pain intensity or related domains (e.g., physical function, mood, sleep) [12]. Many factors exist that can affect the way in which participants interpret the 0 – 10 pain intensity numeric rating scale (NRS). For example, the common anchor of "worst pain imaginable" for the 10 rating is likely interpreted variably by different participants and to our knowledge few instances occur in the

published literature where researchers provide participants with any direction as to how to interpret this anchor [40]. Furthermore, participants are often asked to rate their "average pain" over the past day. Yet they are not provided any instructions regarding how to derive their rating (e.g., should participants with highly fluctuating pain consider periods without any pain in their "average" pain score? Should they include their pain during sleep in their estimate?) [11]. Finally, patients often consider their pain interference with function and affective components of pain when completing their NRS ratings of pain intensity. Enhanced instructions to focus on pain intensity independent of mood and pain interference with activities could minimize the inclusion of these related constructs in NRS pain intensity ratings [11,40]. Currently, little research is available regarding the optimal instructions for participants pertaining to these details of the NRS and certainly no consensus exists; however, clear reporting of instructions provided to participants in RCTs will provide data upon which to base future research. Of note, clear reporting of pain intensity measures has been deficient in recent clinical trials of pain treatments [42].

Example: The pre-specified primary outcome measure was a 0 – 10 numeric rating scale (NRS) [0 = *No pain*, 10 = *Worst pain imaginable*] for average pain over the last 24 hours. Research staff administered training to the participants on completing the pain diary. In brief, participants were asked to (1) complete pain ratings on their own before bedtime; (2) focus on the pain in their legs and feet throughout the entire day considering the intensities felt during different activities when determining their average pain; and (3) avoid considering pain from other sources such as a headache when rating their pain (*Note: author(s) can refer to an Appendix here where they provide the complete training manual*). Pre-specified secondary outcome measures included the pain interference question from the Brief Pain Inventory – Short form [BPI-SF] [7] and the Western Ontario & McMaster Universities Osteoarthritis Index (WOMAC) [2]. The BPI interference question asks patients to circle the number that best describes how, during the past 24 hours, pain has interfered with the following symptoms on a 0 to 10 NRS [0 = does not interfere, 10 = completely interferes]: general activity, mood, walking ability, normal work, relations with other people, sleep, and enjoyment of life. The WOMAC is a self-report scale that has items that fall within 3 domains: pain, stiffness, and physical function. It asks patients to rate their difficult with each item on a 0 – 4 [0 = None, 1 = Slight, 2 = Moderate, 3 = Very, and 4 = Extremely].

4. Blinding

Double-blinding can be challenging in many pain trials because pharmacologic pain treatments often have recognizable side-effects and full blinding of investigators or patients can be impossible with certain behavioral or procedural treatments. Clear reporting of efforts made to maximize blinding and to control for effects not related to the active treatment in behavioral trials (e.g., attention received during study visits) allows the reader to evaluate the methods used in the trial.

Example: This trial compared a physical therapy intervention group to an educational information comparison group. (Note: the active intervention would be described here, see Section 2i for examples). The educational information comparison group received

informational packets outlining similar exercises to the ones performed with the physical therapist in the active treatment group. Participants in this group met with a study therapist to discuss their progress for the same amount of time at the same frequency with which the participants in the active group met with the physical therapist. The study participants were blind to the research hypotheses. They were told that it was unknown whether receipt of an educational packet providing exercise instructions or visits to a physical therapist to perform those exercises was more effective. After the study, the participants were informed of the real study hypothesis and consent to use their data was obtained. Participants were asked to rate their expectations for the outcome of their treatment condition after treatment assignment, but before their first treatment session in order to examine whether expectations between the 2 groups were similar. In addition, the research coordinator administering the outcome measures was blinded with respect to the treatment assignment and the participants were asked not to discuss their study activities with her.

5. Statistical methods

Pre-specification of the primary analysis including identification of the measure(s), time point (if applicable), description of the statistical model and statistical test, groups to be compared, methods for handling multiplicity, methods for accommodating missing data, and sample to be used (e.g., all randomized participants vs. only those that completed the trial according to the protocol) is necessary to enhance trial credibility and minimize the probability of a type I error [20, 24, 44]. Multiple outcomes are often important in pain conditions. For example, investigators may prioritize improvement of pain intensity and physical function equally in an osteoarthritis trial or improvements in pain intensity and fatigue in a trial of fibromyalgia. Furthermore, it is often of interest to evaluate the effects of a treatment on acute and chronic pain or compare more than two treatments, both of which also may lead to multiple analyses of equal importance. If more than one analysis is declared primary, pre-specified methods should be used to adjust for multiplicity, especially in later phase trials that are designed to evaluate efficacy [20]. These methods should be clearly reported. For example, when authors state that there are co-primary analyses, it should be reported whether the protocol specified that the trial would be concluded a success if *both* analyses yielded a result with $p < 0.05$ or if *either* analysis yielded a result with $p < 0.025$. A recent systematic review found deficiencies in identification of primary analyses and methods to adjust for multiplicity in pain trials [20].

In most RCTs, some participants discontinue before the end of the study leading to missing data. Others might remain enrolled in the study, but might not provide some data at some assessment points for other reasons (e.g., missed visit). As a result, missing data is a common problem in many chronic pain trials [18, 27]. A large amount of missing data can lead to bias in treatment effect estimates. Using a statistician-recommended strategy to accommodate missing data, rather than excluding participants whose data are missing, can minimize bias. Such strategies include using multiple, pre-specified methods to accommodate missing data that make different assumptions about the patterns of missingness [29,33]. Methods to accommodate missing data were reported in fewer than half of pain trials reviewed in a recent systematic review [18].

Example: The co-primary outcome measures were the average pain NRS and the Fibromyalgia Impact Questionnaire score [6] measured at 12 weeks after randomization. For each outcome measure, cognitive-behavior therapy was compared to education control using an ANCOVA model that included treatment group as the factor of interest and the corresponding baseline symptom score as a covariate. The primary analyses included all available data from all randomized participants. Missing data were accommodated using the technique of multiple imputation. The imputation procedure for each outcome variable utilized treatment group and outcomes at all time points, along with Markov chain Monte Carlo simulation, to produce 20 complete data sets. These data sets were analyzed separately using ANCOVA and the results combined across data sets using Rubin's rules [29]. A Bonferroni correction was used to adjust for multiplicity; a p-value less than 0.025 for either analysis was considered significant to preserve the overall significance level at 0.05. A secondary analysis compared the percentage of "responders" between groups using a chi-square test. A "responder" was defined as a participant whose NRS pain scores at 12 weeks (1) decreased by at least 30% from baseline and (2) was below 4 out of 10. Participants who prematurely discontinued were defined as non-responders. Pre-specified analyses of secondary outcome variables used similar ANCOVA models to those of the primary analyses. No adjustment for multiplicity was made in the secondary analyses as they were considered exploratory and hypothesis generating.

6. Participant flow

As clearly outlined in the CONSORT guidelines [32], it is imperative for the evaluation of any trial data that the number of participants who were randomized, completed, and whose data were included in the analyses as well as reasons for drop out are outlined for each group. To enhance understanding of generalizability of the trial, we also recommend reporting the numbers of participants who were screened prior to randomization and the reasons for exclusion.

Example: In total, 650 patients were screened for study enrollment; 200 did not meet initial eligibility criteria (authors can refer to the CONSORT diagram for reasons). Another 30 participants were eliminated after the baseline week for the following reasons: participant's mean pain score was < 4 (n = 18), participant did not complete at least 4 diary entries (n = 7), and participant did not return for the randomization visit (n = 5). See CONSORT guidelines for an example of the remainder of the participant flow reporting (items 13a and b [32].)

7. Limitations

It is important to interpret the results of RCTs appropriately based on the statistical analyses performed and overall context of the trial. Overall conclusions of efficacy should be based on the primary between-treatment comparisons. Discussions pertaining to potential efficacy based on changes from baseline in each treatment group in the absence of statistically significant between-treatment differences are discouraged, but if included must be accompanied by an acknowledgement that such analyses do not reflect the level of evidence provided by a RCT and that such effects are possibly due to placebo and other non-specific

effects as well as regression to the mean [5, 19]. It is important to outline the limitations of secondary and subgroup analyses that support a treatment effect in the absence of support by the primary analysis. The degree to which interpretation of these analyses is limited depends on whether there were a limited number of pre-specified secondary and sub-group analyses as compared to many post-hoc analyses and whether attempts were made to control the probability of a type I error in the secondary analyses [24]. It is important to note that post-hoc analyses are valuable for hypothesis generation and, when of interest, should be included in RCT reports with the appropriate caveats. For studies yielding non-significant results, the confidence interval for the treatment effect should be considered when determining whether the data are consistent with the absence of a clinically meaningful treatment effect or are inconclusive [8, 22, 39]. Although confidence intervals can be used to evaluate the possibility that the data support comparable efficacy of two interventions, formal prospective non-inferiority or equivalence studies are necessary to confirm the result [47]. Poor treatment integrity or adherence to the protocol or the absence (or compromise) of blinding can lead to biased treatment effect estimates. A clear discussion of potential effects of low treatment integrity, adherence, or blinding in conjunction with appropriate interpretation of the statistical analyses will provide a balanced interpretation for readers.

Example: The estimated treatment effect from the primary efficacy analysis that compared pain severity between the Treatment A and placebo groups was not significant. However, the confidence interval for the treatment effect included a difference of 3 points on the pain NRS in favor of Treatment A, suggesting that the results of this study cannot rule out a potentially clinically meaningful effect for Treatment A. Additionally, nominally significant differences (i.e., $p < 0.05$) between groups were obtained in secondary analyses comparing the treatment groups with respect to measures of pain interference with function and sleep using items from the BPI interference question. Although these secondary analyses cannot be considered confirmatory, these results in combination with the inconclusive result from the primary analysis suggest that further research may be warranted to determine whether Treatment A is effective for chronic low back pain.

Example: This study failed to demonstrate a significant difference between Treatments A and B with respect to mean pain NRS score in patients with chronic low back pain. The 95% confidence interval for the group difference excluded differences larger than that 1.0 NRS points in favor of either treatment, suggesting that there is no clinically meaningful difference between the treatments. It should be acknowledged, however, that currently no consensus exists for the minimal clinically meaningful between-group treatment difference in pain scores. Additionally, a study with a *pre-specified* hypothesis designed to evaluate the equivalence of the two treatments is required to confirm this conclusion. The endpoint pain score was significantly lower than baseline pain score in both groups, suggesting the possibility of some benefit for each treatment. However, the absence of a placebo group makes it impossible to determine whether the apparent effects of either treatment are due only to placebo effects (e.g., effects from expectation, or increased attention received during a clinical trial), natural history, or regression to the mean.

Example 3: The overall (average) fidelity to the treatment protocol among the study clinicians was 94% (range 85% – 100%), suggesting that the social workers delivered the intervention as it was intended. Thus, deviations from the protocol did not appear to contribute to the lack of treatment efficacy observed for the pain coping skills treatment in this study.

Conclusions

In order to maximize readers' ability to critically evaluate the results and conclusions drawn based on RCTs, it is imperative that authors clearly report the methods and results of those RCTs and carefully interpret those results within the limits of the designs and analyses of the trials. Authors and reviewers of analgesic RCTs should consult the CONSORT guidelines and this checklist to ensure that the issues most pertinent to pain trials are reported with transparency. Although these recommendations are focused on reporting of RCTs, reviewers and readers can also use the information presented here to evaluate the quality of the design and the validity of the results when reading manuscripts reporting the findings from RCTs.

References

1. Altman D, Bland M. Absence of evidence is not evidence of absence. BMJ. 1995; 311:485. [PubMed: 7647644]

2. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988; 15:1833–40. [PubMed: 3068365]

3. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov. 2006; 5:27–36. [PubMed: 16485344]

4. Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. J Public Health Dent. 2011; 71:S52–S63.

5. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA. 2010; 303:2058–64. [PubMed: 20501928]

6. Buckhardt CS, Clark SR, Bennet RM. The Fibromyalgia Impact Questionnaire: development and validation. J Rheumatol. 1991; 18:728–33. [PubMed: 1865419]

7. Cleeland, C. [Accessed 2/15/2017] Brief Pain Inventory. 1991. [http://www.mdanderson.org/education-and-research/departments-programs-and-labs/departments-and-divisions/symptom-research/symptom-assessment-tools/brief-pain-inventory.html]

8. Detsky AS, Sackett DL. When was a "negative" clinical trial big enough? How many patients you needed depends on what you found. Arch Intern Med. 1985; 145:709–12. [PubMed: 3985731]

9. Dragalin V. An introduction to adaptive designs and adaptation in CNS trials. Eur Neuropsychopharmacol. 2011; 21:153–8. [PubMed: 20888739]

10. Dworkin JD, Mckeown A, Farrar JT, Gilron I, Hunsinger M, Kerns RD, McDermott MP, Rappaport BA, Turk DC, Dworkin RH, Gewandter JS. Deficiencies in reporting of statistical methodology in recent randomized trials of nonpharmacologic pain treatments: ACTTION systematic review. J Clin Epidemiol. 2016; 72:56–65. [PubMed: 26597977]

11. Dworkin RH, Burke LB, Gewandter JS, Smith SM. Reliability is necessary but far from sufficient. How might the validity of pain ratings be improved? CJP. 2015; 31:599–602.

12. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, Kerns RD, Stucki G, Allen RR, Bellamy N, Carr DB, Chandler J, Cowan P, Dionne R, Galer BS, Hertz S, Jadad AR, Kramer LD, Manning DC, Martin S, McCormick CG, McDermott MP, McGrath P, Quessy S, Rappaport BA, Robbins W, Robinson JP, Rothman M, Royal MA, Simon L, Stauffer JW, Stein W,

Tollett J, Wernicke J, Witter J. Immpact Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain. 2005; 113:9–19. [PubMed: 15621359]

13. Dworkin RH, Turk DC, Peirce-Sandner S, Burke LB, Farrar JT, Gilron I, Jensen MP, Katz NP, Raja SN, Rappaport BA, Rowbotham MC, Backonja MM, Baron R, Bellamy N, Bhagwagar Z, Costello A, Cowan P, Fang WC, Hertz S, Jay GW, Junor R, Kerns RD, Kerwin R, Kopecky EA, Lissin D, Malamut R, Markman JD, McDermott MP, Munera C, Porter L, Rauschkolb C, Rice AS, Sampaio C, Skljarevski V, Sommerville K, Stacey BR, Steigerwald I, Tobias J, Trentacosti AM, Wasan AD, Wells GA, Williams J, Witter J, Ziegler D. Considerations for improving assay sensitivity in chronic pain clinical trials: IMMPACT recommendations. Pain. 2012; 153:1148–58. [PubMed: 22494920]

14. European Medicines Agency. Guideline on choice of non-inferiority margin. Stat Med. 2006; 25:1628–1638. [PubMed: 16639773]

15. Gallo P, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, Pinheiro J. Adaptive designs in clinical drug development--an Executive Summary of the PhRMA Working Group. J Biopharm Stat. 2006; 16:275–83. [PubMed: 16724485]

16. Gewandter JS, Dworkin RH, Turk DC, McDermott MP, Baron R, Gastonguay MR, Gilron I, Katz NP, Mehta C, Raja SN, Senn S, Taylor C, Cowan P, Desjardins P, Dimitrova R, Dionne R, Farrar JT, Hewitt DJ, Iyengar S, Jay GW, Kalso E, Kerns RD, Leff R, Leong M, Petersen KL, Ravina BM, Rauschkolb C, Rice AS, Rowbotham MC, Sampaio C, Sindrup SH, Stauffer JW, Steigerwald I, Stewart J, Tobias J, Treede RD, Wallace M, White RE. Research designs for proof-of-concept chronic pain clinical trials: IMMPACT recommendations. Pain. 2014; 155:1683–95. [PubMed: 24865794]

17. Gewandter JS, McDermott MP, Mckeown A, Smith SM, Pawlowski JR, Poli JJ, Rothstein D, Williams MR, Bujanover S, Farrar JT, Gilron I, Katz NP, Rowbotham MC, Turk DC, Dworkin RH. Reporting of intention-to-treat analyses in recent analgesic clinical trials: ACTTION systematic review and recommendations. Pain. 2014; 155:2714–9. [PubMed: 25284072]

18. Gewandter JS, McDermott MP, Mckeown A, Smith SM, Williams MR, Hunsinger M, Farrar J, Turk DC, Dworkin RH. Reporting of missing data and methods used to accommodate them in recent analgesic clinical trials: ACTTION systematic review and recommendations. Pain. 2014; 155:1871–7. [PubMed: 24993384]

19. Gewandter JS, Mckeown A, McDermott MP, Dworkin JD, Smith SM, Gross RA, Hunsinger M, Lin AH, Rappaport BA, Rice AS, Rowbotham MC, Williams MR, Turk DC, Dworkin RH. Data interpretation in analgesic clinical trials with statistically nonsignificant primary analyses: an ACTTION systematic review. J Pain. 2015; 16:3–10. [PubMed: 25451621]

20. Gewandter JS, Smith SM, Mckeown A, Burke LB, Hertz SH, Hunsinger M, Katz NP, Lin AH, McDermott MP, Rappaport BA, Williams MR, Turk DC, Dworkin RH. Reporting of primary analyses and multiplicity adjustment in recent analgesic clinical trials: ACTTION systematic review and recommendations. Pain. 2014; 155:461–6. [PubMed: 24275257]

21. Gewandter JS, Smith SM, Mckeown A, Edwards K, Narula A, Pawlowski JR, Rothstein D, Desjardins PJ, Dworkin SF, Gross RA, Ohrbach R, Rappaport BA, Sessle BJ, Turk DC, Dworkin RH. Reporting of adverse events and statistical details of efficacy estimates in randomized clinical trials of pain in temporomandibular disorders: Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks systematic review. J Am Dent Assoc. 2015; 146:246–54. e6. [PubMed: 25819656]

22. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. Ann Intern Med. 1994; 121:200–6. [PubMed: 8017747]

23. Hunsinger M, Smith SM, Rothstein D, Mckeown A, Parkhurst M, Hertz S, Katz NP, Lin AH, McDermott MP, Rappaport BA, Turk DC, Dworkin RH. Adverse event reporting in nonpharmacologic, noninterventional pain clinical trials: ACTTION systematic review. Pain. 2014

24. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005; 2:e124. [PubMed: 16060722]

25. Ioannidis JP, Evans SJ, Gøtzsche PC, O'neill RT, Altman DG, Schulz K, Moher D. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med. 2004; 141:781–8. [PubMed: 15545678]

26. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ. 2001; 323:42–6. [PubMed: 11440947]

27. Katz N. Methodological issues in clinical trials of opioids for chronic pain. Neurology. 2005; 65:S32–49. [PubMed: 16385103]

28. Katz N. Enriched enrollment randomized withdrawal trial designs of analgesics: focus on methodology. Clin J Pain. 2009; 25:797–807. [PubMed: 19851161]

29. Little, RJA, Rubin, DB. Statistical Analyses with Missing Data. John Wiley & Sons, Inc; New York: 1987.

30. Mcquay HJ, Derry S, Moore RA, Poulain P, Legout V. Enriched enrolment with randomised withdrawal (EERW): Time for a new look at clinical trial design in chronic pain. Pain. 2008; 135:217–20. [PubMed: 18258369]

31. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010; 340:c869. [PubMed: 20332511]

32. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. Ann Intern Med. 2001; 134:657–62. [PubMed: 11304106]

33. Molenbergs, G, Kenward, MG. Missing Data in Clincal Studies. Chichester, UK: 2007.

34. Moore RA, Wiffen PJ, Eccleston C, Derry S, Baron R, Bell RF, Furlan AD, Gilron I, Haroutounian S, Katz NP, Lipman AG, Morley S, Peloso PM, Quessy SN, Seers K, Strassels SA, Straube S. Systematic review of enriched enrolment, randomised withdrawal trial designs in chronic pain: a new framework for design and reporting. Pain. 2015; 156:1382–95. [PubMed: 25985142]

35. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979; 35:549–556. [PubMed: 497341]

36. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA. 2012; 308:2594–604. [PubMed: 23268518]

37. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995; 273:408–12. [PubMed: 7823387]

38. Senn SJ. Turning a blind eye: authors have blinkered view of blinding. BMJ. 2004; 328:1135–6.

39. Simon R. Confidence intervals for reporting results of clinical trials. Ann Intern Med. 1986; 105:429–35. [PubMed: 3740683]

40. Smith SM, Amtmann D, Askew RL, Gewandter JS, Hunsinger M, Jensen MP, McDermott MP, Patel KV, Williams M, Bacci ED, Burke LB, Chambers CT, Cooper SA, Cowan P, Desjardins P, Etropolski M, Farrar JT, Gilron I, Huang IZ, Katz M, Kerns RD, Kopecky EA, Rappaport BA, Resnick M, Strand V, Vanhove GF, Veasley C, Versavel M, Wasan AD, Turk DC, Dworkin RH. Pain intensity rating training: results from an exploratory study of the ACTTION PROTECCT system. Pain. 2016; 157:1056–64. [PubMed: 27058680]

41. Smith SM, Chang RD, Pereira A, Shah N, Gilron I, Katz NP, Lin AH, McDermott MP, Rappaport BA, Rowbotham MC, Sampaio C, Turk DC, Dworkin RH. Adherence to CONSORT harms-reporting recommendations in publications of recent analgesic clinical trials: an ACTTION systematic review. Pain. 2012; 153:2415–21. [PubMed: 22985899]

42. Smith SM, Hunsinger M, Mckeown A, Parkhurst M, Allen R, Kopko S, Lu Y, Wilson HD, Burke LB, Desjardins P, McDermott MP, Rappaport BA, Turk DC, Dworkin RH. Quality of pain intensity assessment reporting: ACTTION systematic review and recommendations. J Pain. 2015; 16:299–305. [PubMed: 25637296]

43. Smith SM, Wang AT, Katz NP, McDermott MP, Burke LB, Coplan P, Gilron I, Hertz SH, Lin AH, Rappaport BA, Rowbotham MC, Sampaio C, Sweeney M, Turk DC, Dworkin RH. Adverse event assessment, analysis, and reporting in recent published analgesic clinical trials: ACTTION systematic review and recommendations. Pain. 2013; 154:997–1008. [PubMed: 23602344]

44. Turk DC, Dworkin RH, Mcdermott MP, Bellamy N, Burke LB, Chandler JM, Cleeland CS, Cowan P, Dimitrova R, Farrar JT, Hertz S, Heyse JF, Iyengar S, Jadad AR, Jay GW, Jermano JA, Katz NP, Manning DC, Martin S, Max MB, Mcgrath P, Mcquay HJ, Quessy S, Rappaport BA, Revicki DA,

Rothman M, Stauffer JW, Svensson O, White RE, Witter J. Analyzing multiple endpoints in clinical trials of pain treatments: IMMPACT recommendations. Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials. Pain. 2008; 139:485–93. [PubMed: 18706763]

45. Tuttle AH, Tohyama S, Ramsay T, Kimmelman J, Schweinhardt P, Bennett GJ, Mogil JS. Increasing placebo responses over time in U.S. clinical trials of neuropathic pain. Pain. 2015; 156:2616–26. [PubMed: 26307858]

46. U.S. Health and Human Services. [Accessed 11-22-2016] Adaptive design clinical trials for drugs and biologicals: draft guidance. 2010. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm201790.pdf

47. U.S. Health and Human Services. [Accessed 11-22-2016] Guidance for industry: non-inferiority trials. 2010. http://www.fda.gov/downloads/Drugs/.../Guidances/UCM202140.pdf

Table 1

CONSORT headings

CONSORT checklist categories
Title and Abstract
Introduction
Background and Objectives
Methods
Trial design
Participants
Interventions
Outcomes
Sample size
Randomization
Blinding
Statistical Methods
Results
Participant flow
Recruitment
Baseline data
Numbers analyzed
Outcomes and estimation
Ancillary analyses
Harms
Discussion
Limitations
Generalizability
Interpretation
Other information
Registration
Protocol
Funding

Note: bolded categories are expanded on in the pain-specific supplement checklist.

Table 2

Pain-specific CONSORT supplement checklist

Pain checklist supplement	
Methods	
1. Participants	
2. Intervention	

Treatment definition	*Pharmacologic trials*
	    Treatment dosage, frequency, time of day of administration including relationship to food intake, and titration protocols, including allowances for dosage reduction, criteria for and frequency of rescue dose provision
	Behavioral trials
	    Type of intervention, format (e.g., group, dyad, individual), number, frequency and duration of sessions
	    Individual administering the intervention (e.g., psychologist, physical therapist, self-administered)
	    Location (e.g., outpatient clinic, work, home)
	Procedural trials
	    Type of intervention (e.g., x-ray vs. MRI guided nerve block)
	    Manufacturer of instruments
	    Administrator (e.g., nurse anesthetist)

Investigator training	Details of training protocols for investigators to manage non-specific trial effects on outcomes Methods to maximize treatment integrity, including use of a treatment manual (if applicable, include manual in an appendix) for behavioral and procedural trials
3. Outcomes	Pre-specified primary outcome measure (including type of pain measure (e.g., NRS or VAS), characteristics of pain (e.g., average, worst), time frame of measure, additional instructions provided (e.g., location of pain)) Secondary outcome measures (indicate if pre-specified or not) Any participant training in regards to responding to included patient-reported outcome measures
4. Blinding	Who, if anyone, was blinded (e.g., participants, all investigators, outcome assessors) and what they were blinded to (e.g., treatment assignment, study hypotheses) Efforts made to enhance blinding (e.g., active placebo treatments) Efforts made to maximize the similarities between the active and control study procedures in behavioral and procedural trials, including efforts made to elicit similar outcome expectancies Attempts made to blind investigators to eligibility criteria
5. Statistical methods	Primary analysis (including the time point (if applicable), statistical test(s), groups to be compared, and sample of participants. For a "responder" analysis, provide a clear operational definition of "responder" If multiple primary analyses, methods used to adjust for multiplicity or a statement that no adjustment was made with reasoning Adjustments made for multiplicity in secondary analyses, if any Methods used to accommodate missing data and their underlying assumptions
Results	
6. Participant flow	Numbers screened and summary of major reasons for screen failure and refusal to participate
Discussion	
7. Limitations	Base overall conclusions of efficacy on the primary analysis (should be a between-treatment comparison for RCTs) Acknowledge the limitations of secondary or subgroup analyses that support the treatment effect Acknowledge the extent of missing data or disparities between groups in withdrawal rates as limitations of the trial For studies with non-significant treatment effects, use the CIs to evaluate whether the data are inconclusive or consistent with no treatment effect (for placebo controlled trial) or similar treatment efficacy (trial of 2 active treatments) Acknowledge the limitation of concluding similar treatment efficacy from studies that were not designed to test this hypothesis Discuss assessments of treatment fidelity, adherence, and blinding[*] and how they affect the interpretation of the study results

[*] Only blinding assessments that investigate the primary reason for the treatment guess should be reported and discussed because if participants guess their treatment assignment correctly due primarily to efficacy, then investigators could wrongly conclude that unblinding led to bias in the study results [30].