HHS Public Access
Author manuscript
Pain. Author manuscript; available in PMC 2021 November 01.

Published in final edited form as:
Pain. 2020 November ; 161(11): 2446–2461. doi:10.1097/j.pain.0000000000001952.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Disclosures

The views expressed in this article are those of the authors, some of whom were, or currently are, employees of pharmaceutical, consulting, or contract research companies and may have financial conflicts of interest related to the issues discussed in this article. At the time of the meeting on which this article is based, several authors were employed by pharmaceutical companies and others had received consulting fees or honoraria from one or more pharmaceutical or device companies. Authors of this article who attended the IMMPACT meeting and were not employed by industry or government at the time of the meeting received travel stipends, hotel accommodations, and meals during the meeting provided by ACTTION. ACTTION has received research contracts, grants, or other revenue from the FDA, multiple pharmaceutical and device companies, philanthropy, and other sources. Preparation of background literature reviews and the article was supported by ACTTION. No funding from any other source was received for the meeting, nor for the literature reviews and article preparation. No official endorsement by the FDA, US National Institutes of Health, or the pharmaceutical and device companies that have provided unrestricted grants to support the activities of ACTTION should be inferred. SMS has received in the past 36 months a research grant from the Richard W. and Mae Stone Goode Foundation. RHD has received in the past 5 years research grants and contracts from the US Food and Drug Administration and the US National Institutes of Health, and compensation for serving on advisory boards or consulting on clinical trial methods from Abide, Acadia, Adynxx, Analgesic Solutions, Aptinyx, Aquinox, Asahi Kasei, Astellas, AstraZeneca, Biogen, Biohaven, Boston Scientific, Braeburn, Celgene, Centrexion, Chromocell, Clexio, Concert, Coronado, Daiichi Sankyo, Decibel, Dong-A, Editas, Eli Lilly, Eupraxia, Glenmark, Grace, Hope, Hydra, Immune, Johnson & Johnson, Lotus Clinical Research, Mainstay, Medavante, Merck, Neumentum, Neurana, NeuroBo, Novaremed, Novartis, NSGene, Olatec, Periphagen, Pfizer, Phosphagenics, Quark, Reckitt Benckiser, Regenacy (also equity), Relmada, Sanifit, Scilex, Semnur, Sollis, Spinifex, Syntrix, Teva, Thar, Theranexus, Trevena, Vertex, and Vizuri. In the past 36 months DCT has received research grants and contracts from US Food and Drug Administration and US National Institutes of Health, and compensation for consulting on clinical trial and patient preferences from AccelRx, Eli Lilly, Flexion, GlaxoSmithKline, and Pfizer. MPM has been supported in the past 36 months by research grants from NIH, FDA, NYSTEM, SMA Foundation, Cure SMA, Friedreich's Ataxia Reseach Alliance, Muscular Dystrophy Association, ALS Association, and PTC Therapeutics, has received compensation for consulting from Neuropore Therapies, Inc. and Voyager Therapeutics, and has served on Data and Safety Monitoring Boards (DSMBs) for NIH, Novartis Pharmaceuticals Corporation, AstraZeneca, Eli Lilly and Company, aTyr Pharma, Inc., Catabasis Pharmaceuticals, Inc., Vaccinex, Inc., Cynapsus Therapeutics, Voyager Therapeutics, and Prilenia Therapeutics Development, Ltd. CE does not have any financial conflicts of interest specifically related to the issues discussed in this article. JTF has received research grants and contracts from US Food and Drug Administration, and National Institutes of Health, and consulting fees from Analgesic Solutions, Aptinyx, Biogen, Campbell Alliance, Daiichi Sankyo, DepoMed, Evadera, Jansen, Lilly, Novartis, Vertex, and Pfizer; DSMB services from NIH-NIA and Cara Therapeutics. MCR reports participation in scientific advisory boards with compensation for Site One Therapeutics, and with CODA Biotherapeutics. ZB was a salaried employee of Alexion Pharmaceuticals at the time of the meeting. LBB was a salaried employee of the U.S. Food and Drug Administration at the time of the meeting, and in the past 36 months, she has received compensation for consulting on clinical trial outcomes from Aclaris, Afyx, Alnylam, Bellerophon, Biogen, Brickell, Jazz, Leo, Plethora, Scynexis, Therapeutics MD, and Zynerba. LG does not have any financial conflicts of interest specifically related to the issues discussed in this article. SSE reports service on data monitoring committees for AstraZeneca, Merck, BMS, GeneOne, and Marinus; consulting for Johnson & Johnson, Shionogi, Biomarin, Alkermes, InsMed, Innovative Science Solutions, AbbVie, PTCBio, and Amgen; and an invited lecture for Sanofi. SRE reports compensation from Takeda / Millennium, Pfizer, Roche, Novartis, Achaogen, Huntington's Study Group, ACTTION, Genentech, Amgen, GSK, American Statistical Association, FDA, Osaka University, Nationa Cerebral and Cardiovascular Center of Japan, NIH, Society for Clinical Trials, Statistical Communications in Infectious Diseases (DeGruyter), AstraZeneca, Teva, Austrian Breast & Colorectal Cancer Study Group (ABCSG)/Breast International Group (BIG) and the Alliance Foundation Trials (AFT), Zeiss, Dexcom, American Society for Microbiology, Taylor and Francis, Claret Medical, Vir, Arrevus, Five Prime, Shire, Alexion, Gilead, Spark, Clinical Trials Transformation Initiative, Nuvelution, Tracon, Deming Conference, Antimicrobial Resistance and Stewardship Conference, World Antimicrobial Congress, WAVE, Advantagene, Braeburn, Cardinal Health, Lipocine, and Microbiotix, and grants from NIAID/NIH outside the submitted work. SI was a salaried employee of Eli Lilly and Company at the time of this meeting. MPJ has received in the past 36 months research grants from US the US National Institutes of Health, the Department of Education, the Administration of Community Living, the Patient-Centered Outcomes Institute, and National Multiple Sclerosis Society, the International Association for the Study of Pain, and the Washington State Spinal Injury Consortium, and compensation for consulting from Goalistics. RJ is a salaried employee of Pfizer and holds Pfizer stock. CK does not have any financial conflicts of interest specifically related to the issues discussed in this article. NPK is an employee of Analgesic Solutions, a research and consulting firm with numerous clients in the pharmaceutical and medical device industry. JPK was a salaried employee of Acadia Pharmaceuticals at the time of the meeting. EAK was a salaried employee of Collegium with stock at the time of the meeting. DL is a salaried employee of Scilex Pharmaceuticals, formerly known as Semnur Pharmaceuticals. JDM has had advisory board or consulting agreements with the following entities over the past decade as follows: advisory board for Clexio Biosciences, Esteve Pharmaceuticals, Flexion Therapeutics, Quark Pharmaceuticals, Quartet Medicine, Collegium Pharmaceutical, Biogen, Novartis, Aptinyx, Nektar, Allergan, Grünenthal, Eli Lilly and Company, Depomed, Janssen Pharmaceuticals, Teva, KemPharm, Abbott Laboratories, Plasma Surgical, Chromocell, Convergence Pharmaceuticals, Inspirion, Pfizer, Daiichi Sankyo, and Trevena; has served as a consultant to Trigemina, Editas Medicine, and Plasma Surgical; and has served on data safety monitoring boards for Novartis and Allergan. PJM has received research grants, consulting fees, and speaker fees from Abbvie, Amgen, Boehringer Ingelheim, BristolMyersSquibb, Celgene, Eli Lilly, Galapagos, Gilead, GlaxoSmithKline, Janssen, Merck, Novartis, Pfizer, SUN Pharma, and UCB. KVP has received in the past 36 months research grants and contracts from the US Food and Drug Administration and US National Institutes of Health. SNR has received research grants from US National Institutes of Health and Medtronic Inc., and has served as a consultant for Aptinyx,

Interpretation of chronic pain clinical trial outcomes: IMMPACT recommended considerations

Shannon M. Smith[a,b,c], **Robert H. Dworkin**[a,c,d,e], **Dennis C. Turk**[f], **Michael McDermott**[d,e,g], **Christopher Eccleston**[h], **John T. Farrar**[l,j], **Michael C. Rowbotham**[k], **Zubin Bhangwagar**[l,m], **Laurie B. Burke**[n,o], **Penney Cowan**[p], **Susan S. Ellenberg**[j], **Scott R. Evans**[q], **Roy L. Freeman**[r], **Louis P. Garrison**[s], **Smriti Iyengar**[t], **Alejandro Jadad**[u], **Mark P. Jensen**[v], **Roderick Junor**[w], **Cornelia Kamp**[e,x], **Nathaniel P. Katz**[y,z], **J. Patrick Kesslak**[aa], **Ernest A. Kopecky**[ab], **Dmitri Lissin**[ac], **John D. Markman**[ad], **Philip J. Mease**[ae,af], **Alec B. O'Connor**[ag], **Kushang V. Patel**[f], **Srinivasa N. Raja**[ah], **Cristina Sampaio**[ai,aj], **David Schoenfeld**[ak], **Jasvinder Singh**[al], **Ilona Steigerwald**[am], **Vibeke Strand**[an], **Leslie A. Tive**[ao], **Jeffrey Tobias**[ap], **Ajay D. Wasan**[aq], **Hilary D. Wilson**[ar]

[a]Department of Anesthesiology and Perioperative Medicine, University of Rochester Medical Center, Rochester, NY, USA

[b]Department of Obstetrics and Gynecology, University of Rochester Medical Center, Rochester, NY, USA

[c]Department of Psychiatry, University of Rochester Medical Center, Rochester, NY, USA

[d]Department of Neurology, University of Rochester Medical Center, Rochester, NY, USA

[e]Center for Health and Technology, University of Rochester Medical Center, Rochester, NY, USA

[f]Anesthesiology & Pain Medicine, University of Washington, Seattle, WA, USA

[g]Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY, USA

[h]Centre for Pain Research, The University of Bath, Bath, United Kingdom

[i]Departments of Epidemiology, Neurology, and Anesthesia, University of Pennsylvania, Philadelphia, PA, USA

[j]Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA, USA

[k]CPMC Research Institute, Sutter Health, San Francisco, CA, USA

[l]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA

[m]Rallybio, New Haven, CT, USA

[n]School of Pharmacy, University of Maryland, Baltimore, MD, USA

Allergan, Grunenthal, and Insys Therpaeutics. IS was a salaried employee of Gruenenthal Gmbh at the time of the meeting. LT is an employee of Pfizer and holds Pfizer stock. JT was a salaried employee of NeurogesX, Inc. at the time of the meeting and is a partner and managing director of the Aquila Consulting Group LLC. ADW in the past 36 months has received an investigator initiated grant from Collegium Pharmaceuticals, and has consulted for Pfizer and Analgesic Solutions. HDW was a salaried employee of the University of Washington at the time of the meeting, is currently a salaried employee of the pharmaceutical company Boehringer Ingelheim, and within the past 36 months was previously a salaried employee of a scientific consulting firm Evidera.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[o]LORA Group, LLC, Royal Oak, MD, USA

[p]American Chronic Pain Association, Rocklin, CA, USA

[q]Department of Epidemiology and Biostatistics, The George Washington University, Washington, DC, USA

[r]Department of Neurology, Harvard Medical School, Boston, MA, USA

[s]Department of Pharmacy, University of Washington, Seattle, WA, USA

[t]Eli Lilly and Company, Indianapolis, IN, USA

[u]Dalla Lana School of Public Health and Department of Anesthesia, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[v]Department of Rehabilitation Medicine, University of Washington, Seattle, WA, USA

[w]Eisai Ltd, Hatfield, United Kingdom

[x]Clinical Materials Services Unit, University of Rochester Medical Center, Rochester, NY, USA

[y]Tufts University School of Medicine, Boston, MA, USA

[z]Analgesic Solutions, Natick, MA, USA

[aa]Revance Therapeutics, Newark, CA, USA

[ab]Collegium Pharmaceutical, Inc., Stoughton, MA, USA

[ac]Scilex Pharmaceuticals, Palo Alto, CA, USA

[ad]Neuromedicine Pain Management and Translational Pain Research, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA

[ae]Rheumatology Clinical Research, Swedish Medical Center, Seattle, WA, USA

[af]University of Washington School of Medicine, Seattle, WA, USA

[ag]Department of Medicine, University of Rochester Medical Center, Rochester, NY, USA

[ah]Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[ai]Faculdade Medicinda de Lisboa, Universidade de Lisboa, Lisboa, Portugal

[aj]CHDI Foundation, Princeton, NJ, USA

[ak]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[al]Departments of Medicine and Epidemiology, University of Alabama at Birmingham School of Medicine, Birmingham, AB, USA

[am]Neumentum, Inc., Palo Alto, CA, USA

[an]Division of Immunology/Rheumatology, Stanford University, Palo Alto, CA, USA

[ao]Pfizer Inc, New York, NY, USA

[ap]Aquila Consulting Group, LLC, Petaluma, CA, USA

[aq]Departments of Anesthesiology and Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[ar]Boehringer Ingelheim, Seattle, WA, USA

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

1. Introduction

Incorporating a treatment for chronic pain into clinical practice requires critical evaluation of the treatment's effects based largely on the data submitted to support the approved product labeling, supplemented by publications in the literature. In turn, evaluating the evidence for a particular treatment requires understanding the trial designs (e.g., explanatory trials examining whether a treatment has an effect under carefully controlled conditions, and pragmatic trials examining whether a treatment has an effect in a general clinical population) [100] and methods for presenting data regarding treatment efficacy and safety [69]. When trial designs and reporting methods differ between trials, it may complicate the interpretation of a treatment's clinical effect and the ability to compare results between different treatments for chronic pain and across diverse populations. Under the auspices of the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (ACTTION; http://www.acttion.org/) public-private partnership with the U.S. Food and Drug Administration (FDA), the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT; http://www.immpact.org/) convened a meeting in 2011 to review methods of analyzing and reporting the results of randomized clinical trials (RCTs) of pain treatments and the limitations of each approach. The focus of this meeting was to develop recommendations for improving the understanding and interpretation of RCTs of pain treatments by clinicians and other stakeholders who have limited methodologic or statistical expertise. Topics were selected to be consistent with research that has identified causes of clinical misinterpretations of RCTs, such as understandings of trial summary statistics and defining a clinically meaningful difference [2, 79, 98, 123]. Additional topics were considered during the drafting of this article [13, 66, 69, 70]. This article begins by highlighting explanatory and pragmatic approaches to research. Frequently-used methods for describing the benefits and risks of treatments for chronic pain are then considered, including their advantages and limitations. This is followed by a brief discussion of a selection of methods to generate an integrated summary of a treatment's benefit-risk profile.

2. Methods

IMMPACT organized a meeting to discuss and reach consensus on approaches that facilitate interpretation of analgesic RCTs. International representatives from academia, regulatory and other governmental agencies, industry, and a pain patient advocacy group participated in this meeting, and were included as authors on this manuscript. ACTTION's policy for IMMPACT meetings is to invite all members of the ACTTION Executive and Steering Committees. ACTTION strives to include on these committees individuals from across the world representing diverse stakeholders and with expertise or involvement in clinical trial methods. This list of invitees is supplemented by inviting individuals with particular expertise in the topics to be discussed at the specific IMMPACT meeting. Background

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

lectures were presented by co-authors of this manuscript to facilitate discussion. Topics included: (1) what clinicians want to learn from the results of analgesic RCTs (MCR), (2) responder analyses, cumulative distribution functions, and other approaches to enhancing clinicians' interpretation (JTF), (3) meta-analyses, numbers needed to treat (NNT), and Cochrane systematic reviews and other synthesized evidence regarding healthcare interventions (CE), and (4) interpreting responder analyses and NNTs (available on the IMMPACT website, http://www.immpact.org/meetings/Immpact14/participants14.html). During the meeting, the content to include in the manuscript and the advantages and limitations of various methods used to present RCT results were considered. Multiple revisions to preliminary drafts of this article were made until consensus was achieved among all authors.

3. Explanatory vs. pragmatic trials

When interpreting pain treatment trial results, it is important to consider not only the specific findings and how they are presented by the study authors and sponsors, but also to determine whether the research is designed to answer an explanatory question (i.e., is the treatment efficacious within a carefully controlled study?) or pragmatic question (i.e., is the treatment effective under real world conditions?); see Table 1 for terms and definitions [42, 100, 111]. To answer an explanatory question about a treatment's efficacy and safety (i.e., what causal inferences can be drawn about the treatment's analgesic effect and adverse event profile), the "gold standard" research design is an adequately powered double-blind, placebo-controlled, RCT [24, 53, 92]. RCTs of pain treatments are internally valid to the extent that they are rigorously controlled to minimize effects on outcomes that are not caused by the study treatment. Participants should be selected to maximize the probability of observing a treatment effect if one exists (e.g., including enough patients to have adequate power to detect a clinically meaningful difference between the treatment and control groups, minimizing concomitant pain treatments, comorbid pain conditions, rescue medication, and variability in the pain experience). Identifying a treatment effect if one exists can be accomplished by having strict eligibility criteria, keeping patients and study staff blinded to the protocol (e.g., randomization criteria, treatment group assignments, study start point, analysis time point for the primary efficacy endpoint) when possible, limiting concurrent treatments, performing power calculations with the best available estimates of required parameters (e.g., standard deviation of the outcome variable), measuring outcomes that are meaningful and interpretable in the population studied, and randomizing patients to ensure unbiased assignment to treatment. Explanatory trials are conducted to establish a treatment's efficacy and safety under "ideal" conditions that may not necessarily generalize to the wider population of individuals with a particular chronic pain condition who are treated in clinical practice [39, 41].

Once a treatment has been shown to be efficacious and safe for a specific chronic pain condition within the setting of carefully controlled clinical trials, it is also important to determine whether the treatment is effective in the broader population of individuals with that condition in clinical practice; in other words, evaluating external validity. Evaluating the relative effectiveness compared to other standard treatments may help establish the clinical application of the treatment. Pragmatic trials are designed to evaluate the treatment in a more

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

heterogeneous sample, as may be the case within a clinical patient population in which effect modifiers (e.g., comorbid pain conditions, various concurrent treatments, variability in the pain experience) are more likely to play a role [39, 41]. Such trials may yield a more generalizable estimate of the treatment effect and therefore may be more externally valid, but can be less informative with respect to understanding the potential impact of treatment due to the variability permitted in trial conduct. Identifying factors that might alter the treatment effect in pragmatic trials may enhance the ability to understand the clinical utility of a treatment; however, such investigations are often limited by insufficient sample sizes.

Explanatory and pragmatic trials each contribute distinct information to the evidence available regarding pain treatments. Explanatory trials address whether treatments have health benefits or risks, whereas pragmatic trials explore the bounds within which beneficial analgesic outcomes can be observed in study populations. Establishing whether treatments are efficacious and the broader circumstances under which they work are each important for clinical decision-making [100]. The tradeoffs between precision and external validity in explanatory versus pragmatic trials must be considered when interpreting the results of these two complementary classes of trials. In some cases, study designs that blend these two approaches may accelerate the speed with which pain treatments can be translated into clinical practice [97, 120].

4. Determining treatment benefit

4.1. Hypothesis testing, confidence intervals, and P-values

To adequately evaluate the evidence of efficacy provided by a statistically significant treatment group difference in the primary analysis, several issues must be addressed. Foremost among these are whether the analysis that tested the primary hypothesis of the trial was pre-specified (i.e., decided upon before analyzing the trial data), and whether the problem of multiplicity (i.e., performing multiple analyses), if applicable, was addressed in a satisfactory manner to prevent inflation of the probability of a type I error (see Table 1 for definitions of terms). Additional important considerations include whether there were any flaws or potential for bias in study planning, design, execution, and analysis of the trial data and whether the results for important secondary outcomes were consistent with the primary analysis [60, 94]. It is also important to evaluate the extent to which the results of the RCT suggest that the treatment provides a clinically important benefit, as discussed below.

When the primary analysis of an RCT indicates that the difference between treatment groups is not statistically significant, one possibility is that there was a type II error, that is, the treatment is truly efficacious, but the RCT failed to identify the treatment benefit. This can happen for a variety of reasons, including a sample size too small to detect a treatment effect of minimal clinical importance, or problematic study design or execution that resulted in poor quality data and inadequate assay sensitivity [25]. However, assuming that there was sound study design and execution [93], it is important to decide whether the results should be interpreted as the absence of a clinically meaningful treatment effect or whether they should be considered inconclusive. A useful aid to making this decision is to assess the confidence interval (CI) for the treatment effect (see Figure 1). Specifically, if the CI for the treatment effect does not contain values that would be considered clinically meaningful

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

treatment effects, then the trial results can be interpreted as evidence for the absence of a clinically meaningful treatment effect. Example 1 in Figure 1 is a trial in which the primary analysis did not yield a statistically significant result and the estimated magnitude of the treatment effect (i.e., the range of values that fall within the upper and lower CIs) was not clinically meaningful. In contrast, Example 5 shows a trial in which the treatment effect was found to be non-zero according to the hypothesis test, but the estimated magnitude of the treatment effect was not clinically meaningful [44]. If the CI includes values of the treatment effect considered clinically meaningful, however, the results of a trial in which the treatment effect was not statistically significant would be considered inconclusive (see Examples 2 and 3 in Figure 1). In Example 4, there is evidence for the absence of a beneficial treatment effect, but the evidence regarding superiority of the control is inconclusive. Of course, such results may provide the basis for further study to examine the treatment's hypothesized efficacy [48]. Although for many years biostatisticians have recommended this approach to interpreting results of clinical trials in which the treatment effect was not statistically significant, a recent systematic review found that proper interpretation of CIs occurs infrequently in general medical journals [44].

The importance of examining CIs when interpreting pain RCTs can be seen in trials in which there is a “negative” result. For example, pregabalin is recommended for treatment of painful diabetic peripheral neuropathy (pDPN) by international treatment guidelines and is approved for treatment of neuropathic pain in both the US and Europe [7]. However, certain trials of individuals with pDPN have demonstrated statistically non-significant separation between pregabalin and placebo on pain outcomes, which are interpreted as indicating that the trial is negative. For example, a study conducted by Raskin and colleagues found a treatment effect (pregabalin – placebo) of −0.32 (95% CI, −0.74 to 0.09) on the primary outcome variable (i.e., change in pain intensity) [96]. Based on this result, the authors concluded that the study was negative, citing “the negative primary analysis” in the Discussion. However, consideration of the CI in this study suggests that the trial was inconclusive, rather than negative, given that the CI included results consistent with what could be considered a clinically meaningful decrease in pain intensity for pDPN (i.e., > 0.50) associated with pregabalin [23].

Systematic reviews of RCTs in the general medical literature [12] and for pharmacologic and invasive treatments for pain [45] have also shown that erroneous or misleading interpretations of treatment effects that are not statistically significant are quite common. For example, authors often suggest that two interventions are equivalent when an RCT fails to show that one treatment is superior to another. This is not an appropriate conclusion when a trial has been designed to test superiority rather than equivalence, as is the case for most RCTs of pain treatments. Conclusions of equivalence require that the treatment effect fall within prespecified margins of equivalence that are clinically justified [78, 95]. Additionally, common examples of misleading “spin” in the interpretation of RCTs with disappointing results in the primary analyses include emphasizing statistically significant results of secondary analyses, solely focusing on improvements from baseline in the active treatment group rather than differences between the active group and the placebo or comparison group, or highlighting the upper bound of the CI for the treatment group difference to suggest a meaningful positive effect. It is important for the reader to attend to the primary question

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that the trial was designed to answer and not be misled by secondary outcomes or analyses that only report effects within a treatment group.

The use of p-values in the context of scientific reporting has come under intense scrutiny in recent years due to their misuse and misinterpretation. For example, a common misinterpretation of a p-value is that it is the probability that the null hypothesis (of, say, no effect of treatment) is true. In fact, the p-value is the probability that a treatment effect larger than that observed in the trial would occur under an assumed statistical model if the null hypothesis were true. Also, the interpretation of a trial result as "proof" that a treatment is effective if it is statistically significant (e.g., $p < 0.05$) is flawed, as is the interpretation that a treatment is ineffective if the result is not statistically significant (e.g., $p > 0.05$). Indeed, the sole use of strict dichotomies with respect to p-values to judge whether or not a treatment is effective is highly problematic. For example, the level of evidence with respect to the existence of a treatment effect is certainly not qualitatively different when $p = 0.0499$ and when $p = 0.0501$. Another major problem with the misuse of p-values in clinical trials is the failure to account for multiplicity when several hypotheses are being tested, particularly with respect to secondary analyses (e.g., secondary outcome variables, multiple group comparisons, and subgroup analyses).

Although p-values provide some indication of how compatible the trial data are with a given null hypothesis, they do not convey any information regarding the clinical meaningfulness of the treatment effect. With a large enough sample size or low variability in outcomes, an observed treatment effect that is clinically insignificant can be associated with a statistically significant result [121]. Conversely, as discussed above, a trial with a small sample size or high variability in outcomes can yield a large estimated treatment effect that is not statistically significant. Understanding the potential clinical importance of a treatment effect requires more information, such as the estimated magnitude of the effect and the associated CI around that effect.

Useful discussions of the issues surrounding p-values and their interpretation can be found in recently published series of articles [121, 122] and the references therein. In the spirit of improved reporting, the CONSORT 2010 checklist recommends reporting the effect size estimate and some measure of the precision of that effect estimate [99]. In addition, many journals are moving toward reporting the effect size, the associated confidence interval, and the exact P-value rather than $P < 0.05$ [11, 50, 72], with at least one journal requiring that only effect size estimates and 95% CIs (with no P-values) be reported when researchers do not have a prespecified method to adjust for multiple analyses [50].

4.2. What is a clinically important benefit?

Statistically significant evidence of a treatment's efficacy in a clinical trial is insufficient to indicate that the magnitude of the treatment effect is clinically important. For example, if the sample size is sufficiently large, very small group differences may be "statistically significant" even though they are clinically irrelevant. Evaluations of clinical importance must distinguish between determining whether the mean improvements are important to patients or whether the group differences between treatments in an RCT are clinically important. A third type of evaluation involves determining whether the benefits of a

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

treatment are meaningful to society (e.g., reducing healthcare costs or increasing worker productivity), which is an important consideration but one that is beyond the scope of this article.

4.2.1. Clinical importance of improvements in individual patients—

Determining the magnitude of reduction in pain that is meaningful to patients with acute or chronic pain is important to the field of pain. Results of this research indicate that reduction in pain intensity of 10-20% is considered to be a “minimally important” pain intensity reduction on a patient global impression of change (PGIC) scale, a ≥ 30% reduction corresponds to what patients would consider a “moderately important” improvement in pain intensity, whereas reductions of approximately 50% or more can be considered “substantial” improvements in pain intensity for individuals with acute and chronic pain [15, 27]. However, the importance of such decreases could differ depending on the patient’s baseline pain intensity. For example, a decrease in pain from 8 to 6 on a 0-10 NRS, which can be considered a reduction from severe to moderate pain, might be more important to a patient than a reduction from 3 to 1, both of which are mild levels of pain. Alternatively, as Hanley et al. [49] have shown, individuals with higher pain intensity at baseline require a greater reduction in pain intensity for it to be considered a meaningful decrease. It may also be possible that two individuals experience a reduction in their pain intensity by the same approximate percentage, but for individual A, the pain decreases from a 5 to a 3 on the 0-10 NRS, whereas for individual B, the pain decreases from an 8 to a 5. Although the treatment leads both to experience a 38-40% reduction in pain intensity, individual A may judge the current level of pain to be acceptable, whereas individual B may experience the reduced level of pain as unsatisfactory [109]. Nevertheless, percentage reduction in pain is generally considered a useful approach to determining whether a patient has had a meaningful improvement than an absolute change on an NRS [27, 36, 89, 90].

Decreases in pain intensity, however, do not necessarily correspond to the magnitudes of overall improvement preferred by patients [38, 46]. For example, a clinically important reduction in pain intensity could be accompanied by considerable adverse effects, with health-related quality of life unimproved or even worsened as a result; conversely, treatment might be associated with a modest decrease in pain but substantial improvements in sleep, mood, and function that taken together would be considered a major benefit by the patient.

4.2.2. Clinical importance of group differences in a clinical trial—

The determination of the level of improvement patients consider clinically important is very often confused with evaluation of the group differences between an active and a control treatment. Thresholds for meaningful within-patient change (e.g., a reduction of 2 points on a 0-10 pain intensity NRS) should not be confused with the evaluation of what constitutes a meaningful difference between treatment groups. The determination of the clinical importance of group differences in RCTs depends on a constellation of factors, including: (1) the magnitude of the group difference observed in the trial and its associated CI; (2) the broader context of the disease being treated, including whether other treatments are available; (3) adverse events associated with the treatment, and (4) an overall evaluation of the benefit-risk profile, ideally as assessed by patients, clinicians, researchers, statisticians,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and other stakeholders [20, 23, 47, 71, 98]. For example, a reduction in pain intensity of at least 2 points on a 0-10 NRS could be used to define a clinically meaningful improvement for an individual patient, but the difference in mean change from baseline between an active treatment and placebo does not necessarily need to be ≥ 2 points in order for the effect of the treatment to be considered clinically important. The interpretation of meaningful change depends on whether it is being considered at the group level (where smaller between group differences in changes from baseline may be interpreted as important) or at the individual level, where thresholds for meaningful change are typically based on input from patients regarding what they consider important [10].

A number of factors can be considered when evaluating the clinical importance of group differences in an RCT (see Table 2). The first consideration is that there must be a statistically significant difference between the groups, which is a necessary but not sufficient criterion. In addition, the group difference (e.g., as assessed by the standardized effect size) with respect to the primary outcome variable can be compared with the effects associated with other treatments that are considered to have clinically important benefits. If the treatment effect in an RCT of a new treatment is comparable to, or greater than, the effects seen with established therapies, then the improvement is likely to be clinically important, although studies confirming this finding would be necessary. If the treatment effect found with the new treatment is substantially smaller than what has been found for existing therapies, then it becomes essential to evaluate whether there are any other characteristics of the new treatment that might compensate for the modest treatment effect on the primary outcome variable and make the overall benefit clinically important. Other characteristics to consider include safety and tolerability, results for secondary efficacy outcomes including physical and emotional functioning, limitations of existing treatments, and the other factors listed in Table 2. Importantly, cross-study comparisons of different treatments may not reflect what would occur if the different treatments were compared within the same study.

4.3. Placebo response and placebo effect

The placebo effect, or expectations of a treatment benefit, can play a role in the observed treatment benefit [33]. The placebo effect has neurobiological and physiological mechanisms that are activated by situational effects, interpersonal interactions, verbal suggestion, conditioning processes that include prior experiences with treatments, and other nonspecific effects [17, 33]. This differs from the placebo group response, which captures all changes that occur for patients when an inactive substance is administered, including regression to the mean, disease natural history, and the mechanisms associated with the placebo effect [33]. In RCTs for pain clinical trials, placebo group responses appear to have increased over time, perhaps due to increasing placebo effects, which makes it more difficult to identify efficacious treatments [26, 51, 113, 119]. Some have recommended performing studies to identify and understand the expectations of study participants by conducting a 3-arm trial of which one involves no intervention or by assessing participant expectations [17, 119].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

4.4. Analysis of means

In pain clinical trials, the mean change in the primary outcome measure (i.e., the prespecified measure on which a difference between the treatment and the control group is expected) from the baseline pretreatment period to a designated time point after initiation of treatment for each treatment arm is typically reported (i.e., within-group change). Within-group changes indicate the average change (or no change) that is observed during the course of the study for participants in each treatment arm. However, formal analyses of within-group change do not address the true objective of the RCT – evaluating whether the difference between treatment arms over the course of the study is statistically and clinically significant. In some instances, researchers will employ analysis of means testing such as a t-test or analysis of variance (ANOVA) to compare the mean within-group changes between treatment arms. An analysis strategy that makes more efficient use of the baseline information is analysis of covariance, for which the statistical model includes treatment group and the baseline value of the outcome variable as independent variables. This strategy yields a more precise estimate of the treatment effects than the simple group comparison of mean within-group changes from baseline since the latter strategy incorporates the baseline value in a very limited way (i.e., only in the definition of the outcome variable) [102, 104]. Studies that only report the statistical significance of within-group changes for each treatment arm without statistical comparisons between the groups fail to demonstrate that a treatment provides any benefit beyond a placebo (or other comparator), as changes from baseline can be due to many factors other than the treatment effect (e.g., regression to the mean, symptom fluctuations, contextual influences). It is necessary to show that the changes from baseline are greater for one treatment group than the other in order to demonstrate a benefit of the treatment being studied.

When incorporating a treatment into a clinical setting, it is important to recognize that comparisons of group means indicate what is happening on average across *all participants*, and as is the case for all analyses used in trial designs other than multi-period cross-over studies, they are not informative about the responses of *an individual patient* [27, 68, 101]. For example, although an RCT may show that patients who receive a specific treatment report greater analgesic benefit, on average, than those who receive placebo, patients in the treatment and placebo arms may experience improvement, no change, or even increases in pain. Of course, while this point has often been made about group means, it applies to any group-level estimand such as a propoportion (e.g., of "responders"; see Section 4.5. below).

4.5. Responder analyses

An alternative way to analyze RCT data is to compare the treatment groups with respect to the percentage of patients whose improvements meet a pre-defined threshold. Common examples are categories of severity such as mild, moderate, or severe pain, or categories of reporting changes such as the percentage reporting ≥30% or ≥ 50% reductions in pain intensity. There is a lack of consensus regarding the pros and cons of responder analyses, and the related metric of number needed to treat (NNT). Presenting responder analyses can simplify the interpretation of trial results, allowing for a straightforward comparison of the proportion of patients in each treatment arm who experienced a pre-defined level of improvement on the outcome of interest [34]. However, responder analyses require an

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

understanding of what is clinically meaningful to different stakeholders in order to define what constitutes a "response." In other words, how much within-patient improvement in pain intensity is necessary for patients, clinicians, or other stakeholders to identify the pain reductions as clinically meaningful? As described in section 4.2.1. above, empirical evidence suggests that a reduction of 10-20% on the 0-10 NRS for pain intensity is associated with minimal improvement, a reduction of ≥30% is needed before patients report moderate change in pain intensity, and a ≥50% reduction is considered substantial improvement [27] [49].

The use of the phrase "responder" can be erroneously interpreted to refer to a stable characteristic of the participant, implying that the participant will respond to the treatment regardless of context. In fact, multiple randomized exposures to both the treatment and a control are necessary to determine whether or not a patient responds to the treatment [21, 105]. An important limitation of using proportions in responder analyses is the loss in statistical power associated with dichotomizing continuous data into categorical data (i.e., "responder" vs. "non-responder"). The reduction in power occurs because dichotomization sacrifices information; for example, consider that a patient who has a 29% reduction in pain intensity would not be considered a "responder", whereas a patient with a 31% reduction would be, despite the fact that their pain reductions are nearly identical, and the patient with the 29% reduction may consider that reduction meaningful [4, 103, 108].

4.6. Cumulative distribution functions (CDFs)

A challenge previously discussed regarding responder analyses is that investigators must specify the decrease in pain intensity that must occur for study participants to be categorized as "responders". However, solely reporting the percentage of RCT participants who have reported one or more distinct levels of reduction in pain intensity does not provide complete information about the trial. When contemplating use of a treatment in a clinical setting, clinicians may want to know the percentages of participants in each group who experienced different levels of reduction in pain intensity (e.g., 20% or 75% reduction). Farrar and colleagues proposed an alternative method of reporting RCT results, cumulative distribution functions (CDFs), which graphically depict a continuous plot of the percentages of participants in each treatment arm across the entire range of possible responses (see Figure 2 for an example) [35]. The main advantage of this information is that it provides a visual representation of the treatment group differences in percentage of "responders" at each level of "response," including the full range of improvement. In this way, readers can apply their own definitions of a meaningful improvement when interpreting the results. Multi-tiered information, in conjunction with analysis of means, can be valuable in establishing whether the treatment benefit is clinically meaningful to patients [23, 114]. Presenting CDFs allows the reader to identify the percentages of participants who achieved each level of "response" and how that differed between treatment groups. In addition, the difference between the CDF curves at any "response" threshold is the absolute risk reduction (ARR), which can then be used to calculate the NNT (see below).

CDFs are a useful descriptive tool that can visually depict the data from an RCT. As with any RCT analysis, it is important that the problem of missing data is addressed before

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

computing the CDF. Often, researchers presume that anyone who drops out of an RCT is a "non-responder". However, that is not necessarily the case, particularly when the reason for dropout is unrelated to treatment. Methods to accommodating missing data are briefly addressed in section 7 below.

4.7. Number needed to treat (NNT), number needed to harm (NNH), and relative risk (RR)

The NNT is a value that summarizes a treatment group comparison with respect to the incidence of some event (e.g., "response") between a study's treatment arms [74, 82]. The NNT is calculated as 1/ARR, or the absolute risk reduction. The ARR reflects the difference between treatment groups in the percentages of participants experiencing an event (e.g., difference in the percentages of "responders" between the treatment and placebo arms). The NNT indicates the expected number of people who would need to take the treatment for there to be 1 additional "responder" beyond the number of "responders" in the placebo or other comparator arm [5, 66]. It can be calculated from the difference between the two CDF curves at a particular "response" threshold.

As is the case for responder analyses, the NNT is designed to simplify RCT interpretation and to permit comparisons across studies of different treatments for comparable diagnoses. For example, consider a hypothetical 16-week RCT comparing an analgesic treatment and placebo in which "responders" are those individuals who report a reduction in pain intensity ≥30% from baseline to end of treatment. If the results of the RCT indicate that 40% of the participants in the analgesic treatment arm and 20% in the placebo arm are "responders", the ARR would be 0.40 – 0.20 = 0.20, and the NNT would be 1 / (0.40 – 0.20) = 5. The interpretation would be that if 5 individuals took the treatment for 16 weeks, 2 would be expected to meet the "response" criterion (i.e., 5 × 0.40), whereas if 5 individuals took the placebo for 16 weeks, 1 would be expected to meet the "response" criterion (i.e., 5 × 0.20); there would therefore be 1 additional "responder" among the patients receiving the active treatment versus those administered placebo [66]. The number needed to harm (NNH) is calculated in the same way using the incidence of adverse events (AEs) or other safety outcomes in the treatment and comparator arms over a given time period [5, 28, 82]. The NNH can be misleading, however, particularly if it aggregates AEs of varying severity and seriousness (e.g., if mild dry mouth and death are equally weighted when counting AEs). Thus, NNH calculations are much less frequently found in RCT reporting than NNT calculations.

Relative risk (RR) is another way to summarize the comparative incidence of "response" between two treatment arms in a study, providing information about the likelihood of achieving a "response" [1, 52]. The RR is calculated by dividing the percentage of patients who are "responders" in the analgesic treatment arm by the percentage of "responders" in the placebo arm [1, 52]. Using the same example of an RCT where 40% of the analgesic treatment arm participants were "responders" and 20% of the participants in the placebo arm were "responders", the RR would be 40% / 20% = 2.0, indicating that participants in the analgesic treatment arm were twice as likely to be a "responder." Because the word "risk" usually connotes harm, the interpretation of relative risks could be facilitated if the risks of nonresponse were presented (i.e., (100%−40%)/(100%−20%) = 0.75), indicating that the RR

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of nonresponse is 0.75 rather than that the RR of response is 2. An RR of 1 indicates that there is no difference between the two treatment groups [92]. The ARR may be more informative to clinicians and patients than the RR, however, in that it describes the difference in the incidence of the event of interest between the two treatment groups (e.g., percentages of "responders;" 0.40 – 0.20 = 0.20 or 20% difference in event incidence), rather than describing the relative probability of "response" in the treatment group compared to the placebo group (e.g., 2 times more likely to meet the "response" criterion) [1]. An RR of 2.0 could reflect very different ARRs, such as 20% (40% – 20%) or 5% (10% – 5%). Relative risks presented without the absolute risk are a common cause of confusion, in which the risk to an individual can appear exaggerated.

The limitations of the NNT mirror those of responder analyses, given that the NNT is based on categorizing study patients as having met a pre-defined threshold of improvement from baseline. Additionally, the NNT is frequently misinterpreted as indicating the number of individuals who need to be treated in order for 1 patient to be a "responder" [66], rather than the number of patients who would need to take the treatment for there to be 1 additional "responder" beyond the number of "responders" that would occur in the placebo arm. Clinically, this distinction is important. Using the example above, the NNT of 5 does not mean that we would expect 1 person to be a "responder" among 5 patients taking the active treatment, but rather that there would be 1 more "responder" among 5 patients receiving the active treatment than there would be among 5 patients taking placebo. It is also important to recognize that an NNT that is calculated using a specific threshold can be incorrectly compared to an NNT calculated with a different threshold (e.g., an NNT calculated using a 30% reduction in pain intensity to define "responder" would be interpreted differently than an NNT calculated using a 50% reduction) [35, 108].

There is a lack of consensus among the authors regarding whether NNTs contribute to the interpretation of RCTs and the extent to which they are incorrectly interpreted. As an example, "consider a trial comparing paracetamol to a placebo for treating tension headache. After 2 hours, 50% of people treated with the placebo are pain-free, as are 60% of those who were treated with paracetamol. The difference is 10% and the NNT is 10. However, if paracetamol works for 100% of participants in 60% of the times they are treated, it will give the same NNT as if it works for 60% of the participants 100% of the time. A high NNT should not be taken to imply that a drug works really well for a specific, narrow subset of people. It could simply mean that a drug is just not that effective across all individuals" (pp. 620-1) [106]. For additional reading on the limitations of NNTs, see also [66, 79].

4.8. Time to effect and duration of effect

Beyond identifying the effect of a treatment, data on the temporal course of that effect can provide information regarding the time to onset of a beneficial effect and how long the beneficial effect lasts. These details provide clinicians and patients with a more comprehensive understanding of a treatment's overall clinical effect [22]. Although presenting a treatment's time course data may be valuable for interpreting treatment effect, there is no standardized method to assess time to effect or duration of effect. One possibility, for example, is to graphically present each treatment arm's mean pain intensity and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

variability across each week of the RCT, allowing the reader to interpret the time course. It may also be informative to present data indicating when the mean pain intensity in the treatment arm first demonstrates a statistically significant or clinically meaningful separation from the placebo arm, although this may conflate time to onset with sample size. Additionally, data regarding time to a clinically relevant event (e.g., minimal pain, discontinuation of treatment) in each treatment arm can be used to assess the time to onset of a beneficial effect. In acute pain RCTs, for example, researchers may use the double stopwatch method to capture the onset of first pain relief, as well as the onset of meaningful pain relief [73]. Clinicians might also want data on the proportion of "responders" at each assessment period separated by treatment arm. Reports of RCTs may use different methods to present data on a treatment's time course, making it difficult to compare the temporal course of various treatments. As yet, this approach has not been frequently used in chronic pain RCTs. A further limitation is that few studies are conducted to examine the long-term durability of a treatment's effect. Frequently the only available data on the duration of a treatment's effect comes from RCTs that are 12 to 16 weeks in length (and sometimes shorter; see [16, 64, 83] showing that the double-blind period of opioid analgesic trials is frequently 6 weeks or less); such data cannot speak to the effect of the treatment when used over an extended period of time as might be expected in many chronic pain conditions.

One approach that may provide information about the duration of analgesic effect is an adaptation of the randomized withdrawal design. In randomized withdrawal studies, all patients initially receive active treatment and are then randomized to stay on active treatment or receive placebo [84]. At the end of the treatment period, individuals in both the treatment and placebo arms who meet some pre-defined threshold (e.g., ≥30% reduction in pain intensity) could be followed to determine the duration of the treatment effect in a double-blind long-term efficacy study [75].

5. Treatment risks

A complete description of a treatment's clinical effect requires not only reporting efficacy results, but safety (i.e., AEs identified through patient symptoms or clinically assessed signs) as well. AEs provide important information regarding the tolerability and safety of a treatment, and can have implications for patients' perceptions of the effectiveness of the treatment. For example, adverse treatment effects have been shown to be associated with increased reports of pain interference beyond the effect of pain intensity itself [77], suggesting that it is not a treatment's analgesic benefit alone that affects pain outcomes. The CONSORT group has published guidance for comprehensive and transparent reporting of AEs occurring in an RCT [58] (see also [43]). It is important to recognize that AEs are not all equivalent, in that the seriousness and severity of the AEs affect when a clinician might opt to introduce the treatment. For AEs that are more serious or severe, the treatment might only be considered when all other treatment options have been exhausted and the patient is debilitated by their symptoms. For AEs that are mild or moderate in severity, the treatment might be discussed with a patient at an earlier stage in the context of balancing the treatment's benefits and side effects.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In addition to documenting the types and numbers of AEs occurring in an RCT, it is essential for reports of RCTs to describe the methodology for acquiring data regarding AEs. For example, with passive capture, AEs may be collected solely when study participants self-disclose without any prompting or in response to a question such as, "Have you had any problems since the last visit?" With active capture, one could have a checklist of potential AEs and ask the subject if (s)he has experienced each AE. Research has shown that passive capture can result in underestimation of the harms experienced by study participants [6, 9, 67], although active capture has the potential to overestimate the number of harms observed. The methods used to capture AEs should be prespecified and described in articles reporting RCTs so the adequacy of the methods can be evaluated. Other critical details regarding potential treatment harms that should be reported are: (1) the number and nature of the specific AEs that were identified and reported throughout the trial, (2) the severity of the AEs (i.e., mild, moderate, severe), (3) definitions of each severity grade, (4) occurrence of serious AEs (SAEs; e.g., hospitalization or death) [115], (5) whether or not the AEs were considered to be plausibly associated with the treatment (i.e., adverse reactions), and (6) whether standardized coding methods (e.g., MedDRA; http://www.meddra.org) were used to classify AEs (e.g., "feeling nauseated", "feeling queasy") into meaningful categories (e.g., nausea). These characteristics of a treatment's risks are particularly important because they have implications for comparisons across trials. It is also necessary to consider the length of the RCT, given that the risk of harm may occur after long-term use, rather than in 12-16 week trials.

Systematic reviews of pain trials have shown that investigators do not always report all study AEs [54, 107, 124]. Instead, subsets of AEs are frequently reported (e.g., "common" AEs, or AEs for which treatment group differences in incidence were statistically significant). Such truncated AE reporting is likely due in part to a desire to briefly summarize study AEs, as well as to meet limitations imposed by journal publishers. One drawback of reporting subsets of AEs is that rare AEs with possible serious clinical implications may not be adequately disclosed. Solely reporting the AEs that demonstrate a statistically significant difference between treatment arms can be misleading because the RCT may not have been sufficiently large to detect differences in the incidence of certain AEs between treatment arms [3, 56, 57, 63, 112]. Alternately, apparently statistically significant differences in AEs between treatment arms may be false positives that arise from multiple statistical tests without any adjustments for multiplicity [40]. Adequate AE reporting, therefore, involves reporting the denominator for all AE data and reporting both the number of events and the number of study participants who experience each AE that causes study withdrawal, as well as the moderate, severe, and serious AEs that occurred during the study [58, 107]. Additional AE detail could then be made available in online journal supplements [107].

6. Summarizing and integrating treatment benefits and risks

Interpreting the overall effect of a treatment necessitates considering the treatment's benefits alongside its risks. Benefit-risk evaluations can be used to guide individual treatment decisions by patients and their clinicians, and can also be made at the societal level as a basis for regulatory approvals, reimbursement decisions, and medical policies. However, it can be practically difficult to weigh the benefits against the risks. Developing a standardized

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

method to integrate treatment benefits and risks in order to provide an easily interpretable metric that represents the treatment's benefit-risk profile is a complicated endeavor and currently no single, well-accepted method exists. Furthermore, determining whether treatment benefits outweigh the risks may require information about the medical history and preferences of the individual patient who will receive the treatment. Despite these challenges, there are several approaches to synthesizing the body of evidence across RCTs regarding a treatment's effects that are important to highlight.

Aggregating data on a treatment's overall benefits and risks can be done through a systematic review, meta-analysis, or integrated benefit-risk methods. Conducting a systematic review to identify all research published on a specific analgesic treatment, combining the efficacy data and the AE data across all trials, and analyzing those data is an effective way of consolidating the available evidence. There are limitations to meta-analyses, however. One concern is that the quality of the meta-analysis depends upon the quality of the research that goes into it. Poorly designed and executed studies are likely to be biased in a variety of ways (e.g., selection bias, performance bias), which affects the validity of their results, and this risk of bias should be accounted for in the meta-analysis [52]. Meta-analyses also require a method to aggregate trial outcomes that may not have been assessed in the same manner (e.g., continuous outcomes, time to effect outcomes). Furthermore, meta-analyses may be no better than post-marketing safety surveillance at identifying rare events that have not been previously identified [81]. Meta-analyses also tend to include only published data. Given a well-known reporting bias whereby study results that fail to show a treatment effect are less likely to be published, meta-analyses can be biased toward demonstrating a treatment effect that is larger than the true treatment effect [52, 80]. Despite these limitations, meta-analyses can provide relatively complete information regarding a treatment's benefit and risk profile, which can help to inform clinical practice. Cochrane has developed comprehensive guidance regarding the conduct of systematic reviews and meta-analyses for treatment interventions [52], including the adoption of Grades of Recommendation, Assessment, Development, and Evaluation (GRADE). GRADE is an approach designed to evaluate the risk of bias in individual studies, and to provide a rating of confidence in the overall estimate of any effect and the likelihood that the estimate could be changed by additional data [8].

In the past decade, several initiatives have focused on advancing the methodology for integrating the assessment of a therapy's benefit and risk into a single framework. Integrated benefit-risk frameworks may be qualitative, quantitative, or include both qualitative and quantitative components [86]. Qualitative frameworks include visual displays (i.e., tables, figures) that list key benefit and risk attributes [86]. One example of a qualitative benefit-risk framework is the FDA's Benefit-Risk Integrated Assessment [116]. This structured framework is in a tabular format that allows reviewers to synthesize the evidence of the therapeutic context (i.e., analysis of the condition and current treatment options) and evidence supporting the benefit and risk and risk management strategies of the product that weighed in their decision-making. The European Medicines Agency (EMA) has issued a guidance document that addresses benefit-risk assessment that does not recommend a specific quantitative methodology, but is open to considering these approaches on a case-by-case basis [29].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Another method to integrate and summarize a treatment's benefit-risk assessment that includes both qualitative and quantitative components is the Benefit Risk Action Team (BRAT) framework [18]. The important features of this approach include identifying the research context (e.g., condition being treated, treatment comparator) and the essential benefits and risks across studies, and then presenting the data regarding those benefits and risks in a way that can be easily understood (e.g., plots of differences in benefits and risks between treatment groups; see [18] for an example). This may assist clinicians in interpreting the overall benefit-risk profile, which can be integrated with their clinical expertise when treating patients. A more recent review identified 49 different methodologies to conduct quantitative and systematic benefit-risk assessment of medications [85]. These methods range from descriptive to more quantitative. The problem, objectives, alternatives, consequences, trade-offs, uncertainty, risk and linked decisions framework (PrOACT-URL) provides another method to descriptively report the risks associated with a treatment [85] (see http://protectbenefitrisk.eu/PrOACT-URL.html for examples). Other more quantitative methods include multi-criteria decision analysis (MCDA) that compares treatment options based on weighted benefit and risk criteria [76, 88], stochastic multicriteria acceptability analysis (SMAA; derived from MCDA) [85], benefit-risk ratio (BRR) that reflects the treatment risks divided by the benefits [85], stated-choice surveys of willingness to accept risks [61] or discrete choice experiments [85], and health outcomes modeling using the quality-adjusted life-year (QALY) [19].

Typical benefit-risk analyses involve separate intervention comparisons for each efficacy, safety, and quality-of-life outcome. Outcome-specific effects are tabulated and combined (systematically or unsystematically) in benefit-risk analyses so that such analyses can describe the totality of effects on patients. However, such approaches do not incorporate associations between outcomes of interest, fail to summarize the cumulative nature of different outcomes on individual patients, and suffer from competing risk challenges when interpreting individual outcomes. In addition, because efficacy and safety analyses are conducted on different subsets of participants, the population to which these benefit-risk analyses apply is unclear. New benefit-risk methodologies continue to be developed such as the desirability of outcome ranking (DOOR) and partial credit which attempt to address the limitations of prior methods by ranking various study outcomes using predetermined criteria [30-32].

7. Select Areas of Advancement for Clinical Trials

In 2010, the National Research Council (NRC) produced a report on The Prevention and Treatment of Missing Data in Clinical Trials [87], and in 2019 an addendum to the ICH E9 guidance on Statistical Principles for Clinical Trials [55] was released, leading to a major shift in how clinical trialists think about trial design and analysis. The NRC report drew attention to the significant limitations of existing simplistic methods for dealing with missing data, such as carrying forward the last (or baseline) observation. They also emphasized the use of more principled methods to deal with the problem such as direct likelihood methods (e.g., mixed model repeated measures, or MMRM), multiple imputation, or generalized estimating equations. Because any method to deal with missing data is based on untestable assumptions, the NRC report and the ICH E9 addendum further emphasized

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the importance of performing sensitivity analyses (i.e., analyses that make different assumptions concerning the distribution of the missing values given the observed data) to determine the degree of the dependence on the inference concerning the treatment effect on these assumptions.

The ICH E9 draft addendum discusses the importance of precise formulation of the estimand(s) of interest based on the study objectives [55]. The estimand consists of (1) the patient population of interest, (2) the outcome variable, (3) how post-randomization events (intercurrent events) will be handled, and (4) the population-level summary for the outcome variable. In particular, much thought needs to be given to how to deal with intercurrent events in formulating the estimand. The most common intercurrent events include dropout, discontinuation of the study intervention, use of prohibited medications and other protocol violations, and use of rescue medication [14]. The choice of an estimand depends on the characteristics of the treatment (e.g., disease modifying vs. symptom control), setting of treatment use (e.g., the ability to monitor outcome over time), and the choice of the control group [55]. The estimand has a major influence on the study design, the data to be collected, and how the data should be analyzed [14].

The vast majority of clinical trials in pain have used standard parallel group or cross-over designs. Some advances in trial design that are beginning to be used in the pain field include cross-over trials with multiple periods, enrichment, methods to reduce the amount of improvement in placebo groups, adaptive designs, and master protocols. A summary of these trial designs is provided in Table 3.

8. Research Agenda

Improving the interpretation of trial data in clinical or public health decision-making research would make an important contribution in a number of related areas. First, models of shared decision making that include data from meta-analyses of trials could usefully be developed for pain treatments, including a focus on the values and preferences of both patients and clinicians. Second, when asked how data should be presented, people often prefer that complexity be reduced as much as possible. A recent survey of clinicians across 8 Western countries found that clinicians differentially understand various approaches to presenting treatment effects. Methods that employ dichotomized continuous outcomes (e.g., risk reduction) were considered by clinicians to be the most accessible, although the percentage of clinicians who correctly interpreted these methods was still below 50% [62]. Efforts to extend these findings to evaluate the information stakeholders want from RCTs, and how to educate stakeholders on the judicious use of both continuous and dichotomized data and the limitations of these data, would be valuable. Additionally, experimental research comparing different communication strategies and their effects on decision making and clinical outcomes would help us to understand the potential risks of these communication strategies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

9. Conclusions

Interpreting RCTs and their implications for clinical practice can be complicated due to the variety of methods that researchers use to report their findings. Becoming familiar with typical reporting approaches, as well as their strengths and limitations (see Table 4 for considerations regarding efficacy reporting methods), may assist clinicians in understanding a treatment's observed benefits and risks and how those effects might translate to a clinical setting.

Acknowledgments

We thank Valorie Thompson and Andrea Speckin for their assistance in organizing the meeting on which this article is based. We acknowledge Stephen Senn for his participation in the IMMPACT meeting on which this manuscript is based, as well as to this manuscript. We further acknowledge Stephen Morley for his participation in the IMMPACT meeting, and who passed before publication.

References

1. Alhazzani W, Walter SD, Jaeschke R, Cook DJ, Guyatt G. Does Treatment Lower Risk? Understanding the Results In Guyatt G, Rennie D, Meade MO, & Cook DJ, editors. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd ed. New York, NY: McGraw-Hill Education, 2015.
2. Altman DG, Bland JM. Improving doctors' understanding of statistics. Journal of the Royal Statistical Society: Series A (Statistics in Society) 1991;154:223–248.
3. Altman DG, Bland JM. Statistics notes: Absence of evidence is not evidence of absence. BMJ 1995;311:485. [PubMed: 7647644]
4. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332:1080. [PubMed: 16675816]
5. Andrade C. The numbers needed to treat and harm (NNT, NNH) statistics: what they tell us and what they do not. J Clin Psychiatry 2015;76:e330–3. [PubMed: 25830454]
6. Atkinson TM, Rogak LJ, Heon N, Ryan SJ, Shaw M, Stark LP, Bennett AV, Basch E, Li Y. Exploring differences in adverse symptom event grading thresholds between clinicians and patients in the clinical trial setting. J Cancer Res Clin Oncol 2017;143:735–743. [PubMed: 28093637]
7. Azmi S, ElHadd KT, Nelson A, Chapman A, Bowling FL, Perumbalath A, Lim J, Marshall A, Malik RA, Alam U. Pregabalin in the Management of Painful Diabetic Neuropathy: A Narrative Review. Diabetes Ther 2019;10:35–56. [PubMed: 30565054]
8. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401–6. [PubMed: 21208779]
9. Basch E. The missing voice of patients in drug-safety reporting. N Engl J Med 2010;362:865–9. [PubMed: 20220181]
10. Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, Strand V, Shea B. Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal Clinically Important Difference. J Rheumatol 2001;28:400–5. [PubMed: 11246687]
11. British Medical Journal. BMJ Guidance for Authors. 2018; Available from: https://www.bmj.com/sites/default/files/attachments/resources/2018/05/BMJ-InstructionsForAuthors-2018.pdf, accessed 1/17/2020.
12. 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]
13. Busse JW, Bartlett SJ, Dougados M, Johnston BC, Guyatt GH, Kirwan JR, Kwoh K, Maxwell LJ, Moore A, Singh JA, Stevens R, Strand V, Suarez-Almazor ME, Tugwell P, Wells GA. Optimal

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Strategies for Reporting Pain in Clinical Trials and Systematic Reviews: Recommendations from an OMERACT 12 Workshop. J Rheumatol 2015;42:1962–1970. [PubMed: 25979719]

14. Cai X GJ, He H, Turk DC, Dworkin RH, McDermott MP. Estimands and Missing Data in Clinical Trials of Chronic Pain Treatments: Advances in Design and Analysis. under review.

15. Cepeda MS, Africano JM, Polo R, Alcala R, Carr DB. What decline in pain intensity is meaningful to patients with acute pain? Pain 2003;105:151–7. [PubMed: 14499431]

16. Chou R, Clark E, Helfand M. Comparative efficacy and safety of long-acting oral opioids for chronic non-cancer pain: a systematic review. J Pain Symptom Manage 2003;26:1026–48. [PubMed: 14585554]

17. Colloca L. The Placebo Effect in Pain Therapies. Annu Rev Pharmacol Toxicol 2019;59:191–211. [PubMed: 30216744]

18. Coplan PM, Noel RA, Levitan BS, Ferguson J, Mussen F. Development of a framework for enhancing the transparency, reproducibility and communication of the benefit-risk balance of medicines. Clin Pharmacol Ther 2011;89:312–5. [PubMed: 21160469]

19. Cross JT, Veenstra DL, Gardner JS, Garrison LP Jr. Can modeling of health outcomes facilitate regulatory decision making? The benefit-risk tradeoff for rosiglitazone in 1999 vs. 2007. Clin Pharmacol Ther 2011;89:429–36. [PubMed: 21289618]

20. 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]

21. Dworkin RH, McDermott MP, Farrar JT, O'Connor AB, Senn S. Interpreting patient treatment response in analgesic clinical trials: implications for genotyping, phenotyping, and personalized pain treatment. Pain 2014;155:457–60. [PubMed: 24071599]

22. 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]

23. Dworkin RH, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P, Farrar JT, Hertz S, Raja SN, Rappaport BA, Rauschkolb C, Sampaio C. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. Pain 2009;146:238–44. [PubMed: 19836888]

24. Dworkin RH, Turk DC, Peirce-Sandner S, Baron R, Bellamy N, Burke LB, Chappell A, Chartier K, Cleeland CS, Costello A, Cowan P, Dimitrova R, Ellenberg S, Farrar JT, French JA, Gilron I, Hertz S, Jadad AR, Jay GW, Kalliomaki J, Katz NP, Kerns RD, Manning DC, McDermott MP, McGrath PJ, Narayana A, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Reeve BB, Rhodes T, Sampaio C, Simpson DM, Stauffer JW, Stucki G, Tobias J, White RE, Witter J. Research design considerations for confirmatory chronic pain clinical trials: IMMPACT recommendations. Pain 2010;149:177–93. [PubMed: 20207481]

25. 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]

26. Dworkin RH, Turk DC, Peirce-Sandner S, He H, McDermott MP, Farrar JT, Katz NP, Lin AH, Rappaport BA, Rowbotham MC. Assay sensitivity and study features in neuropathic pain trials: an ACTTION meta-analysis. Neurology 2013;81:67–75. [PubMed: 23700332]

27. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, Haythornthwaite JA, Jensen MP, Kerns RD, Ader DN, Brandenburg N, Burke LB, Cella D, Chandler J, Cowan P, Dimitrova R, Dionne R, Hertz S, Jadad AR, Katz NP, Kehlet H, Kramer LD, Manning DC,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

McCormick C, McDermott MP, McQuay HJ, Patel S, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Revicki DA, Rothman M, Schmader KE, Stacey BR, Stauffer JW, von Stein T, White RE, Witter J, Zavisic S. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105–21. [PubMed: 18055266]

28. Edelsberg J, Oster G. Summary measures of number needed to treat: how much clinical guidance do they provide in neuropathic pain? Eur J Pain 2009;13:11–6. [PubMed: 18456524]
29. European Medicines Agency. Guidance document for the content of the <co-> rapporteur day 80 critical assessment report. 2014.
30. Evans SR, Bigelow R, Chuang-Stein C, Ellenberg S, Gallo P, He W, Jiang Q, Rockhold F. Presenting risks and benefits: helping the data monitoring committee do its job. Annals of Internal Medicine in press.
31. Evans SR, Follmann D. Using Outcomes to Analyze Patients Rather than Patients to Analyze Outcomes: A Step toward Pragmatism in Benefit:risk Evaluation. Stat Biopharm Res 2016;8:386–393. [PubMed: 28435515]
32. Evans SR, Rubin D, Follmann D, Pennello G, Huskins WC, Powers JH, Schoenfeld D, Chuang-Stein C, Cosgrove SE, Fowler VG Jr., Lautenbach E, Chambers HF. Desirability of Outcome Ranking (DOOR) and Response Adjusted for Duration of Antibiotic Risk (RADAR). Clin Infect Dis 2015;61:800–6. [PubMed: 26113652]
33. Evers AWM, Colloca L, Blease C, Annoni M, Atlas LY, Benedetti F, Bingel U, Buchel C, Carvalho C, Colagiuri B, Crum AJ, Enck P, Gaab J, Geers AL, Howick J, Jensen KB, Kirsch I, Meissner K, Napadow V, Peerdeman KJ, Raz A, Rief W, Vase L, Wager TD, Wampold BE, Weimer K, Wiech K, Kaptchuk TJ, Klinger R, Kelley JM. Implications of Placebo and Nocebo Effects for Clinical Practice: Expert Consensus. Psychother Psychosom 2018;87:204–210. [PubMed: 29895014]
34. Farrar JT. What is clinically meaningful: outcome measures in pain clinical trials. Clin J Pain 2000;16:S106–12. [PubMed: 10870749]
35. 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;31:369–77. [PubMed: 16632085]
36. Farrar JT, Young JP Jr., LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149–58. [PubMed: 11690728]
37. Fava M, Evins AE, Dorer DJ, Schoenfeld DA. The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. Psychother Psychosom 2003;72:115–27. [PubMed: 12707478]
38. Ferreira ML, Herbert RD, Ferreira PH, Latimer J, Ostelo RW, Grotle M, Barrett B. The smallest worthwhile effect of nonsteroidal anti-inflammatory drugs and physiotherapy for chronic low back pain: a benefit-harm trade-off study. J Clin Epidemiol 2013;66:1397–404. [PubMed: 24021611]
39. Flay BR, Biglan A, Boruch RF, Castro FG, Gottfredson D, Kellam S, Moscicki EK, Schinke S, Valentine JC, Ji P. Standards of evidence: criteria for efficacy, effectiveness and dissemination. Prev Sci 2005;6:151–75. [PubMed: 16365954]
40. Fleming TR. Identifying and Addressing Safety Signals in Clinical Trials. New England Journal of Medicine 2008;359:1400–1402. [PubMed: 18768938]
41. Gartlehner G, Hansen RA, Nissman D, Lohr KN, Carey TS. Criteria for Distinguishing Effectiveness From Efficacy Trials in Systematic Reviews, in Technical Review 12. AHRQ Publication No. 06–0046. 2006 Agency for Healthcare Research and Quality: Rockville, MD.
42. Gartlehner G, Hansen RA, Nissman D, Lohr KN, Carey TS. A simple and valid tool distinguished efficacy from effectiveness studies. J Clin Epidemiol 2006;59:1040–8. [PubMed: 16980143]
43. Gewandter JS, Eisenach JC, Gross RA, Jensen MP, Keefe FJ, Lee DA, Turk DC. Checklist for the preparation and review of pain clinical trial publications: a pain-specific supplement to CONSORT. PAIN Reports 9000;Latest Articles.
44. Gewandter JS, McDermott MP, Kitt RA, Chaudari J, Koch JG, Evans SR, Gross RA, Markman JD, Turk DC, Dworkin RH. Interpretation of CIs in clinical trials with non-significant results: systematic review and recommendations. BMJ Open 2017;7:e017288.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

45. 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]

46. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. Bmj 1995;311:1356–9. [PubMed: 7496291]

47. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. BMJ 1998;316:690–3. [PubMed: 9522799]

48. Hackshaw A, Kirkwood A. Interpreting and reporting clinical trials with results of borderline significance. BMJ 2011;343:d3340. [PubMed: 21727163]

49. Hanley MA, Jensen MP, Ehde DM, Robinson LR, Cardenas DD, Turner JA, Smith DG. Clinically significant change in pain intensity ratings in persons with spinal cord injury or amputation. Clin J Pain 2006;22:25–31. [PubMed: 16340590]

50. Harrington D, D'Agostino RB Sr., Gatsonis C, Hogan JW, Hunter DJ, Normand ST, Drazen JM, Hamel MB. New Guidelines for Statistical Reporting in the Journal. N Engl J Med 2019;381:285–286. [PubMed: 31314974]

51. Häuser W, Bartram-Wunn E, Bartram C, Reinecke H, Tölle T. Systematic review: Placebo response in drug trials of fibromyalgia syndrome and painful peripheral diabetic neuropathy—magnitude and patient-related predictors. PAIN® 2011;152:1709–1717. [PubMed: 21429668]

52. Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (eds). Cochrane handbook for systematic reviews of interventions 2019; Version 6.0 (updated July, 2019):[Available from: www.training.cochrane.org/handbook; accessed 10/10/19.

53. Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, Moschetti I, Phillips B, Thornton H, Goddard O, Hodgkinson M. The Oxford 2011 levels of evidence. 2011 11 10, 2016]; Available from: http://www.cebm.net/index.aspx?o=5653.

54. 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;155:2253–62. [PubMed: 25123543]

55. International Committee for Harmonisation. Addendum on Estimands and Sensitivity Analysis in Clinical Trials. 2019; Available from: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf, accessed 1/17/20.

56. Ioannidis JA, Mulrow CD, Goodman SN. Adverse events: The more you search, the more you find. Annals of Internal Medicine 2006;144:298–300. [PubMed: 16490917]

57. Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. Arch Intern Med 2009;169:1737–9. [PubMed: 19858427]

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

59. Ivanova A, Tamura RN. A two-way enriched clinical trial design: combining advantages of placebo lead-in and randomized withdrawal. Statistical Methods in Medical Research 2015;24:871–890. [PubMed: 22143405]

60. Jadad AR, Enkin MW. Randomized controlled trials: questions, answers, and musings. Oxford, UK: Blackwell Publishing, 2007.

61. Johnson FR, Hauber AB, Ozdemir S, Lynd L. Quantifying women's stated benefit-risk trade-off preferences for IBS treatment outcomes. Value Health 2010;13:418–23. [PubMed: 20230550]

62. Johnston BC, Alonso-Coello P, Friedrich JO, Mustafa RA, Tikkinen KA, Neumann I, Vandvik PO, Akl EA, da Costa BR, Adhikari NK, Dalmau GM, Kosunen E, Mustonen J, Crawford MW, Thabane L, Guyatt GH. Do clinicians understand the size of treatment effects? A randomized survey across 8 countries. CMAJ 2016;188:25–32. [PubMed: 26504102]

63. Jonville-Béra A, Giraudeau B, Autret-Leca E. Reporting of drug tolerance in randomized clinical trials: When data conflict with authors' conclusions. Annals of Internal Medicine 2006;144:306–307.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

64. Kalso E, Edwards JE, Moore RA, McQuay HJ. Opioids in chronic non-cancer pain: systematic review of efficacy and safety. Pain 2004;112:372–80. [PubMed: 15561393]

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

66. Katz N, Paillard FC, Van Inwegen R. A review of the use of the number needed to treat to evaluate the efficacy of analgesics. J Pain 2015;16:116–23. [PubMed: 25419989]

67. Katz NP. The measurement of symptoms and side effects in clinical trials of chronic pain. Contemp Clin Trials 2012;33:903–11. [PubMed: 22561389]

68. Kelley JM, Kaptchuk TJ. Group analysis versus individual response: the inferential limits of randomized controlled trials. Contemp Clin Trials 2010;31:423–8. [PubMed: 20624483]

69. Knottnerus JA, Tugwell P. The way in which effects are analyzed and communicated can make a difference for decision making. J Clin Epidemiol 2016;72:1–3. [PubMed: 26946104]

70. Knottnerus JA, Tugwell P. We must further reduce the room-for-improvement gap in producing, reporting and summarizing clinical evidence for better care. J Clin Epidemiol 2016;74:1–3. [PubMed: 27296837]

71. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. Biol Psychiatry 2006;59:990–6. [PubMed: 16368078]

72. Lancet. Information for Authors. 2020 1/17/2020]; Available from: https://marlin-prod.literatumonline.com/pb-assets/Lancet/authors/tl-info-for-authors.pdf.

73. Laska EM, Siegel C, Sunshine A. Onset and duration: measurement and analysis. Clin Pharmacol Ther 1991;49:1–5. [PubMed: 1988234]

74. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988;318:1728–33. [PubMed: 3374545]

75. Mallinckrodt C, Chuang-Stein C, McSorley P, Schwartz J, Archibald DG, Perahia DG, Detke MJ, Alphs L. A case study comparing a randomized withdrawal trial and a double-blind long-term trial for assessing the long-term efficacy of an antidepressant. Pharm Stat 2007;6:9–22. [PubMed: 17238129]

76. Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. Pharmacoeconomics 2014;32:345–65. [PubMed: 24504851]

77. Martel MO, Finan PH, Dolman AJ, Subramanian S, Edwards RR, Wasan AD, Jamison RN. Self-reports of medication side effects and pain-related activity interference in patients with chronic pain: a longitudinal cohort study. Pain 2015;156:1092–100. [PubMed: 25782367]

78. Mascha EJ. Equivalence and noninferiority testing in anesthesiology research. Anesthesiology 2010;113:779–81. [PubMed: 20808211]

79. McAlister FA. The “number needed to treat” turns 20--and continues to be used and misused. Cmaj 2008;179:549–53. [PubMed: 18779528]

80. McGauran N, Wieseler B, Kreis J, Schüler Y-B, Kölsch H, Kaiser T. Reporting bias in medical research - a narrative review. Trials 2010;11:37–37. [PubMed: 20388211]

81. McIntosh HM, Woolacott NF, Bagnall A-M. Assessing harmful effects in systematic Reviews. BMC Medical Research Methodology 2004;4:19. [PubMed: 15260887]

82. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. Ann Intern Med 1997;126:712–20. [PubMed: 9139558]

83. Moore RA, McQuay HJ. Prevalence of opioid adverse events in chronic non-malignant pain: systematic review of randomised trials of oral opioids. Arthritis Res Ther 2005;7:R1046–51. [PubMed: 16207320]

84. 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]

85. Mt-Isa S, Hallgreen CE, Wang N, Callreus T, Genov G, Hirsch I, Hobbiger SF, Hockley KS, Luciani D, Phillips LD, Quartey G, Sarac SB, Stoeckert I, Tzoulaki I, Micaleff A, Ashby D, participants I-Pb-r. Balancing benefit and risk of medicines: a systematic review and classification of available methodologies. Pharmacoepidemiol Drug Saf 2014;23:667–78. [PubMed: 24821575]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

86. Mt-Isa S, Ouwens M, Robert V, Gebel M, Schacht A, Hirsch I. Structured Benefit-risk assessment: a review of key publications and initiatives on frameworks and methodologies. Pharm Stat 2016;15:324–32. [PubMed: 25981683]

87. National. The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: The National Academies Press, 2010.

88. Nutt DJ, King LA, Phillips LD, Independent Scientific Committee on D. Drug harms in the UK: a multicriteria decision analysis. Lancet 2010;376:1558–65. [PubMed: 21036393]

89. Olsen MF, Bjerre E, Hansen MD, Hilden J, Landler NE, Tendal B, Hrobjartsson A. Pain relief that matters to patients: systematic review of empirical studies assessing the minimum clinically important difference in acute pain. BMC Med 2017;15:35. [PubMed: 28215182]

90. Olsen MF, Bjerre E, Hansen MD, Tendal B, Hilden J, Hrobjartsson A. Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors: systematic review of empirical studies. J Clin Epidemiol 2018;101:87–106.e2. [PubMed: 29793007]

91. Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, Holmes J, Mander AP, Odondi Lo, Sydes MR, Villar SS, Wason JMS, Weir CJ, Wheeler GM, Yap C, Jaki T. Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Medicine 2018;16:29. [PubMed: 29490655]

92. Piantadosi S. Clinical trials: a methodologic perspective. Hoboken, NJ: John Wiley & Sons, Inc., 2005.

93. Pocock SJ, Stone GW. The Primary Outcome Fails - What Next? N Engl J Med 2016;375:861–70. [PubMed: 27579636]

94. Pocock SJ, Stone GW. The Primary Outcome Is Positive - Is That Good Enough? N Engl J Med 2016;375:971–9. [PubMed: 27602669]

95. Powers JH, Fleming TR. Noninferiority trials: clinical understandings and misunderstandings. Clin Investig (Lond) 2013;3:215–218.

96. Raskin P, Huffman C, Toth C, Asmus MJ, Messig M, Sanchez RJ, Pauer L. Pregabalin in patients with inadequately treated painful diabetic peripheral neuropathy: a randomized withdrawal trial. Clin J Pain 2014;30:379–90. [PubMed: 23887339]

97. Rowbotham MC, Gilron I, Glazer C, Rice AS, Smith BH, Stewart WF, Wasan AD. Can pragmatic trials help us better understand chronic pain and improve treatment? Pain 2013;154:643–6. [PubMed: 23541132]

98. Ruyssen-Witrand A, Tubach F, Ravaud P. Systematic review reveals heterogeneity in definition of a clinically relevant difference in pain. J Clin Epidemiol 2011;64:463–70. [PubMed: 21109400]

99. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332. [PubMed: 20332509]

100. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis 1967;20:637–48. [PubMed: 4860352]

101. Senn S. Individual response to treatment: is it a valid assumption? BMJ 2004;329:966–8. [PubMed: 15499115]

102. Senn S. An unreasonable prejudice against modelling? Pharmaceutical Statistics 2005;4:87–89.

103. Senn S. Statistical issues in drug development. Hoboken, NJ: John Wiley & Sons, Ltd, 2007.

104. Senn S. Being Efficient About Efficacy Estimation. Statistics in Biopharmaceutical Research 2013;5:204–210.

105. Senn S. Mastering variation: variance components and personalised medicine. Statistics in medicine 2016;35:966–977. [PubMed: 26415869]

106. Senn S. Statistical pitfalls of personalized medicine. Nature 2018;563:619–621. [PubMed: 30482931]

107. 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]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

108. Snapinn SM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. Trials 2007;8:31. [PubMed: 17961249]

109. Strand V, Boers M, Idzerda L, Kirwan JR, Kvien TK, Tugwell PS, Dougados M. It's good to feel better but it's better to feel good and even better to feel good as soon as possible for as long as possible. Response criteria and the importance of change at OMERACT 10. J Rheumatol 2011;38:1720–7. [PubMed: 21807792]

110. Sudhop T, Brun NC, Riedel C, Rosso A, Broich K, Senderovitz T. Master protocols in clinical trials: a universal Swiss Army knife? The Lancet Oncology 2019;20:e336–e342. [PubMed: 31162107]

111. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, Tunis S, Bergel E, Harvey I, Magid DJ, Chalkidou K. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol 2009;62:464–75. [PubMed: 19348971]

112. Tsang R, Colley L, Lynd LD. Inadequate statistical power to detect clinically significant differences in adverse event rates in randomized controlled trials. Journal of Clinical Epidemiology 2009;62:609–616. [PubMed: 19013761]

113. 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]

114. United States Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009; Available from: http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf, accessed 12/6/19.

115. United States Food and Drug Administration. Guidance for Industry and Investigators: Safety Reporting Requirements for INDs and BA/BE Studies. 2012; Available from: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM227351.pdf, accessed 12/6/19.

116. United States Food and Drug Administration. Benefit-risk assessment in drug regulatory decision-making: draft PDUFA VI implementation plan (FY 2018–2022). 2018; Available from: https://www.fda.gov/media/112570/download, accessed 10/24/2019.

117. United States Food and Drug Administration. Adaptive designs for clinical trials of drugs and biologics: guidance for industry. 2019; Available from: https://www.fda.gov/media/78495/download, accessed 1/31/20.

118. United States Food and Drug Administration. Enrichment strategies for clinical trials to support determination of effectiveness of human drugs and biological products: guidance for industry. 2019; Available from: https://www.fda.gov/media/121320/download, accessed 1/31/20.

119. Vase L, Wartolowska K. Pain, placebo, and test of treatment efficacy: a narrative review. British Journal of Anaesthesia 2019;123:e254–e262. [PubMed: 30915982]

120. Wasan AD. Efficacy vs Effectiveness and Explanatory vs Pragmatic: Where Is the Balance Point in Pain Medicine Research? Pain Medicine 2014;15:539–540. [PubMed: 24716587]

121. Wasserstein RL, Lazar NA. The ASA Statement on p-Values: Context, Process, and Purpose. The American Statistician 2016;70:129–133.

122. Wasserstein RL, Schirm AL, Lazar NA. Moving to a World Beyond "$p < 0.05$". The American Statistician 2019;73:1–19.

123. West CP, Ficalora RD. Clinician attitudes toward biostatistics. Mayo Clin Proc 2007;82:939–43. [PubMed: 17673062]

124. Williams MR, McKeown A, Pressman Z, Hunsinger M, Lee K, Coplan P, Gilron I, Katz NP, McDermott MP, Raja SN, Rappaport BA, Rowbotham MC, Turk DC, Dworkin RH, Smith SM. Adverse Event Reporting in Clinical Trials of Intravenous and Invasive Pain Treatments: An ACTTION Systematic Review. J Pain 2016;17:1137–1149. [PubMed: 27522950]

125. Woodcock J, LaVange LM. Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. New England Journal of Medicine 2017;377:62–70. [PubMed: 28679092]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

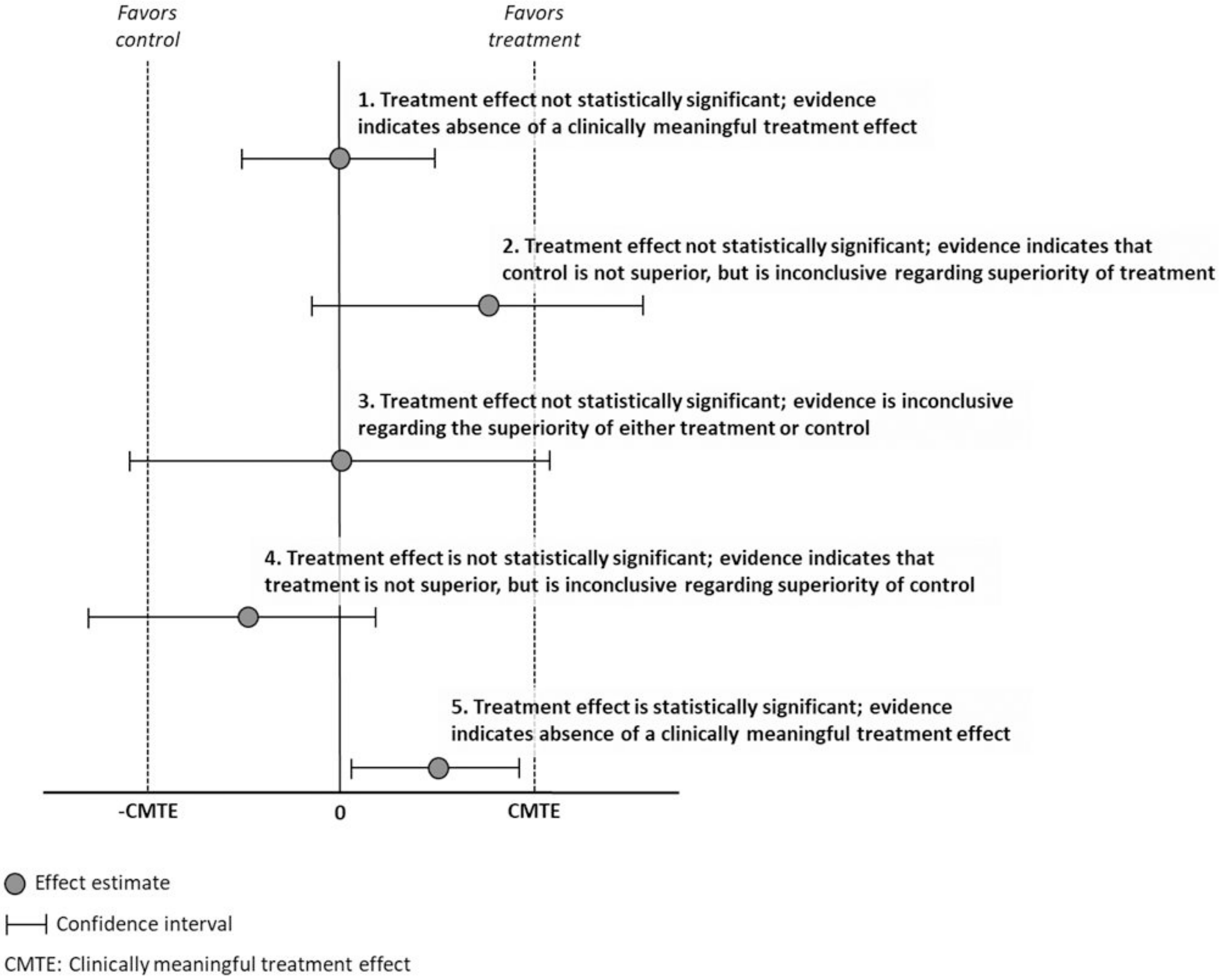
Favors control
Favors treatment
1. Treatment effect not statistically significant; evidence indicates absence of a clinically meaningful treatment effect
2. Treatment effect not statistically significant; evidence indicates that control is not superior, but is inconclusive regarding superiority of treatment
3. Treatment effect not statistically significant; evidence is inconclusive regarding the superiority of either treatment or control
4. Treatment effect is not statistically significant; evidence indicates that treatment is not superior, but is inconclusive regarding superiority of control
5. Treatment effect is statistically significant; evidence indicates absence of a clinically meaningful treatment effect
-CMTE
0
CMTE
● Effect estimate
├─┤ Confidence interval
CMTE: Clinically meaningful treatment effect

Figure 1.

Interpretation of confidence intervals

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

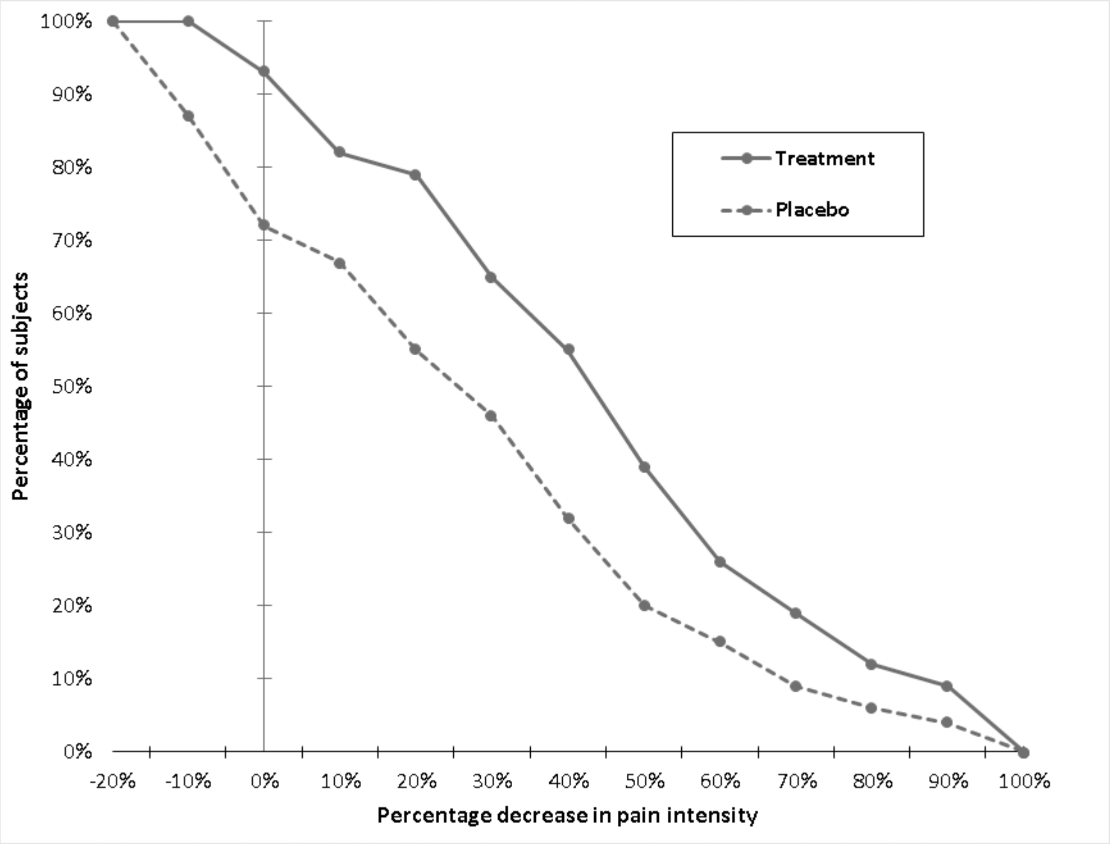

Figure 2.

Continuous distribution function (CDF) example

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Key terms

Term	Explanation and example
Absolute risk reduction (ARR)	The treatment group difference in the percentages of participants experiencing an event (e.g., the difference in the percentages of "responders" between the treatment and placebo arms).
Analysis of means	Statistically analyzing the between-group difference in mean outcome.
Between-group difference	The difference in mean outcome between two groups. Example: The mean of the change between baseline and week 12 on a 0-10 NRS for participants in the active treatment group minus the same quantity for participants in the placebo group.
Between-group minimal clinically meaningful difference	The between-group difference in an RCT (i.e., amount of additional pain reduction in the treatment group beyond that observed in the comparator group) that is meaningful to patients or other stakeholders. There is no universally accepted difference between treatment and comparator that is considered to be clinically important.
Confidence intervals (CIs)	At a given level of confidence (e.g., 95%), the range of possible values that are expected to contain the true treatment effect. For example, if the RCT were replicated a large number of times, 95% of the 95% CIs from the RCTs would contain the true treatment effect
Cumulative distribution function (CDF)	Plots of the percentage of "responders" in each study arm across the range of possible responses (see Figure 2).
Duration of effect	The length of the treatment benefit.
Explanatory trials	Trials designed to test whether the treatment is efficacious in more highly controlled settings (e.g., in a relatively homogeneous population).
Number needed to harm (NNH)	Identical to NNT, except NNH evaluates percentages of patients with harms.
Number needed to treat (NNT)	The reciprocal of the treatment group difference in the percentages of participants experiencing an event, calculated as 1/ARR. This number can be used, for example, to indicate the number of patients who would need to be treated to find 1 more "responder" in the treatment arm than in the comparator arm.
Power	Probability of rejecting the null hypothesis of no treatment effect when the treatment actually has an effect of a specified magnitude; calculated as 1 – Prob(Type II error).
Pragmatic trials	Trials designed to test whether a treatment that has been shown to have analgesic efficacy is effective in more real-world settings (e.g., in a heterogeneous population, concomitant medications allowed).
Primary outcome	The prespecified measure on which the effect of the treatment is being evaluated.
Relative risk (RR)	The ratio of the participants experiencing an event in the treatment arm to that in the placebo arm (e.g., the percentage of "responders" in the treatment arm divided by the percentage of "responders" in the placebo arm).
"Responder" analysis	A comparison between treatment groups of the percentage of "responders" (i.e., the individuals who have had a certain percentage improvement in pain intensity from baseline to end-of-study).
Treatment risks	All adverse events (AEs) associated with a treatment as identified by subject symptom reports and clinician-observed signs.
Type I error	Probability of rejecting the null hypothesis of no treatment effect (e.g., no treatment group difference in outcome) when the treatment actually has no effect; typically set at $\alpha = 0.05$.
Type II error	Probability of failing to reject the null hypothesis of no treatment effect when the treatment actually has an effect of a specified magnitude; typically set at $\beta = 0.10 – 0.20$.
Within-group difference	The mean change within one treatment group between baseline and a defined follow-up time period. Example: The mean of the change between baseline and week 12 on a 0-10 NRS for participants in the placebo group.
Within-patient minimal clinically meaningful change	The within-person change in pain intensity that is meaningful to the individual. Typically considered to be ≥ 10-20% improvement on the 0-10 NRS, with > 30% improvement on the 0-10 NRS considered moderate improvement, though baseline levels of pain can affect this percentage.

Notes: NRS – numerical rating scale; RCT – randomized clinical trial

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Major factors to consider in determining the clinical importance of group differences (adapted from Dworkin et al. [23])

- Statistical significance of the primary efficacy analysis (typically necessary but not sufficient to determine that the group difference is clinically meaningful)
- Availability of alternative therapies and their benefit-risk profiles
- Treatment effect size for the primary outcome variable compared to that of available treatments
- Safety and tolerability
- Rapidity of onset of treatment effect
- Durability of treatment effect
- Results for secondary efficacy endpoints (e.g., improvements in physical or emotional functioning)
- Limitations of available treatments
- Different mechanism of action vs. existing treatments
- Cost, convenience, and patient adherence
- Other benefits (e.g., few or no drug interactions, availability of a test that predicts a good therapeutic response)

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 3

Innovative clinical trial designs that can be considered in studies of pain treatments

Multi-period cross-over trials – *Cross-over trials with multiple periods (e.g., 2 active treatment periods, 2 placebo or comparator periods) allow for determination of the extent to which the effect of a treatment relative to placebo varies among patients [21]*

Enrichment clinical trials – *Clinical trials in which patients are selected based on a given characteristic that is expected to increase the likelihood of detecting a treatment effect [118], for example:*

- Enriched Enrollment Randomized Withdrawal (EERW) – *Clinical trials in which participants are initially administered the treatment of interest and those who reach a threshold of improvement and/or tolerability are then randomized to either remain on treatment or be withdrawn from treatment [65, 84]*

Designs that might reduce placebo group improvement and increase assay sensitivity

- Sequential Parallel Comparison Design (SPCD) – *Clinical trial design in which participants are randomized to active treatment or placebo, and then participants who do not "respond" to placebo are re-randomized to active treatment or placebo [37]*
- Two-way Enriched Design (TED) – *Clinical trial design in which participants are randomized to active treatment or placebo, and then participants who do not respond" to placebo are re-randomized to active treatment or placebo and participants who "respond" to treatment are re-randomized to active treatment or placebo [59]*

Adaptive designs – *Clinical trial designs that prospectively plan for modifications to the design based on the available evidence from the trial obtained at interim analyses without compromising the integrity or validity of the trial [117], for example:*

- Interim sample size re-estimation – *Most commonly practiced as estimating nuisance parameters (e.g., standard deviation) while the study is ongoing in order to determine whether the assumptions underlying the original sample size calculation are reasonable and, if not, increasing the sample size as needed [91]*
- Interim efficacy/futility analyses – *Comparing treatment arms while the study is ongoing to determine whether there is overwhelming evidence supporting either the efficacy or futility of the active treatment based on pre-specified stopping rules [91]*

Master protocol – A protocol containing multiple sub-studies that examine combinations of treatments, patient types, or diseases to increase the efficiency of drug development [110, 125]:

- Basket – *Study of 1 treatment in multiple conditions [110, 125]*
- Umbrella – *Study of a predetermined set of multiple treatments in 1 condition [110, 122]*
- Platform – *Perpetually adding treatments into an umbrella trial [110, 117]*

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4

Considerations for reporting and interpreting efficacy outcomes from chronic pain clinical trials

Analysis of means

- Requires specific pre-specified outcome variable at a specific time period
- Identifying the benefit of a treatment beyond that of the comparator requires statistically comparing treatment groups (e.g., active treatment vs. placebo)
- Tables or figures reporting a week-by-week analysis of means data may provide insight into the onset and durability of treatment response
- If figures are presented for secondary outcomes, they should not detract from the presentation of the primary endpoint analyses
- Sample size and amount of missing data should be reported, and appropriate statistical methods should be used to account for missing data along with reporting the assumptions underlying those methods

Responder analyses

- Requires understanding of a clinically meaningful threshold of change
- Those with a response meeting the pre-established threshold should not be considered a ‘responder’ unless there are data on repeated exposure to both active treatment and placebo at the individual level
- Dichotomizes continuous data into categorical data and, hence, reduces information
- Typically requires larger sample sizes to have sufficient power to detect a statistically significant difference between the treatment and comparator arms
- May be valuable as a secondary analysis to assist in interpreting RCT results

Number needed to treat (NNT), number needed to harm (NNH), relative risk (RR)

- See considerations for responder analyses above
- NNTs represent the number needed to be treated to achieve 1 additional “responder” in the treatment arm beyond the number of “responders” in the comparator arm (frequently misinterpreted to represent the number needed to treat to achieve 1 “responder”)
- RRs may be easier to interpret than NNTs or NNHs – represents the percentage of “responders” in the treatment arm divided by the percentage of responders” in the placebo arm
- Patient characteristics must be considered (i.e., NNTs, NNHs, RRs are specific to the sample used to generate them and may not generalize)
- RRs are frequently interpreted as an exaggerated risk of benefit when presented without the absolute risk.

Cumulative distribution functions (CDFs)

- Continuous plot of percentages of participants in each treatment arm who experience a particular change in pain intensity across the range of possible change
- Visually reflects the difference between the treatments, and may be valuable alongside analysis of means
- CDFs should be complete, not truncated, and properly account for missing data

Within-patient minimally clinically important difference

- Reflects a within-patient change (i.e., amount of change required from study baseline to end-of-study that is considered to be clinically important), not a between-group difference

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

- May not equate to well-managed pain
- Determined empirically with patient input

Time to effect and duration of effect

- Data on time to effect and duration of effect may contribute to a more complete understanding of the overall treatment effect
- Definitions of a “beneficial effect” vary, and should be clearly stated
- Various methods exist to report time to effect and duration of effect
- Research typically does not examine long-term treatment efficacy, despite extended use in the treatment of chronic pain conditions