ACTTION Special Issue on Clinical Trials of Pain Treatments
Review
OPEN
PAIN REPORTS®

Design and conduct of confirmatory chronic pain clinical trials

Nathaniel Katz

Abstract:
The purpose of this article is to provide readers with a basis for understanding the emerging science of clinical trials and to provide a set of practical, evidence-based suggestions for designing and executing confirmatory clinical trials in a manner that minimizes measurement error. The most important step in creating a mindset of quality clinical research is to abandon the antiquated concept that clinical trials are a method for capturing data from clinical practice and shifting to a concept of the clinical trial as a measurement system, consisting of an interconnected set of processes, each of which must be in calibration for the trial to generate an accurate and reliable estimate of the efficacy (and safety) of a given treatment. The status quo of inaccurate, unreliable, and protracted clinical trials is unacceptable and unsustainable. This article gathers aspects of study design and conduct under a single broad umbrella of techniques available to improve the accuracy and reliability of confirmatory clinical trials across traditional domain boundaries.

Keywords: Chronic pain, Clinical trial design and conduct, Measurement error, Confirmatory trials

Every physiological experiment should be performed under such circumstances as will secure a due observation and attestation of its results, and so obviate, as much as possible, the necessity for its repetition

—Marshall Hall (1790-–1857)

There is a peculiar paradox that exists in trial execution—we perform clinical trials to generate evidence to improve patient outcomes; however, we conduct clinical trials like anecdotal medicine: (1) we do what we think works; (2) we rely on experience and judgment; and (3) have limited data to support best practices.

—Monica Shah

1. Introduction

In considering "confirmatory" clinical trials, as opposed to earlier clinical trials where the purpose is "learning,"[216] it is tempting to imagine that everything is already known about the study treatment before the trial begins and that the confirmatory trial is some type of formality. In reality, confirmatory studies are often the first studies with a sufficient sample size to adequately characterize the performance of a treatment and the first to do so in a more heterogeneous multicenter population that strenuously challenges efforts to minimize measurement error. In the area of drug studies, confirming efficacy observed in earlier studies is no simple matter: over half of phase 3 failures across therapeutic areas are due to failure to confirm efficacy, despite the fact that many phase 3 studies evaluate reformulations of drugs already known to be efficacious.[114] Performing a robust clinical trial requires an understanding of the factors that determine whether a trial will succeed or fail to demonstrate the truth about an investigational treatment, whether drug, device, behavioral treatment, or other treatment type.

There are important differences between learning (phase 2) and confirming (phase 3) studies. Phase 2 trials embody a tension: the purpose is to determine whether the treatment relieves pain, as intended, compared with a control condition, which is best accomplished under conditions that minimize sources of variability to the extent possible. Another purpose is to prepare for phase 3, where in general conditions are more heterogeneous; thus, there is value in phase 2 in examining the impact of sources of variability (eg, baseline participant characteristics, enrollment criteria, endpoints, analysis methods, and covariates) on outcome. Approaches to accomplishing these goals in phase 2 are reviewed in this series by Campbell et al.,[33] as well as in a previous IMMPACT review[97]; this article will be oriented towards principles of study design and conduct in phase 3 trials, although as indicated these principles are applicable to all randomized controlled trials (RCTs).

This chapter is meant to address confirmatory studies of any type of treatment, whether conducted for regulatory purposes or

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

Corresponding author. Address: WCG Analgesic Solutions, 321 Commonwealth Rd, Suite 204, Wayland, MA 01778. Tel.: 781-444-9605. E-mail address: nkatz@analgesicsolutions.com (N. Katz).

PR9 6 (2021) e854

http://dx.doi.org/10.1097/PR9.0000000000000854

not. The terms phase 2 and phase 3 are applied primarily to trials of drug treatments performed for regulatory purposes; however, in this chapter, when these terms are used, the intended application is to learning and confirming trials, respectively, regardless of the regulatory intent.

In the early days of the RCT, the main concern of trialists was *false-positive* results—that is, the finding that a treatment is effective when in reality it is not. Therefore, an emphasis was placed on understanding and addressing biases that cause false-positive trials. More recently, the increasing rate of trial failure has led to the examination of *false-negative* results: the treatment is effective, but the trial fails to show it. These concerns have led to the emergence of a new clinical trial science focused on characterizing and optimizing the factors that allow trials to accurately measure the efficacy (and safety) of new therapeutics, which essentially amounts to minimizing measurement error.[62] The relative importance of these goals can be debated. False-positive studies may lead clinicians to expose patients to risks and society to costs, without a compensatory benefit. On the other hand, millions of individuals are already suffering from chronic pain, which has not responded to available treatments, and delays or cessation of the development of improved therapeutics because of a falsely negative trial perpetuates their suffering. Importantly, improving assay sensitivity alone runs the risk of detecting small signals of efficacy that will not meaningfully change patients' lives, which may be compounded by delays in detecting important safety issues before large-scale use. The science of clinical trials thus must achieve goals on several fronts that may compete with each other: avoiding false conclusions of efficacy, avoiding missing an efficacy signal, and advancing therapies whose magnitude of benefit is meaningful.

The purpose of this article is to provide readers with a basis for understanding the emerging science of clinical trials and to provide a set of practical, evidence-based suggestions for designing and executing clinical trials in a manner that minimizes measurement error. Although the evidence base supporting these advances is constantly expanding, studies evaluating the impact of study design or conduct methods on the outcome of clinical trials are still scant.[62] Therefore, common sense approaches to reducing measurement error will be provided for the reader's consideration even when evidence for their impact is limited. Moreover, the boundary between science and operations, regulation, budgeting, and other seemingly prosaic domains is blurry. This article deliberately gathers aspects of study design and conduct under a single broad umbrella of techniques available to improve the accuracy and reliability of clinical trials even when these techniques cross traditional domain boundaries.

For background and inspiration, readers are referred to the classic monograph on analgesic clinical trials by Max, Portenoy, and Laska[164] as well as key IMMPACT papers on this topic.[63,64]

2. Principles of experimental design relating to confirmatory studies

2.1. Clinical trials and measurement

The purpose of a clinical trial is to measure some attribute of a treatment, most commonly efficacy (and always, safety). Thus, one can think of a clinical trial as a measurement instrument, like a weighing scale or pH meter. The goal of study design and conduct is to measure the treatment attribute as accurately and reliably as possible, so that the observed result is neither exaggerated nor diminished in comparison with the true magnitude of that attribute. As such, it may be useful to examine what is known about measurement in other areas of science and engineering and to determine whether insights from those areas could inform how we conceptualize, describe, design, and conduct clinical trials. In the context of confirmatory clinical trials, the most common design is the randomized controlled clinical trial, where the measure of efficacy is the magnitude of benefit of the study treatment compared with controls.[168] In the engineering arena, a measurement instrument, such as a weighing scale, also produces estimates of some attribute of that which is being measured. Accurate and reliable output from a measurement instrument requires that each component of the instrument be calibrated. In the case of a clinical trial where the primary endpoint is subjective (eg, pain intensity), the components include human beings—one approach to optimizing clinical trial design is to consider how these humans can be calibrated so that as a whole the clinical trial produces reliable results. Thus, one can consider measurement performance at the level of the entire clinical trial (Is the measurement of the treatment effect accurate?), or at the level of the individual components of the trial, which must be accurate for the results of the entire trial to be accurate (Is that diagnostic assessment being applied accurately? Is outcome being measured accurately?). Importantly, accuracy of measurement of a subjective state such as pain depends on the performance of the measurement instrument (eg, a pain intensity scale) and of the person using the instrument (eg, the patient, in the context of patient-reported outcome measures, or the clinician, in the context of clinician-reported measures). These concepts are developed further below.

2.2. Terminology

Although there have been extensive efforts to define and harmonize terminology related to the science of measurement (metrology) across the physical and biological sciences, these efforts have not extended systematically to clinical trials. The following lexicon (**Table 1**) is based on the international consensus of the Joint Committee for Guides in Metrology and the ISO 5725 standard on measurement principles and definitions.[122,126] *Measurement* is the assignment of a number to a characteristic of an object or event of interest; in our case, this characteristic may be the efficacy or safety of a treatment. A *measurement system* is a set of one or more measuring instruments and other devices assembled to generate measured quantity values. A clinical trial is a measurement system, consisting of many individual measurement activities conducted within a well-defined experiment. Measurement terminology can be applied to a *measurement method* (an overall clinical trial methodology or a specific assessment performed in such a trial) or the *result* of a measurement method (eg, the observed treatment effect in a clinical trial or the result of a specific assessment). The general term *accuracy* covers 2 related concepts: *precision* (often called *reliability*) and *trueness* (often, confusingly, called *accuracy*). When applied to a measurement method, *precision (reliability)* refers to the closeness of agreement between the results of replicate measurements of the same or similar objects under specified conditions.[126,153] Readers are likely to be familiar with test–retest reliability evaluations of specific clinical outcomes assessments; the same reasoning can be applied to the evaluation of test–retest reliability of a method of conducting a clinical trial, such as a third molar extraction model. Dispersion of repeated measures from each other is considered *random error*. The term precision does not apply to the results of a single measurement because by definition it relates to repeated measurements.

Table 1

Terminology related to measurement as applied to clinical trials.

Term	Definition
Measurement	The assignment of a number to a characteristic of an object or event of interest; in clinical trials, this characteristic may be the efficacy or safety of a treatment.
Measurement system	A set of one or more measuring instruments and other devices assembled to generate measured quantity values. A clinical trial is a measurement system, consisting of many individual measurement activities conducted within a well-defined experiment.
Measurement method	A specific set of equipment or procedures designed to produce measurements. This may refer to a type of measurement (eg, inflatable cuffs to measure blood pressure), a clinical trial method (eg, third molar extraction pain model), or the measurements executed by a specific laboratory.
Measurement result	The measured quantity value generated by a specific measurement event
Accuracy	*When applied to a measurement method:* The closeness of the average of a large number of measurements to the true value or an accepted standard (if available) *When applied to a single measurement result:* The closeness of a single measurement result to the true value or an accepted standard (if known) Note that the term accuracy is also used as a general term for measurement performance that covers both accuracy (as defined above) and *reliability* (see below)
Reliability (precision)	*When applied to a measurement method:* The closeness of the results of multiple measurements to each other *When applied to a single measurement result:* Not applicable Note that while the term reliability applies only to repeated measurements and therefore cannot apply to the results of a single clinical trial, the reliability of the critical activities within a clinical trial that determine whether the ultimate result is accurate, such as reliability of a diagnostic procedure or outcome assessment, can be measured.
Measurement error	The opposite of measurement accuracy (taken in its overall sense): the deviation of a measurement or set of measurements from the true value, if known
Assay sensitivity	The ability of a type of experiment, or an individual experiment, to discriminate groups that are known to be different (eg, active drug vs placebo). Assay sensitivity requires both accuracy and reliability of the measurement method and is often used as a proxy for these measurement characteristics, which may not be directly measurable.
False-negative trial	A clinical trial that fails to demonstrate the efficacy of a truly efficacious treatment
False-positive trial	A clinical trial that demonstrates the efficacy of a treatment that is truly not efficacious

This lexicon is based on the international consensus of the Joint Committee for Guides in Metrology (JCGM) and the ISO 5725 standard on measurement principles and definitions.[120,122]

When applied to a measurement method, the term *trueness (accuracy)* refers to the closeness of agreement between the average of a large number of repeated measurements and the true value of what is being measured or an accepted reference value (if available).[122,126,153] The related concept in psychometrics is *validity*, which is generally defined as the extent to which a specific outcome assessment instrument measures what it purports to measure, often determined by a de facto gold standard.[78] Dispersion of a set of repeated measurements from the true value is considered *systematic error* or *bias* (although the term "bias" has other meanings in statistics, psychometrics, and legal or philosophical contexts).[117,122,126] When applied to a single result of a measurement, accuracy refers to the difference between the observed result and the true value (if known).

Measurement error is the opposite of measurement accuracy. For a set of results, measurement error refers to the sum of random and systematic error components. For a specific result, measurement error is the difference between the observed value and the true value or reference standard (if known). For an individual measurement to be accurate, or to differentiate quantities that are truly different, the measurement method must be both accurate (true) and precise (reliable).

In this article, the concept of the proximity of a set of measured results to the true value will be referred to as *accuracy* because it is more familiar to readers, although ISO prefers *trueness*. The proximity of a set of results to each other will be referred to as *reliability*, for the same reason, although ISO prefers *precision*. Loss of either accuracy or reliability will be referred to together under the umbrella term *measurement error*.

For the purpose of this article, a *failed* or *false-negative* study will refer to a study in which a treatment that is "known" to be effective does not statistically differentiate from an inert treatment, and a successful (*true-positive*) study is one that does demonstrate a statistically significant difference between a "known" effective treatment and an inert treatment. It must be acknowledged that any trial that reveals the "truth" about an investigational treatment, whether the treatment is effective or not, can be regarded as successful.

2.3. Measurement error, assay sensitivity, and the performance of clinical trials

In the physical sciences, an accepted "true value" of a measurand, such as the weight of an object, can be used to determine the accuracy of a measurement instrument, such as a weighing scale. In clinical trials, we do not have an easily grasped "true value" for the efficacy of an investigational treatment; therefore, the accuracy of the results of an individual trial can seldom be measured directly like that of a weighing scale. It can be difficult to discern whether variability of results from one trial to the next is due to true differences (eg, based on the study population, treatment context, disease-related issues, or other factors) or unreliability of measurement. Alternatively, one can imagine evaluating the reliability of a clinical trial as a measurement system by examining the results of multiple repetitions of the same trial type. Ideally, this would be done based on studies performed as identically as possible at highly specialized centers that regularly repeat stereotypical study designs (eg, dental pain centers); this type of data is generally not available.

Several methods can be used to gain insight into the accuracy and reliability of clinical trials. The first method is to evaluate, among trials attempting to answer the same question the same way (eg, same treatment, dose, population, and general protocol), how reliable are the results? One can explore this issue using meta-analyses where the individual studies are similar enough to be deemed combinable. A meta-analysis of acetaminophen for osteoarthritis of the knee, for example (**Fig. 1**), illustrates the wide range in observed results typical of repeated clinical trials of the same treatment for the same condition and includes positive trials, negative trials, and 1 trial where placebo was numerically superior to active treatment.[267] Quantifying the reliability of the results of such

studies as one would a measurement instrument in engineering yields a standard deviation of 0.09, a range of 0.26, an average deviation of 0.07, and a coefficient of variation of 69%. Thus, there are large differences in the results of similar clinical trials of the same treatment for the same disorder.

A second approach is to evaluate so-called replicate trials in which the same exact protocol is executed simultaneously by 2 different groups of investigators. This approach accounts for differences between one protocol and another, such that differences in results must be attributable to study conduct alone, rather than protocol design. Replicate trials may fail to produce similar results. For example, replicate trials were performed on lamotrigine for painful diabetic peripheral neuropathy (DPN), where the same exact protocol executed by 2 different groups of investigators led to widely divergent results.[256] Even the placebo arms of clinical trials in neuropathic pain differ widely in observed efficacy.[198]

A third approach is based on the concept that, for an individual clinical trial to discriminate 2 treatments that are known to be different (eg, ibuprofen and placebo for dental pain), the trial methodology must have been rigorous enough to produce accurate and reliable results. In other words, *discrimination* requires both *accuracy* and *reliability*. *Assay sensitivity*, the ability of a clinical trial to differentiate between an efficacious treatment and a control treatment, is thus a useful indicator that the methods of that trial were accurate and reliable because neither of these individual concepts can be easily measured directly in practice. A clinical trial that successfully discriminates between an effective treatment and a control condition is said to have demonstrated assay sensitivity.[65,163,237] Therefore, the main method for evaluating the measurement performance of a clinical trial is examining its ability to discriminate between a positive and negative control. To accomplish this, an active comparator of "known" efficacy can be added as a third treatment in a trial comparing an investigational treatment with a negative control.[237] If an investigational treatment fails to differentiate from placebo, and a known active comparator also fails, this suggests it was the study and not necessarily the drug that failed.[237] If the active comparator generates a much larger or smaller difference compared with placebo than what is normally observed, this aids in interpretation of the observed effect of the study treatment.

A fourth approach for evaluating the measurement performance of a clinical trial is to evaluate the performance of critical processes within the trial that have an impact on the accuracy of the final results. This requires a determination of what the key processes are and specification of a method for assessing their accuracy and reliability. This approach is equivalent to examining the gears or springs in a scale and determining whether they are functioning as intended. The power of this approach is that it provides actionable options to improve the rigor of individual clinical trials, rather than just lamenting about the reliability of the overall results. The remainder of this article will be largely devoted to enumerating these measurement components and providing methods for improving their accuracy and reliability.

Key measurement components of clinical trials include diagnostic assessments for inclusion into the trial or clinical outcome assessments (COAs).[78,153] These assessments are examples of the "components" of a clinical trial whose performance must perform accurately and reliably in order for the trial as a whole to produce accurate and reliable results. For example, the reliability of a COA has a direct impact on the sample size requirement to detect a specified treatment difference in a clinical trial (**Table 2**). In this case, a decrease in reliability from 1 (perfect) to 0.6 forces an increase in a sample size of 67%.[153] The accuracy of diagnostic inclusion criteria is also subject to measurement error and can influence the accuracy and reliability of the results of the clinical trial as a whole.[153] Multiple types of error may act synergistically to compromise overall trial results and statistical power. Thus, increasing attention has been directed to identifying and controlling sources of measurement error that undermine clinical trials.[31,50,64,130,147,153,189,223] It is also important to recognize the ethical requirement of performing clinical trials in a manner that generates the most accurate and reliable results

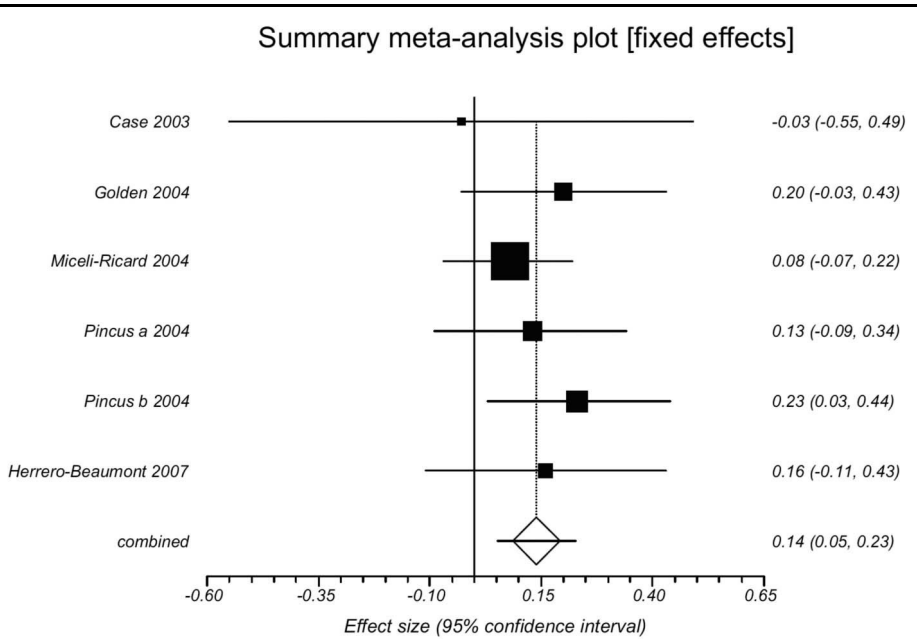

Figure 1. Meta-analysis of randomized controlled trials of acetaminophen for the treatment of hip and knee osteoarthritis: Observations range from placebo being numerically superior to acetaminophen to a standardized effect size of about 0.25, with most studies failing to show statistically significant superiority. The meta-analysis indicated that acetaminophen provides a standardized effect size of about 0.15.[267]

Table 2

Relationship between reliability of the primary outcome assessment in a clinical trial and statistical power and sample size requirements for a fixed detectable difference.

Reliability*	Power	Increase in N required for 80% power
1.0	80%	0%
0.9	76%	11%
0.8	71%	25%
0.7	65%	43%
0.6	58%	67%
0.5	51%	100%
0.4	43%	150%
0.3	34%	233%
0.2	24%	400%
0.1	14%	900%

As the reliability of the endpoint measure decreases from "perfect," statistical power decreases and sample size requirements increase.
Adapted from Muller and Szegedi.[174]
* Intraclass correlation coefficient, kappa.

because clinical trials are burdensome to patients and expose them to risks.[73]

3. Study framework: purpose, objectives, assessments, and endpoints

The study framework consists of a *purpose*, *objectives*, *assessments*, and *endpoints* and represents the skeleton of any trial. Many failed studies can be traced to defects in the study framework. Designing a successful clinical trial begins with defining the purpose of the study: in plain language, why are you doing the study? Potential purposes include: "to perform a required pivotal trial for regulatory submission," "to broaden a label from postherpetic neuralgia to pain associated with peripheral neuropathy," or "to generate data that will convince payers to pay for study treatment." Lack of consensus regarding the actual purpose of a trial may lead to regret years later, when key stakeholders discover that a beautifully designed and executed trial has not fulfilled their view of its original purpose.

The *objective* is the specific scientific aim of a study, which may be framed as a refutable hypothesis. Examples include: "to evaluate the effect of treatment vs placebo on pain intensity in patients with chronic nonradicular low back pain" or "to compare the effects of 2 different treatment regimens on signs and symptoms of osteoarthritis." An objective should not include the specific assessments or endpoints that will be used to achieve the objective; this comes later. Objectives can be specific or broad (eg, "effect on pain intensity" vs "effectiveness in patients with osteoarthritis). The advantage of specific objectives (eg, "to compare the effect on pain intensity…") is that it is easier to account for the relationship between each objective, assessment, and endpoint. In this example, an objective focused on pain intensity can be matched to a single assessment of pain intensity and an endpoint wrapped around that assessment instrument. The advantage of broader objectives ("to compare effectiveness in patients with osteoarthritis") is that multiple assessments and endpoints can be linked to a single objective, which makes the list of objectives more concise. In either case, the goal is to ensure a complete mapping of all objectives, assessments, and endpoints, with nothing left unattached.

An *outcome assessment* is a measurement instrument that generates a score intended to represent aspects of a patient's health status.[260] Outcome assessments fall into 2 categories: COAs are those that depend on someone's judgment, such as the patient, a clinician, or a caregiver. By contrast, a *biomarker* is a "defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions;" its measurement properties are minimally subject to human influence.[84] Biomarkers can in principle be used to measure treatment effects when they can be demonstrated to predict some meaningful clinical outcome but do not measure that clinical outcome directly. Biomarkers can, however, be useful in the development of therapeutics as measures of important attributes of the biological effect of the treatment, such as target engagement, which can be useful for development purposes even when falling short of being able to substitute for a COA as an outcome measure. In studies of pain treatments, the primary outcome assessment is almost always a COA because the goal of treatment is clinical benefit. Because the use of biomarkers has increased the success of therapeutic development across a range of indications,[114] there is a growing interest in the use of biomarkers in clinical trials of pain, particularly early trials. At present, no biomarkers have been validated as surrogate endpoints that can replace COAs as outcome measures for clinical trials of analgesics.

The terms endpoint and assessment have been used loosely and synonymously and are still used synonymously by many. This chapter will use the definition of endpoint from a recent FDA-NIH working group[84]:

> *Endpoint: A precisely defined variable intended to reflect an outcome of interest that is statistically analyzed to address a particular research question. A precise definition of an endpoint typically specifies the type of assessments made, the timing of those assessments, the assessment tools used, and possibly other details, as applicable, such as how multiple assessments within an individual are to be combined.*

For example, if the primary assessment is "average pain intensity in the past 24 hours on a 0 to 10 numerical rating scale (NRS)," an endpoint could be "the change from baseline in the average 24-hour pain NRS averaged over week 12 minus the average of the baseline week." Some would go farther and include the statistical analysis method in the definition of the endpoint.[260] Numerous nuances drive the exact composition of different elements of an endpoint; the important distinction drawn in this recent clarification of terminology is that the assessment is not the endpoint, but the endpoint contains the assessment comprehensively specified. In this article, "COA," "assessment," "measure," and "instrument" will be used synonymously.

The reader is cautioned against simply picking assessments from a list or previous trial; assessments are often propagated from one trial to another without examination of their measurement properties. New information about measures is continuously emerging. Therefore, a thorough and updated measure review should be performed to justify measure selection for every study. Principles governing COA selection are treated in detail in the article by Patel in this series.[186]

Readers should also be aware of the commonly used PICO framework in designing clinical trials and, more often, searching for and extracting data from published studies: patients/populations, interventions, comparators/controls, and outcomes/endpoints.[201]

4. Preventing false-positive studies: randomization, blinding, and placebo controls

4.1. Randomization

Allocation bias occurs when investigators manipulate the assignment of patients to study arms, usually assigning patients with better prognostic factors to the experimental arm. Allocation bias can occur in nonrandomized trials or in randomized trials when randomization is compromised. Dozens of studies have evaluated the impact of inadequate randomization on observed effect sizes. Initially, allocation bias was found to primarily inflate observed treatment effects,[213] presumably owing to patients with better prognostic factors being assigned to the more effective treatment. More recent studies have found that the impact of allocation bias is more complex and can either magnify or shrink observed treatment effects, with the greatest effects seen in studies with subjective outcomes such as pain.[182,211]

The main motives for investigators to manipulate treatment assignments are to ensure the best care for patients, to respect patient treatment preference, and to ensure the "best outcome" for the study.[186] The most common methods are tampering with randomization envelopes, deciphering the nature of blinded medication in treatment kits, and predicting future assignments based on past assignments. These observations suggest that envelopes should not be used to determine treatment assignments; successful masking of treatments should be documented; randomization blocks at sites should not be so small that investigators can guess the next treatment assignment; and size of randomization blocks should not be revealed to site personnel.

The most common randomization ratio is 1:1 active to control, as this ratio generates the greatest statistical power. The probability of assignment to an active treatment increases in several situations: asymmetric allocation ratios in 2-arm studies where a greater proportion of patients are assigned to active treatment; this is usually performed to generate more safety exposure data or to facilitate patient recruitment. Other situations include the addition of an active comparator to a standard 2-arm placebo-controlled study or the addition of multiple doses or regimens of study treatment. Asymmetric allocation ratios potentially create *expectation or observer bias*: if patients (or investigators) feel that they are more likely to receive active treatment, they may be biased to report (or observe) greater improvement or more adverse events. These effects augment the placebo response and thereby decrease the observed net treatment effect.[158,184,222,251] For this reason, some groups have cautioned against multiarm studies and asymmetric randomization ratios.[64] Yet, avoiding these design features when they are needed to fulfill study objectives for fear of expectation bias is the proverbial tail wagging the dog. An alternative approach is to design studies as necessary to accomplish the study objectives and use available techniques to manage expectation bias. These techniques include masking the randomization ratio[64] or using patient and staff training programs that focus on neutralizing expectation (described further below).[8,241,269]

Stratified randomization refers to the method of selecting a baseline participant characteristic thought to predict outcome (eg, high baseline pain intensity, sex, and pain phenotype), subgrouping patients based on a cutoff, and randomizing patients to treatment within each stratum. The purpose is to ensure balance between treatment groups on important covariates so that estimates of efficacy are attributable to the treatment assignment and not to imbalances in baseline prognostic factors. A large number of strata can create imbalances within strata and can impose an operational burden. Therefore, stratification should generally be limited to small trials in which treatment outcomes may be affected by known factors that have a large effect on prognosis and in trials when interim analyses are planned with small numbers of patients, and only a small number of strata should be used.[106,141]

4.2. Blinding and placebo controls

Expectation and *observer biases* can be conscious (eg, the patient consciously underreports their pain intensity because they want to please the investigator) or unconscious (eg, the participant actually perceives less pain because of the neural mechanisms triggered by the therapeutic context).[128] Expectations of the investigator can be transmitted to the participant consciously and unconsciously, as well as verbally and nonverbally.[5,105] The margins between observer bias and expectation bias are blurry because in the context of patient-reported outcome measures the patient is the observer, and observer bias on the part of the investigator can lead to expectation on the part of the patient.

Blinding has been the gold-standard method for limiting observer bias since the first double-blind placebo-controlled study.[113] Under *single blinding*, the investigator is aware of the treatment assignment, but the participant is not (or the reverse); under *double blinding*, neither the investigator nor the participant is aware of the treatment assignment. In studies in which the participant is not told the treatment assignment, but research staff in contact with the participants are aware of the treatment assignment, the expectation of the research staff has been well documented to be transmitted to the participant, with a major impact on treatment effects.[36,85,107] Therefore, single-blind studies can be viewed as essentially unblinded.

Blinding is a means to an end, but not an end in itself: the end is balance of patient and researcher expectations across treatment groups, and effective double blinding is a practical method for achieving this goal. However, the effectiveness of blinding for achieving this aim is almost always assumed and rarely evaluated. Moreover, alternative approaches to maintaining balanced expectations are seldom implemented, either in addition to or instead of blinding. Common situations in which alternatives to double blinding are implemented are medical device studies or studies of physical or psychological treatments, where it can be challenging to fully blind participants and staff. When double blinding is not feasible, study designers can choose from a variety of alternative methods to mitigate observer and expectation bias. One approach is to have an unblinded team perform the treatment procedure and a blinded third party perform all patient assessments. In these situations, a specific blinding plan and adherence to the plan should be documented for the overall study and at each clinical site. A second approach is to ensure that potential sources of bias, such as information about study treatments, ancillary support, and wording of the informed consent form, reinforce equipoise about which treatment is better.

The effectiveness of the chosen approaches can be evaluated by assessing participants' expectation of benefit at the beginning of the trial, and whether they know what treatment group they were in at the end of the trial (for blinded studies), and the reason for their guess (correct guesses due to efficacy should not be counted as unblinding).[67,87,88,149] Post hoc analyses of the relationship between factors that could have produced unblinding (eg, side effects) and efficacy can be performed. These procedures should also be used as appropriate in studies which cannot be blinded, such as certain behavioral treatments or

invasive procedures.[67,140] When blinding and its impact on outcome have been evaluated in RCTs, generally patients are not very good at guessing what treatment they are on, and correct guesses generally have not predicted the outcome. For example, in a crossover trial of dextromethorphan and memantine in neuropathic pain, participants' guesses were not significantly better than chance and did not predict outcome.[209] In another crossover study, comparing nortriptyline, gabapentin, and their combination in neuropathic pain, only a minority of patients guessed correctly.[102] Nonetheless, in specific cases, it is possible that functional unblinding could occur and bias responses.

5. Common study designs: strengths and weaknesses

5.1. Parallel group design

The gold standard for confirmatory studies is the classic prospective parallel group design (**Fig. 2A**).[63] Advantages of the parallel group design include simplicity; ease of analysis, interpretation, and communication of the results; comparability with other similarly designed studies; and familiarity to stakeholders. The basic active vs placebo design is easily augmented by adding additional doses/regimens, active comparators, or even combination treatment arms. Because each participant is only exposed to 1 treatment in this design, the duration of individual participant participation is shorter than in designs in which participants are exposed to sequential treatments of equivalent duration.

The main disadvantage of the prospective parallel design is statistical inefficiency: sample size requirements are larger than those needed for some alternative designs. This is a larger problem than it seems because large sample sizes may necessitate increasing the number of research sites, which in turn may add variability to the data, shrink observed effect sizes, and thereby decrease statistical power, subsequently necessitating an even larger sample size in a vicious cycle that may lead to study failure.[169] Large numbers of sites may also lead to global programs in a diversity of languages, cultures, healthcare systems, and research quality, which further undermine assay sensitivity. Although methods are available to address site variability (see below), these methods are in the early phases of adoption. Therefore, it is worthwhile to consider designs that have better statistical efficiency, especially for disorders in which patient recruitment is particularly challenging. An additional disadvantage of parallel designs, from the perspective of the patient who is interested in trying a new treatment, is that patients assigned to placebo may never get the opportunity to try the new treatment.

A treatment duration of 3 months has typically been accepted as a proxy for long-term use in studies of treatments for chronic pain[82]; longer-term studies are typically performed when indicated for specific treatments, such as 6-month studies for intra-articular injections in patients with knee osteoarthritis,[37] 16-week studies for antibodies that are administered every 8 weeks,[214] and longer durations for clinical trials of behavioral[67] or invasive[140] treatments where effectiveness may take longer to establish, assessment of loss of effectiveness over time is important, or safety issues may take longer to observe. Although it seems logical to study a long-term treatment over a long period of time, lengthy studies have considerable technical challenges such as adherence to treatment, avoidance of prohibited treatments, participant retention, documentation of extraneous care, and occurrence of unanticipated health events that contribute to missing data and confounding, ultimately limiting the interpretability of long-term data compared with higher quality, short-term studies. A variety of compromises can be considered, such as a tightly controlled 3-month treatment period (with strict prohibitions of activity that might compromise assay sensitivity and comprehensive documentation of these issues), followed by a less tightly controlled long-term phase that primarily focuses on safety and accepts compromises on assay sensitivity for efficacy endpoints.

5.2. Randomized withdrawal design

The randomized withdrawal design exists under several aliases, including the randomized discontinuation design and the

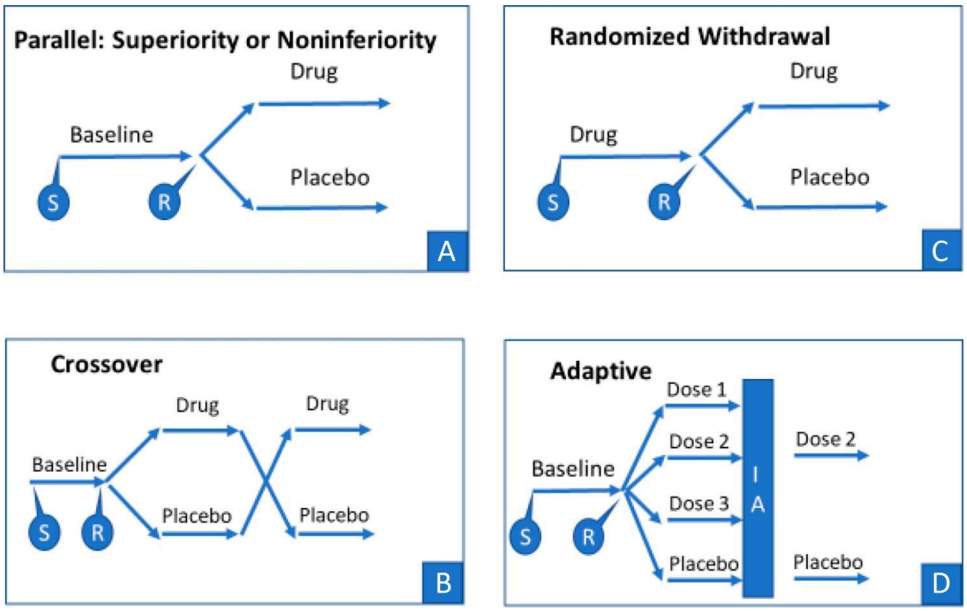

Figure 2. Schematics of common study designs. A. Parallel study (superiority or non-inferiority). B. Randomized withdrawal. C. Crossover (in this case a two-treatment, two-period design is shown). D. Adaptive design (in this case a dose-truncation option is shown, where after an interim analysis two doses are dropped). IA, interim analysis; R, randomization; S, screening.

enriched enrollment randomized withdrawal (EERW) design.[131,171] Typically, patients enter the study already receiving treatment (as part of clinical practice or a previous trial) or receive study treatment in a single-arm open-label fashion (**Fig. 2C**). Patients who do not benefit from or tolerate treatment are discontinued, enriching the population for potential responders (Patients can improve during the enrichment phase of an enriched enrollment trial for a variety of reasons other than the treatment they received, such as a placebo response or natural history of their disease. In this article, the term "responders" will be used to describe those who improve during enrichment recognizing this limitation.). Eligible participants are then randomized to either continued treatment or placebo, often with a blinded taper to prevent withdrawal when applicable. The primary endpoint can be pain intensity at a specified time point after randomization (eg, 12 weeks) or time to loss of pain control.[80] The open-label phase can accommodate flexible dosing. Flexible dosing is not specific to the EERW design; flexible dosing can be used in other types of designs; and EERW studies do not require flexible dosing. This is an important point because trialists may be drawn to the EERW design by the flexible dosing feature that is actually available in other designs.

One of the motivations for the randomized withdrawal design was to minimize the time that patients spend on ineffective or harmful treatments after randomization.[80,131] For example, patients stable on an antiepileptic or antihypertensive medication who enter a randomized withdrawal phase can be exited from the study as soon as they fail treatment and be placed back on their original treatment.[37] The primary endpoint in such a study would be time to exit. Because the exit due to treatment failure in these designs is informative, exiting does not create a problem of missing data for that endpoint (there may still be missing data when patients exit for other reasons and missing data relevant for other endpoints). Time to exit in pain studies has been shown to be a more statistically powerful endpoint in pain studies with an EERW design than differences in pain intensity at a fixed time point.[131] The randomized withdrawal design also addresses the problem of evaluating the long-term efficacy of treatments such as antidepressants or antineoplastic agents: instead of performing an impossible or unethical multiyear placebo-controlled RCT, investigators can recruit patients who have been on seemingly beneficial treatment for years, randomize them to continue treatment or receive placebo, and see whether patients worsen when the treatment is withdrawn without endangering them. Thus, the randomized withdrawal (or discontinuation) design can be used to evaluate the efficacy of long-term treatment already in use long before the study starts. Randomized withdrawal designs generally show larger group mean differences in pain intensity and fewer adverse events after randomization than traditional parallel designs because nonresponders are excluded, decreasing the risk of trial failure.[131,171]

The main disadvantage of the randomized withdrawal design is the complexity of interpreting and communicating the results. This makes them easy to politicize, as has been done with opioid studies.[159] Confusion arises when critics argue that randomized withdrawal studies are invalid because they cannot be directly compared with traditional parallel treatment studies. It is true that the results of EERW studies cannot be compared directly with the results from parallel or other types of designs: endpoints may be different (eg, hazard ratios or time to exit), which are not directly comparable with parallel group studies (which usually use group mean differences); by definition, the enriched enrollment study is enriched, which is not comparable with an unselected population; and pain worsening upon withdrawal of a treatment is not the same experience as reduction of pain upon implementation of a treatment. Therefore, the effect sizes from randomized withdrawal studies should not be directly compared with traditional parallel studies; this does not, however, undermine the validity of the randomized withdrawal design. Another challenge for the randomized withdrawal design is the comparison of 2 or more active treatments because various biases can affect the treatment results,[41] although there are methods to overcome these issues.[108] For example, one could imagine an EERW study where during the enrichment phase patients were given drug A, then were randomized into 1 of 3 arms: drug A, drug B, or placebo (3-arm trial). This study would be biased in favor of drug A because patients were enriched for effectiveness and tolerability of drug A, not drug B. A final potential disadvantage occurs when the criteria used to establish responder status are met by only a small percentage of the initial population, thus ballooning sample size requirements and undermining the applicability of the results.

5.3. Crossover studies

In a crossover design study (**Fig. 2B**), participants first receive one treatment and are then crossed over to receive another treatment in a random sequence. The classic design is the 2-treatment, 2-period, AB-BA approach, although this design can also use 3 or more periods, and in incomplete block designs, patients can be exposed to some but not all study treatments.[214] Many successful crossover studies of analgesics have been published.[99] The main advantage of a crossover design is that each participant's response to one treatment is directly compared with their response to another treatment, and low intrasubject variability rather than higher between-subject variability is used in the effect size (and *P* value) calculation. Accordingly, a crossover study requires considerably fewer participants than a parallel study to achieve adequate statistical power under most conditions. As long as treatment periods are not too long and the treatments do not have prolonged effects, additional study periods can be added without major increases in clinical trial size, duration, or budget.[165,199] Another advantage of the crossover design is that it accommodates patient treatment preference as an endpoint. A final advantage relates to patient recruitment: many patients prefer to enter a clinical trial in which they are guaranteed, at some point, to receive the study treatment.

Crossover studies are seldom used in confirmatory clinical trials of prolonged treatments because multiple 3-month study periods would be too long for most patients. Crossover studies are perceived to create regulatory risks, although crossover studies can be accepted by regulators as long as certain crossover-specific methodological issues are not problematic (eg, treatment-by-period interactions). The risk of carryover effect, where the effect of treatment given in one period is "contaminated" by treatment received in a previous treatment period, can be minimized by the use of washout periods, which is a treatment-free time period designed to be long enough to eliminate the effect of the previous treatment.[214] Crossover studies cannot be used to study resolving pain syndromes or treatments that produce long-term or disease-modifying effects. Crossover designs are useful for proof-of-concept studies where shorter treatment periods are adequate, in situations in which the treatment is used intermittently (eg, cancer breakthrough pain), when studying recurrent episodes of stereotypical attacks such as dysmenorrhea or migraine, and when patient recruitment is so difficult that the disadvantage of a long duration of individual participant participation is outweighed by the advantage of a smaller required sample size.

5.4. Noninferiority designs

A noninferiority study compares a test treatment with an active comparator with the goal of demonstrating that the test treatment is not inferior to the comparator, within a specified noninferiority margin. Noninferiority of the new treatment with respect to the reference treatment is of interest on the premise that the new treatment has some other advantages, such as greater availability, reduced cost, less invasiveness, fewer adverse effects, or greater ease of administration.[191] Several approaches are available to select the noninferiority margin,[162] including choosing a fixed margin based on a portion of the net effect size of the active treatment in previous studies, the lower bound of the 95% confidence interval around the treatment effect of a single placebo-controlled trial or a meta-analysis of such trials, or the synthesis method, which accounts for the variability of observed treatment effects.[79,162]

The validity of the noninferiority design depends on the assumption that the active comparator would have been more efficacious than a placebo control in the present clinical trial had a placebo control been included and therefore the trial has sufficient assay sensitivity to demonstrate inferiority if the study treatment were actually inferior. This cannot be assumed in pain studies because of the high variability of results for both active and placebo treatment seen between trials. According to the FDA Guidance on noninferiority designs, "If the intent of the trial is to show similarity of the test and control drugs, the report of the study should assess the ability of the study to have detected a difference between treatments. Similarity of test drug and active control can mean either that both drugs were effective or that neither was effective."[79] Thus, conclusions from noninferiority studies of treatments for pain (or other therapeutic areas in which trial results vary substantially between trials), which only compare the test treatment and an active control, are not valid for evaluating efficacy.[63,79] Only studies that incorporate an internal demonstration of assay sensitivity, ie, superiority of one treatment to another or a placebo, can be used to support conclusions of noninferiority. Examples of demonstration of assay sensitivity include a 3-arm study where an active control is demonstrated to be superior to placebo. Despite the limited interpretability of noninferiority designs without internal demonstrations of assay sensitivity, they have been used to support medical device approvals in the United States, although standards for the approval of medical devices appear to be evolving.

5.5. Adaptive designs

An adaptive design (Fig. 2D) is a prospectively designed study that plans for future design modifications depending on data accrued during the trial while controlling for type I error and minimizing operational biases.[83] Aspects of trial design that are potentially modifiable include the randomization strategy (eg, changing the allocation ratio), number of dose arms (ie, dropping a dose arm), sample size, stopping for efficacy or futility, inclusion/exclusion criteria, the primary endpoint, and transitioning from one study into the next.[23,185] Adaptive designs require some form of *interim analysis*, which is defined as any examination of data obtained from participants in a trial while the trial is ongoing. *Comparative* interim analyses consider unblinded treatment group assignments in the analysis, whereas *noncomparative* analyses do not.[83]

A general reason to use an adaptive design is statistical efficiency. Adaptive designs can provide greater statistical power (or decreased sample size requirements) compared with nonadaptive designs. Another important reason is ethics: studies can be stopped early if an interim analysis shows that the treatment is effective or futile. Adaptive designs involving sample size re-estimation can be useful in phase 3 studies, especially considering that sample size estimates based on phase 2 data may not accurately predict sample size requirements in phase 3 because of changes in study design and conduct. Finally, adaptive designs may offer economic efficiency: sponsors can stop trials sooner than planned if the result becomes apparent.

Adaptive designs also introduce several challenges.[83] Achieving consensus about whether type I error has been controlled can be difficult because of varying approaches to calculation of the impact of interim analyses on overall type I error. Adaptive designs can require extensive planning, specialized statistical consultants, computer simulations, operational integration, multiple statistical teams, specialized documentation, and multiple rounds of regulatory consultation. If there is a strong desire to use more than the simplest type of adaptive design for regulatory approval, it is often useful to begin discussions about the study a year in advance. Sometimes gains in statistical efficiency are associated with increased statistical risks (eg, the *expected* sample size may be lower than that in comparable fixed-design studies, but the *maximum* sample size may be higher) or are otherwise neutralized by operational realities (eg, enrollment may be complete before the interim analysis is performed). Finally, other considerations such as a need to accumulate sufficient safety data to characterize the risks of treatment may limit the usefulness of adaptive approaches.

5.6. Enriched designs

There are 3 fundamental purposes for enrichment studies: (1) to decrease measurement error (eg, by enriching for patients who are able to report their symptoms accurately or will be more adherent to study procedures, discussed further below), (2) to include patients with a greater likelihood of experiencing an endpoint or level of symptom intensity, and (3) to include patients with a potentially better biological response to treatment (greater benefit or less risk). These efforts increase assay sensitivity and statistical power. Enrichment strategies are applied before randomization, unlike an adaptive design. Therefore, enrichment efforts do not undermine internal validity but may change external validity, ie, to whom the results apply. Although every clinical trial is in principle enriched (not everybody with the target indication is included), studies of chronic pain often use explicit enrichment strategies.[63,64,131]

Methods for selecting patients who are potentially more responsive to study treatment include selection of those with a history of positive responses to the treatment or class; those with responsive phenotypes (if known); or those who respond positively to a direct challenge with study treatment, often called *empiric enrichment*. Enriching a study sample with patients who have a high likelihood of positive response has 2 main advantages: a smaller postrandomization sample size requirement and greater risk–benefit balance for the studied sample.[80] The most common empiric enrichment method is to give patients open-label study treatment and identify those with the desired outcome (usually effectiveness and tolerability), who are then randomized. After enrichment, there are 2 options. Treatment can be withdrawn, and patients can then be randomized to treatment or control once their pain has returned; this is known as an *enriched enrollment randomized treatment* study, of which few have been conducted in the field of pain.[197] A more common alternative for analgesics is the EERW design discussed above. The enriched enrollment randomized treatment approach

has the advantage of functioning like a prospective parallel treatment design, but disadvantages include operational complexity, patient attrition between the open-label phase and the double-blind phase, and little general experience with the design.

Some investigators have used double-blind enrichment phases in which participants who respond to treatment but not placebo are rerandomized to receive drug or placebo in a subsequent phase (eg, Byas-Smith et al.[32]; see the review by Quessy[197] for others). It is unclear whether the additional complexity of a blinded enrichment phase is worthwhile. Enrichment can also be attempted by simply asking patients about their previous responses to a treatment; however, this historical information is usually unreliable, even with medical records. One review comparing studies of pregabalin and gabapentin based on whether the studies excluded or included patients with previous study treatment failures found that efficacy was ultimately similar between them.[233]

If enrichment is successful, the effect size of treatment in an enriched group of patients will be larger than that in an unenriched group.[80] Of course, in an unenriched study, the observed effect size is driven entirely by the subgroup that improved, which is typically around 40% of randomized participants in pain studies.[173] The question often arises of whether enriched enrollment studies are generalizable or can be extrapolated to the general population compared with an unenriched study. The answer depends on how the question is constructed: in both cases, the results are driven by a subgroup that improved. Although the literature on this topic is mixed, the author's view is that effect sizes in an enriched study will be larger because the nonresponsive subgroup has been excluded.[131,170,171]

5.7. Flare designs

The flare design was initially developed for the study of nonsteroidal anti-inflammatory drugs (NSAIDs) for osteoarthritis. Investigators who were interested in studying the natural "inflammatory flare" that occurs periodically in patients with osteoarthritis attempted to produce this phenomenon by stopping chronic high-dose NSAID treatment, and patients whose pain worsened were then randomized.[243] Thus, the flare design is a type of pharmacologic enrichment that identifies responders by symptom worsening after withdrawal of treatment rather than symptom reduction after treatment initiation. The term "flare" is used somewhat loosely: while strictly speaking, "flare" refers to an *increase* in pain once treatment is withdrawn, sometimes the term has been used to indicate a *minimum* pain intensity after washout, regardless of whether it increases. Two meta-analyses have demonstrated that flare designs substantially increase the assay sensitivity of trials of NSAIDs.[20,243] Moreover, there appears to be a "dose–response" relationship between flare strictness and observed effect size: studies requiring only a minimum pain intensity at randomization yield a smaller standardized effect size (SES) than those requiring a flare, and studies requiring both a flare and a minimum pain intensity yield an even higher SES. There is too little experience with flare designs for the study of other classes of analgesics to know the extent to which the flare design would improve assay sensitivity.

6. Trial phases and treatment groups

6.1. Prerandomization: screening and baseline

The prerandomization phase includes a screening period to establish eligibility for randomization and a baseline period to establish the patient's baseline status for the computation of change-from-baseline metrics and to assess the comparability of treatment groups at baseline. Trial designers must decide what types of treatments patients can use during the prerandomization period. In general, the same types of rescue or concomitant treatments that will be allowed during the postrandomization period should be allowed during the baseline period so that changes can be attributed to study treatment. Much effort has been dedicated to discussing the merits of placebo run-in periods, which attempt to eliminate placebo responders by excluding patients who report a decrease in pain during the placebo run-in period. Placebo run-in periods do not reduce postrandomization placebo responses (at least in psychiatric studies) and should not be performed for that purpose.[64,77,249] However, a baseline period in which patients can be evaluated for compliance with study procedures (eg, e-diary compliance, adherence to [placebo] study treatment, and recording the use of rescue medication) and undergo assessments of pain variability is useful for patient selection.[63,107,166] Other options to reduce bias in the prerandomization phase including double-blind lead-ins, where patients start their active treatment at various times that are kept double-blinded, have been used in depression studies and merit further evaluation for pain studies.[53,54,77,104]

The optimal duration of the baseline period is unknown; successful pain studies have used baseline observation periods ranging from no baseline period (patients randomized on the day of screening) to 2 weeks.[111,173] Whether a longer baseline period would better establish stable symptom intensity, as appears to be the case with other disorders such as migraine and epilepsy, deserves further exploration.

6.2. Treatment phase

Once patients are randomized, the structure of the treatment phase depends on the nature of the treatment. For treatments that are administered as a single dose (eg, intra-articular injections), treatment is best administered on the day of randomization to avoid the possibility of events occurring between randomization and treatment, and the remainder of the treatment phase simply consists of observation. Oral medications that will be administered at fixed doses can also begin immediately. Sometimes, it is useful to administer the first dose of treatment at a (roughly) fixed time for all patients (eg, 9 AM on the day of randomization) and use the opportunity for in-clinic observation, to capture samples for pharmacokinetic studies, to evaluate first-dose analgesic effects, or to train patients on self-administration techniques of more complex products (eg, self-injectors, topical gels, patches, and pills in "smart packaging"). Some drugs require titration to the therapeutic dose to minimize side effects and early dropouts; in such cases, the treatment period can be divided into titration and maintenance periods or combined into a single period. This decision is generally based on regulatory negotiations: shorter studies are usually better from the perspective of the sponsor as they are subject to fewer dropouts and generally have better assay sensitivity than longer studies, but regulators may require fixed (eg, 12 weeks or longer) maintenance periods, in which case the titration period must precede the maintenance period, adding to the duration of postrandomization treatment.

For behavioral interventions, the intervention will be administered by trained clinicians according to a manual, on which they have been trained, and competency demonstrated.[67] The protocol will need to include ongoing assessments of treatment provider and patient adherence to the treatment regimen, which

may be complex and consist of multiple components.[22] Adherence to plans for maintaining blinding should be documented.[24] For invasive treatments, such as spinal cord stimulation or nerve blocks, similar considerations apply.[39] Careful documentation should be captured of clinician adherence to the treatment protocol and patient adherence to the prescribed treatment regimen (eg, using their device). Ongoing documentation of concomitant treatments and activity that may impact outcome assessments, such as use of prohibited pharmacologic and nonpharmacologic treatments, level of physical activity, changes in work and personal status, must be captured. Control groups will need specific management depending on their nature; for example, attention will need to be paid to documenting care in "treatment as usual" or "waitlist control" groups.[67]

The time point at which the primary endpoint is determined often coincides with the end of the double-blind treatment period, which addresses regulatory and scientific interest in the outcome of treatment, compared with control, at the longest observation time available, which is felt to best reflect the likely long-term impact of treatment. The limitation of this approach is that it essentially ignores the patient's response over time, which may provide useful information. Indeed, some regulatory agencies prefer assessing efficacy over the entire duration of treatment, which has led in some cases to different primary endpoints specified for different regulatory jurisdictions in the same clinical trial.[2] An additional risk of assessing efficacy only at the very end of treatment is the "hockey stick" phenomenon, where a clinical trial is positive every week until the final week of treatment (which unfortunately was specified as the primary endpoint). During the final week, the placebo group suddenly improves, the treatment group suddenly worsens, or both, leading to a convergence of the 2 treatment curves resembling a hockey stick.[161] Further research is needed to evaluate the prevalence and characteristics of this phenomenon. Meanwhile, one possible approach is to disconnect the capture of the primary endpoint from the end of treatment and blind investigators and patients to the difference. For example, in a study that seeks to demonstrate efficacy at 12 weeks, treatment might be continued until week 14 and, unbeknownst to participants or investigators, the primary endpoint determined based on pain scores captured at week 12.

6.3. Posttreatment

The main questions at the end of treatment are whether to stop the study treatment abruptly or use a taper; whether to attempt to capture what happens to symptoms after treatment stops; whether to continue participants into another study phase (eg, an open-label or blinded extension); and how to transition patients back into routine clinical care.

For drug studies, abrupt discontinuation is generally preferred, even with drugs with some potential for a withdrawal syndrome, to characterize the incidence and severity of withdrawal (if it occurs), because the clinical trial is the best opportunity to safely characterize these phenomena. Of course, if abrupt discontinuation is known to be unsafe or not easily treated, such as from earlier trials, a blinded taper is preferred. This issue is by no means limited to the end of the study: patients commonly stop their own medication for short or long periods of time during trials. Investigators should be instructed to evaluate the possibility of withdrawal as the cause of adverse events during trials, and know how to manage it. In trials of behavioral or invasive treatments, after the primary double-blind treatment period patients may be allowed to either cross into the other arm,[178] or all patients may get access to open-label treatment. Although it may be possible to glean further efficacy information from observations made after the end of the primary observation periods, such as from optional crossovers or entries into open-label treatment, careful consideration should be given to biases, statistical power, and operational complexity in planning such analyses.

Transitioning patients back to normal care after the study is not often given much consideration and can leave investigators and patients struggling at the study end. For example, if a patient discontinues a prestudy analgesic (eg, an opioid) to participate in a clinical trial and wishes to restart the therapy after the trial, who will prescribe it? In the author's experience, physicians may refer complex and challenging patients into clinical trials to offset difficulties in caring for such patients and may not welcome them back when the trial is completed, especially when decisions must be made about continuing complex, risky, or burdensome treatment (eg, opioids). These issues have both medical and ethical implications and are best sorted out before the study starts.

6.4. Treatment groups

To fulfill study objectives, confirmatory clinical trials must carefully consider the choice of control groups. The default design, performed to address the question of whether a treatment is better than receiving an inert treatment in a clinical trial, is addressed by using a placebo or sham control group. In analgesic studies, placebo controls are almost always expected and have generally been viewed as being not only ethically appropriate but ethically necessary to yield the most robust characterization of the efficacy and safety of the test treatment.[63,64,72,82] In cases in which recruiting patients into a placebo arm is unethical or impractical, other designs can be considered. A low-dose control (eg, Rowbotham et al.[208]) offers the benefit of improving blinding (low doses of a drug often produce some degree of the side effects of full doses). However, low-dose control designs must be handled with care because low doses are sometimes more effective than anticipated and can compromise the primary study objective.[206,219] Another option is the add-on design, where all participants can continue prestudy analgesic treatments, have such a treatment deliberately added, or receive the "standard of care" in addition to study treatment or placebo. The latter approach can be problematic in the context of a confirmatory clinical trial because participants are often treated with a variety of different therapies, decreasing assay sensitivity despite randomization.[64,130,173]

Sometime, studies use "active placebos"—medications that produce low-grade side effects—to maintain blinding. Options include diphenhydramine, benztropine, benzodiazepines, and loperamide.[46,101,209]

Active placebos are not recommended except in unusual circumstances[63,64] because patients are effectively exposed to the risks of a medication without any benefit, and the issue of unblinding can be addressed in other ways as discussed above.

With respect to the active treatment arm, care must be taken to ensure that it represents the investigator's best "shot on goal" and the treatment regimen that is intended to be used in practice. If there are significant areas of uncertainty surrounding the dose, titration, and frequency of administration, components of a complex intervention, administration methods, or other important treatment parameters, it is better to first design a trial to resolve these issues rather than jump to a confirmatory study. A common issue in drug studies is that early trials may support a "good enough" dosage and administration regimen that used the highest possible doses or frequency of administration. Although

these dosing paradigms may have been appropriate to maximize the chance of successfully identifying efficacious treatments in early studies, when carried into confirmatory trials they may result in low adherence, excessive side effects, or (if the trial is positive) an overly burdensome prescribed regimen in the marketplace. In the realm of behavioral interventions, pilot studies are often performed to refine these nuances before confirmatory studies.[66,205] In the realm of neuromodulation or other devices, open-label uncontrolled studies are often performed before RCTs to sort out procedural techniques and optimal regimens.[81]

Studies should generally include an active comparator for benchmarking without an attempt to directly compare the study treatment with the comparator because such comparisons will almost always be underpowered. In some regions, active comparators are required for marketing authorization.[74,75,79] The main motivation for including an active comparator with well-established efficacy is that, when used alongside a placebo group, it enables the assessment of assay sensitivity; if the investigational treatment fails, the best way to judge whether the treatment failed or the trial failed is to examine the performance of an active comparator.[64] An informative example is a RCT of a gabapentin prodrug in painful DPN that included pregabalin as an active comparator; in this study, neither drug significantly separated from placebo, suggesting that the trial lacked assay sensitivity.[266] Ironically, active comparators are routinely included in highly reliable acute pain models such as dental pain, whereas they are seldom included in far less reliable chronic pain models.

Active comparators can also be used to directly support study objectives. When active comparators are included for regulatory approval or pricing considerations, the comparator must be chosen carefully, as pricing may be based on the price of the comparator or on cost-effectiveness calculations vs the active comparator. Challenging technical issues may arise when considering how to incorporate an active comparator into a clinical trial. For example, incorporating an active comparator that requires titration (eg, an opioid) into a trial of a study drug that does not require titration (eg, a NSAID) may create challenging logistical problems (eg, how to provide blinded drug supply when the dose is adjusted based on tolerability) or interpretation problems (eg, determining whether the titration algorithm for the drug that required titration was optimized).

6.5. Protocol complexity

Protocol complexity can be quantified based on the number of objectives and endpoints, eligibility criteria, study procedures, burden of work on the site, and case report form pages per protocol. Execution complexity can be quantified as the number of countries, sites, patients screened, patients randomized, and data points collected.[93,94] In one study, more complex protocols were associated with half the screen-to-completion rate, 12% longer time to first patient first visit, 73% longer time to last patient last visit, and 68% more amendments relative to less complex protocols.[95,96] Thus, protocols should be simplified to the extent possible, with the goal of prioritizing the ability of the trial to address its primary objective.

7. Patient selection, recruitment, and retention

A clinical trial must balance 2 opposing forces: *patient recruitment*, getting patients in the door, and *patient selection*, choosing the right ones. Choosing the right patients is both art and science; the goal is to select participants who represent the disorder of the target population; have relatively stable symptom severity that is unlikely to fluctuate dramatically, or resolve, during the study; are not anticipated to have major lifestyle perturbations; will comply with the protocol; report their symptoms accurately; adhere to study treatment; and avoid prohibited activities or treatments. Although all study personnel in principle want the study completed quickly and rigorously, in practice, study teams are usually divided into operations personnel who prioritize getting the study performed quickly and focus on recruitment and clinical scientists who prioritize scientific rigor and focus on selection.

7.1. Patient recruitment

Recruitment in clinical trials is a major source of delays in bringing new treatments to market,[175] with the cost of delays estimated between $600,000 and $8 million per day per product in development, based primarily on lost time in the marketplace.[110] Twenty-five percent of all drug development delays are said to be due to slower-than-expected patient enrollment,[1] nearly 80% of clinical trials fail to meet enrollment timelines, and up to 50% of research sites enroll one or no patients.[43] Recruitment is not just about money and time; delayed time to market also deprives patients of potentially useful new treatments. Clinical trials seldom report recruitment details, undermining the ability to assess the generalizability of research results and the effectiveness of recruitment methods.[220]

At the patient level, factors that impact recruitment include the patient's health status: pain, immobility, and fatigue limit willingness to participate, while the prospect of personal benefit, altruism, access to medical care and diagnostic assessments, and potential access to new treatments motivate participation.[120] Primary care physicians view their patients as vulnerable and attempt to protect them from potentially destabilizing interventions, often discouraging patients from participation in clinical trials.[120] Attitudes towards research also govern recruitment, with some populations suspicious of the research establishment.[215] Altruism is a motivator but seldom by itself.[120] Previous negative experience with the study treatment, randomization, and placebo controls are demotivators. Patient engagement also influences study participation. Key factors include communication between patients, gatekeepers, and the trial team; marketing and presentation of the trial; trust in the trial team; the opinions and endorsements of others; and the opinion of the referring doctor.[58] An illuminating study found that the most important preferences of patients with chronic pain for clinical trial features were no invasive procedures (including blood tests), ability to continue current pain medications, higher monetary compensation, and fewer in-person visits (but more phone contacts).[227]

Patient engagement can help align research topics to patient priorities, improve data collection tools, increase patient participation, and improve the dissemination of results. An effective strategy for participant engagement is involving patients in the design and implementation of clinical trials, particularly when identifying strategies to overcome recruitment obstacles.[156] Yet, patient engagement requires time, financial support, and energy from patients, stakeholders, and researchers to yield mutual benefit.

Research sites are selected for several main reasons, among which a primary reason is the number of patients in their research and/or clinical databases of patients with the target disorder, because this is perceived to accelerate recruitment. These databases are rarely reviewed and validated by the research sponsor and are often overestimated. Related issues include failure to quickly and effectively contact potential participants,

overestimating how many patients can be enrolled from clinical practice, overestimating the usefulness of electronic health records for identifying patients, overestimating the usefulness of advertisements, assuming that previously effective recruitment strategies will continue to be useful, and failing to set up systems to track prescreening and screening activities. These obstacles are not generally addressed in the clinical trial contract or budget. This is discussed further under site selection.

Numerous site-based patient recruitment strategies have been developed and implemented with variable success.[25,176,202] Patient databases can be a tremendously useful resource when they contain correct, current information. Effective sites stay in contact with patients listed in their database and perform periodic updates. Effective sites also have dedicated recruitment teams who can focus uninterruptedly on contacting patients in the database as soon as the study begins and on responding to inquiries from advertising, especially on nights and weekends when patients are home. Advertising is necessary for recruitment in most pain studies. Many sites focus on obtaining referrals from external physicians; this tends to be more useful when external physicians have incentives to refer patients, which may be challenging to achieve.[57]

7.2. Patient retention

Recruitment accomplishes little without retention. Dropouts create missing data and can render the study uninterpretable, leading to questions about the ethics of having experimented on patients in the first place. Although patients should discontinue participation in a trial when it is no longer in their best interest, the focus of patient retention efforts is to overcome avoidable obstacles to continued participation.[25]

The first step is to make the study protocol as patient-friendly as possible.[93,94,96] Investigators and sponsors are scientifically curious and generally interested in answering as many questions as possible, often resulting in excessively burdensome protocols.[93] Incorporating patient input upfront can illuminate the burden of participation in a manner that may suggest opportunities for streamlining a protocol and supporting patient engagement.[227] Sponsors may be penny-wise and pound-foolish in supporting patient participation; it is not unusual to spend tens of thousands of dollars per participant in a trial but fail to support relatively inexpensive measures (eg, cab fare and meals) to ensure sufficient retention to support interpretability of final study results.

Simple positive feedback is often a major source of emotional energy for continued patient participation in trials. Comments, postcards, emails, text messages, phone calls, and small gifts can all buoy the patient's spirit and remind them that their contribution is greatly appreciated, so long as it is done without inflating the patient's expectation of clinical benefit beyond an appropriate level.

7.3. Patient selection

Most trialists agree that patient selection is the key to successful clinical trials, yet have trouble explaining exactly what that means, because it appears to transcend the list of specific eligibility criteria. The pioneers of pain research ascribed a great deal of importance to the process of selecting patients with "the right stuff" and hiring nurse study coordinators with a track record of selecting the right patients, which seems to have been borne out by the large effect sizes seen in clinical trials at their sites.[163,164] In the early days, most studies were performed in single centers by investigators who dedicated their careers to advancing pain research; if the trial succeeded or failed, it was obvious who was responsible. Nowadays, the clinical research machine scarcely resembles its youthful version of a half century ago. Large multicenter trials are conducted across multiple continents and in multiple languages by organizations with variable expertise in the pain area and at sites managed by investigators who are seldom specialists.[125,210] In the modern era, improving patient selection requires systematizing the clinical intuition that characterized the successful work of early pain research pioneers.

7.3.1. Baseline participant characteristics

An appropriate distribution of age, sex, and race/ethnicity of participants is expected in a manner consistent with the disorder being studied. Socioeconomic status, education, occupational status, and baseline quality of life can all influence treatment outcomes[63] and should be documented in chronic pain studies to ensure balance between groups, allow stakeholders to understand the study sample, and explore the impact of these factors on outcomes. Health literacy and numeracy are complex and evolving concepts that focus broadly on an individual's capacity to seek, understand, and use health information and are worth further consideration in selecting participants who can contribute interpretable data in clinical trials.[28] The issue of representativeness of the study sample to the target population is controversial. There is little disagreement that the study sample should have the same disorder, with similar clinical characteristics, as the target population. On the other hand, to accurately characterize the pharmacologic or other characteristics of the study treatment, sources of measurement error and variability must be controlled. Thus, participants who will not be able to comply with the study protocol, or have extraneous conditions that may add error or noise to the observed relationship between treatment and outcome, are generally excluded from clinical trials. Finding a balance between achieving internal validity, by controlling these sources of variability, and achieving external validity, by selecting patients who represent the target population, requires careful consideration.

7.3.2. Diagnosis

Given that clinical trials endeavor to study patients with specific disorders, it is surprising how little effort is expended to ensure that diagnostic assessments are performed reliably and are clearly documented. One study reported a wide range of reliability of implementing diagnostic criteria for osteoarthritis of the hip, knee, and hand using either clinical criteria (kappa 0.0–0.65) or combined clinical, radiological, and laboratory criteria (kappa 0.31–0.85), indicating that raters often disagree about the presence or absence of these disorders.[16] Reliability was substantially higher among experienced rheumatologists, suggesting that reliability can be improved with training and experience. In a study of the performance of the Michigan Neuropathy Screening Instrument for diagnosing diabetic neuropathy, the best cutoff score correctly classified approximately 80% of patients, indicating that 20% were misclassified.[116] In a recently published clinical trial in posttraumatic neuropathic pain, a diagnostic assessment that was centrally reviewed by a team of trained neurologists ended in the rejection of approximately 30% of patients deemed eligible by the investigators; this rate improved over time, once more suggesting that investigators are responsive to training.[161] Better training, performance monitoring, and central reviews of the patient's diagnosis are "low-hanging fruit" for improving clinical trial reliability.

A related issue is that the rate of positive studies is appreciably higher for some chronic pain conditions than others. For example, the rate of positive studies tends to be higher in postherpetic neuralgia and DPN than for HIV-associated neuropathy or lumbosacral radiculopathy. The reasons for this are not entirely clear and may relate to clarity of diagnosis, presence of non-neuropathic pain elements, different pathophysiologies of neuropathic pain, concomitant comorbidities or treatments, or other issues.

7.3.3. Pain intensity and duration

Patients must have a minimum pain *intensity* for a sufficient *duration* of time to demonstrate a difference between active treatment and control. The traditional cutoff for pain intensity in clinical trials is a score of 4 on the 0 to 10 NRS, which is equivalent to a score of 40 on the 100-mm visual analog scale (VAS), or "moderate" on a 4-point verbal categorical scale (none, mild, moderate, and severe).[63] Little empirical research has been published on the performance of various cutoffs for distinguishing analgesic from placebo. One study of duloxetine in patients with painful DPN identified a larger SES in patients with baseline pain ≥6/10 compared with those with baseline pain <6/10.[268] Alternatively, a meta-analysis of clinical trials of a high-concentration topical capsaicin patch identified a higher SES in patients with baseline pain ≤4 than those with pain 4 to 7.[136] Higher cutoffs for pain intensity screen out a larger proportion of participants, such that even if the impression that higher pain cutoffs are associated with greater assay sensitivity is true, the real question becomes what cutoff optimizes trial efficiency as a whole (eg, it may be better to have a patient with a pain score of 4 than not have a patient at all). A reasonable approach is to use a minimum cutoff of 4/10 and focus on other important aspects of patient selection such as ensuring that patients can report their pain accurately and do not exhibit baseline score inflation.[63–66] If this is done in early trials, empiric analyses can be performed later to see whether excluding patients with baseline pain <5 would improve assay sensitivity in subsequent studies. It has also become commonplace to exclude patients with average daily pain of >9/10 because these patients may have extreme psychological distress or such severe pain that they should not be in a trial.[63,64]

There are only a few reports of the impact of patients' prestudy pain duration on study outcome, and these assessments generally suggest that it does not matter much (eg, Ziegler et al.[269]). However, of the 7 RCTs reported in a meta-analysis of the capsaicin 8% patch, the only study that failed to show efficacy was one that allowed patients with a minimum prestudy pain duration of only 3 months; all other studies required at least 6 months.[136] The average prestudy pain duration in studies of chronic pain is typically in the range of 10 years[37,132]; thus, given that patients with relatively short durations of prestudy pain are more likely to experience spontaneous resolution of pain during the study, a minimum duration of prestudy pain of 6 months, if not 12, appears appropriate.[63,64] Importantly, most definitions of chronic pain require only a 3-month duration of pain; thus, the above recommendation could potentially exclude some patients who meet criteria for a chronic pain condition yet do not meet study eligibility criteria. While in practice this does not seem to impact recruitment, because patients with such recent onset of pain generally do not present themselves for clinical trials, and including patients whose pain may resolve spontaneously undermines the goals of characterizing the efficacy of treatments, investigators should consider whether in a specific trial a shorter duration may be appropriate. With respect to maximum prestudy pain duration, the current IMMPACT recommendation is to not impose a cap.[63,64]

Another imperative is to ensure that patients' pain intensity is relatively stable for a period of time before randomization. The current standard is to establish a minimum intensity based on the average of daily pain scores over 1 week before randomization.[63,64] More research is necessary to determine whether a longer period, potentially up to a month, would improve performance because there is a large rate of decline of pain intensity in pain studies, as evidenced by the placebo response.

7.3.4. Pain variability

A major recent advance in pain research methodology is the discovery of a relationship between pain variability, placebo response, and assay sensitivity. Harris et al.[111] first reported that patients with high variability in baseline daily pain also had high placebo responses (whereas responses to treatment were unaffected); this was subsequently confirmed in a meta-analysis of 10 other studies.[76] Therefore, variability of daily pain provides an opportunity to identify "preferential placebo responders" (patients who will have a high response to placebo but not active treatment),[239–241] which is more relevant to clinical trials than simply predicting a high placebo response alone. High variability of experimental pain predicts preferential placebo responsiveness even better than variability in clinical pain.[240] Thus, it has become commonplace to exclude patients who exhibit high baseline pain variability based on the SD of daily scores from an electronic diary, and some studies have excluded patients based on high variability in experimental pain.[165,241]

7.3.5. Baseline score inflation

Several studies in psychiatry have found that clinician-based assessments reported by investigators at baseline (ie, when eligibility for randomization is determined) are often scored higher than assessments reported by independent third-party raters, presumably because of intentional or unintentional inflation of baseline scores to facilitate clinical trial entry.[148,154] This phenomenon has been recognized as a source of measurement error in pain studies in the analgesic drug development guideline of the European Medicines Agency[71] and is considered routinely when designing clinical trials of analgesics. That said, there is no consensus on the best way to prevent this kind of bias. Available approaches include blinding investigators and patients to the minimum pain intensity criterion by using masked protocols; randomizing patients regardless of their baseline score and prespecifying the primary analysis cohort as those meeting the minimum pain intensity criterion; and requiring that patients satisfy minimum requirements for 2 different pain measures (which can also be masked).

7.3.6. Medical and psychiatric comorbidities

Patients with chronic pain commonly have medical and psychiatric comorbidities that can potentially affect the safety and efficacy of a new treatment, the patient's ability to report symptoms accurately, and their ability to follow the protocol. Patients with multifocal chronic pain syndromes are common in clinical practice; however, in clinical trials, it is desirable to exclude other painful conditions. A reasonable approach is for all patients to be screened with a body diagram on which the patient can record any part in which they have experienced substantial recent

pain.[113] The medical history associated with each painful area can be documented and a more informed and auditable judgment applied. Patients with moderate or severe pain in areas outside that being studied can be excluded, except where the intention is to study patients with widespread pain, such as fibromyalgia.

Anxiety, depression, insomnia, substance abuse, posttraumatic stress disorder, and other psychiatric issues are highly prevalent among patients with chronic pain and to some extent define the chronic pain syndrome.[63,64] Patients with significant psychiatric comorbidities are typically excluded from clinical trials of analgesics for safety reasons and because of the presumption that these patients may undermine the trial due to poor adherence to treatment and study procedures; inaccurate reporting; fluctuations in pain intensity due to psychological, social, and other influences; and perhaps the very nature of their pain syndrome. Published data on the impact of these comorbidities on the effect size of analgesics are scarce, although one study did show that patients with negative affect had higher placebo responses and smaller responses to active treatment compared with control patients.[261] On the other hand, complete exclusion of these patients makes it nearly impossible to conduct a pain study and raises questions about the population to which the study results apply.

An additional complication is that patients must have sufficient cognitive function and language ability to complete questionnaires or other assessments accurately. This is rarely assessed at screening, and in the author's view, tests of cognitive function and health literacy should be more routinely implemented. Patients with severe or unstable concomitant medical disorders are also typically excluded because they may experience adverse events more likely related to the comorbid disorder than study treatment, confusing the assessment of safety.

Current recommendations are to exclude patients with significant psychiatric comorbidities from chronic pain studies by using validated screening instruments.[64] Patients with substance abuse problems should also be excluded using both a validated questionnaire and quantitative urine drug tests at screening and during the study. Investigator judgment and patient self-report correlate poorly with urine drug screens and therefore are an unreliable basis for eligibility.[133,138] For trials involving drugs with abuse potential, a lifetime exclusion for past addiction is probably appropriate, unless the intent of the study is to examine abuse potential in a higher risk population. For drugs without abuse potential, a shorter period (eg, 2 years) may be appropriate. To the author's knowledge, prescription drug monitoring data have not been used to detect stigmata of substance abuse in clinical trials, but should be.

7.3.7. Concomitant and rescue analgesics

Ideally, in an RCT of an investigational pain treatment vs some control, patients would not use any additional pain treatments, so that any differences between groups could be clearly ascribed to the investigational treatment. In reality, during trials of pain treatments, whether pharmacologic, invasive, or behavioral in nature, patients often use nonstudy treatments for pain, which can also be pharmacologic, invasive, or behavioral. To complicate matters further, these additional pain treatments may be for the index pain being studied in the clinical trial, or may be for a new pain syndrome (eg, headache and backache), or a preexisting pain syndrome, which has continued to cause pain or which has flared. Because interpretation of the study results requires understanding the degree to which additional pain treatments have been used, and why they have been used, it is important to have clear terminology, and method of quantification, for these extra treatments. Unfortunately, no consensus on either the terminology or the best method for measuring these extraneous treatments exists.

In general, medications that are used on a fixed-dose basis for the patient's index chronic pain are referred to as *concomitant analgesics*, and medications that are used on an as-needed basis for the index pain (which are often provided by the sponsor in an attempt to achieve some control over these medications) are referred to as *rescue analgesics*.[63,64] Protocols typically specify what concomitant and rescue medications are "allowed," although patients commonly transgress these prohibitions, which is often understandable in view of the need to treat exacerbations of the index or nonindex pains. It is important to capture any use of concomitant and rescue analgesics, and the exact amounts used, as well as whether any new medications, or increased doses of existing medications, were for the index pain, a new pain syndrome, or exacerbation of a preexisting pain syndrome. It is also important to determine whether in the latter 2 cases the new medication use represents an adverse event. These attributions often figure critically into the calculation of the primary or key secondary endpoints of the clinical trial because patients may be considered responders to study treatment depending on whether new medications were used for their index pain; for that reason, independent confirmation of this attribution may be useful.

Several studies have shown that allowing multiple concomitant and rescue medications decreases the observed effect size of treatment,[130,137] presumably because patients in the placebo arm are actually receiving treatment, thus undermining the primary study objective. Safety data can also be confounded by allowing patients to take multiple additional medications. Providing an abundance of rescue treatments does not seem to be necessary in most pain studies because most patients on placebo do fine with minimal rescue medication. On the other hand, many research participants are taking one or more concomitant analgesics and are averse to stopping them and must be assured that their clinical state will not be destabilized during the trial. Current practice is to limit the number of concomitant pharmacologic treatments to the minimum that is consistent with the realities of recruitment, retention, and reasonable medical practice.[63,64]

A related consideration is concomitant nonpharmacologic treatments such as physical therapy, acupuncture, psychological support, and ice and heat. Standard practice has been to allow enrolled patients to continue using these modalities if the nonpharmacologic treatments remain stable during the study. Although this seems reasonable, the author has never seen a clinical trial in which the use of nonpharmacologic treatments was rigorously quantified before or during the study, or where it informed the analysis. The degree to which this is a source of measurement error in clinical trials is unknown, suggesting that further efforts to capture this information and evaluate its impact on treatment would be useful. Some sponsors have considered providing a light but consistent program of physical and psychological support during clinical trials to ensure consistency across participants and treatment arms, which may further offer advantages for ethics, recruitment, and retention. Several evidence-based online programs of this type are available.[207]

7.3.8. Sensory phenotyping

The mechanism of pain may differ between patients with pain of similar etiology (eg, PHN or osteoarthritis), leading to the idea of

"mechanism-based pain treatment."[265] Most efforts to define pain mechanisms for clinical trials have focused on quantitative sensory testing to divide patients into "phenotypes" that presumably reflect independent pain mechanisms.[11] A comprehensive review of phenotyping recommended that a sensory phenotyping approach be considered for all pain studies.[69] Much of this research has been conducted by the German Research Network on Neuropathic Pain,[203] and several groups have developed simple bedside sensory testing approaches to classify patients for clinical trials based on putative pain mechanism.[85,183,200] Several studies have suggested that sensory phenotype does predict the net treatment effect in patients with neuropathic[47,48] or musculoskeletal pain.[183]

7.3.9. Response to previous treatments

Investigators and sponsors may be interested in avoiding patients who have been refractory to previous treatments under the presumption that such patients will decrease the effect size of a new study treatment. Alternatively, sponsors may wish to select these types of patients to evaluate whether a new treatment is effective in refractory patients. At the time of this writing, there is no evidence that past treatment failure predicts treatment efficacy; recent studies of anti-Calcitonin gene-related peptide CGRP antibodies for migraine[245] and anti nerve growth factor-NGF antibodies for musculoskeletal pain,[103,132,155,212] which enrolled only patients who had failed previous treatments, were consistently positive. Limiting enrollment to patients who have failed simpler treatments is appropriate when the study treatment is known to have significant safety risks, thus providing better justification for the risk in view of the lack of available alternatives for such patients. Selecting patients who have been refractory to specific previous treatments may cause bias either in favor of or against the study treatment. For example, in a study comparing drug A with drug B, excluding patients who have been refractory to drug B (or its class) favors drug A.

As already noted, it is challenging to determine exactly what treatments a patient has received in the past and how they responded. At minimum, all pain studies should attempt to document and report past treatments and responses so that consumers of the data can understand the study sample, evaluate potential biases, and understand the impact of past responses on treatment effects. Having said that, such information is subject to recall biases and other forms of imprecision, even when medical records are available. In specific cases, including or excluding patients based on responses to previous treatments can be necessary to accomplish the study objectives. Further efforts are needed to determine how best to acquire this information.

7.3.10. Professional and duplicate subjects

A distressing number of patients have been found to disguise their identities and enter the same study at multiple centers, enter different studies simultaneously without disclosing it, or fabricate symptoms and create or hide medical histories to meet enrollment criteria.[98] A study of psychiatric trials in southern California found that 3.5% of participants were duplicates, whereas other estimates as high as 12% have been reported.[218] A survey of 100 experienced clinical research participants[56] found that 75% reported concealing some health information to avoid exclusion, one-third concealed health problems, 28% concealed the use of prescribed medications, 20% concealed recreational drug use, 25% exaggerated symptoms to qualify, and 14% pretended to have a health condition to qualify. The negative impact of these practices on the credibility and assay sensitivity of clinical trials is obvious. Several strategies have been recommended to prevent patients from entering trials under false pretenses[55] including confirming diagnoses from records or referring physicians, designing prescreening and screening scripts that conceal exclusion criteria, minimizing the information provided on the internet during study recruitment, and downplaying compensation for participation. Several clinical trial registries designed to detect duplicate patients are available, including Verified Clinical Trials (www.verifiedclinicaltrials.com) and CTSdatabase (www.ctsdatabase.com). These registries indicate that they are fully regulatory compliant, and studies indicate that they can successfully identify duplicate subjects.[218] The use of one or more of these registries should now routinely be considered in pain trials.

7.3.11. Pharmacogenomics

The pharmacokinetics and pharmacodynamics of most classes of analgesics are, to some varying extent, mediated by genotype.[172] Genetic polymorphisms and activity levels vary based on numerous factors including race, ethnic background, and tobacco abuse as well as interactions with other medications.[3] Major drug–drug and drug–gene interactions are common.[252] Some studies have described the influence of genetic polymorphisms on the pharmacodynamics of analgesics and on clinical manifestations of pain, although these data are less consistent than data for genotype–pharmacokinetic relationships.[231] Pharmacogenetics is genetic testing that assesses a patient's risk of an adverse response or a likelihood of responding to a given drug, thereby informing drug selection and dosing.[232] Genetic testing can be performed with blood, saliva, or buccal swabs. A variety of panels are available, and the cost of this testing has decreased over the years. Specific consideration should be given to performing pharmacogenetic testing in clinical trials of analgesics, especially when metabolism of the study drug is affected by these genotypes, as well as in cases in which a body of data suggests that a genetic variant may affect responses to the study drug or its class.

8. Assessment of adverse events

The value of treatment to a patient is based on a balance between benefits and harms; characterizing harms therefore remains a fundamental obligation in every clinical trial. Several methods are available for characterizing harms and differ in important ways.[61,135] Reporting guidelines for adverse events AEs in clinical trials are available from the Consolidated Standards of Reporting Trials (CONSORT) group.[123] At present, compliance with these guidelines in pain studies is suboptimal.[225]

8.1. Passive AE capture

The most basic form of accounting for harms (and a minimum expectation for all clinical trials) is passive capture of spontaneously reported events. This ambiguous method is associated with a high degree of variability within and across sites, time, and studies and is easily biased by the nature of interactions between staff and participants. Improved consistency can be achieved by scripting researcher–patient interactions with a protocol-specified nonleading prompt stated at every clinic visit such as "How have you been feeling?" (as opposed to a leading prompt such as "What side effects have you been experiencing?").

Similarly, protocols usually specify definitions for the different severity levels, but these are seldom controlled for quality, leading to haphazard ratings of AE severity. These simple approaches to improving the accuracy and reliability of AE reporting and characterization are "low hanging fruit" for improving the usefulness of clinical trial data.

8.2. Reporting of passively captured AEs

The incidence of AEs should be reported for each treatment group, including the percentages of participants who experienced one or more events. The severity of AEs should also be reported, as this may differ among treatments that have a comparable incidence of AEs.[68] It is a commonplace and entirely unacceptable practice to present only AEs that occurred above a certain arbitrary frequency (eg, 3% or 5%) or AEs that occurred more frequently in the active treatment group than in the control group by some threshold. These methods may hide severe AEs that occur at low rates (sometimes just below the chosen reporting threshold). Arbitrary threshold–reporting approaches may also hide AEs reported using different terms that mean much the same thing; if the terms were summated, the true incidence of the event might be considerably larger. For example, paresthesias, dysesthesias, numbness, neuropathy, neuritis, and allodynia are individual AEs that can reflect the same root problem—if each occurred at an incidence of 4% in a trial where only events of greater than 5% were reported, the overall incidence of the underlying problem would go unreported. Responsible investigators and sponsors can create categories of important events of interest that aggregate related individual events.[132] Although summary tables of AEs occurring above a specified frequency are appropriate for summarizing, complete AE data must be made available, potentially as supplemental materials for publications and certainly in complete study reports.

8.3. Deriving more insight from passively captured AEs

Even without capturing additional data, passively captured AE data can provide additional insights. The typical AE form contains fields for the start date, stop date, intensity (mild, moderate, or severe), and attribution of each reported AE. The integration of incidence, severity, and duration can further inform between-treatment differences. For example, in a recent analysis of data collected from a clinical trial comparing tapentadol with oxycodone, an integrated measure of AE severity and duration revealed impressive between-treatment differences, because of greater severity and longer duration of individual AEs in the oxycodone group, that were not apparent from scanning the standard AE tables, which only present the incidence of individual AEs.[137]

8.4. Adverse events of special interest

A more specific approach is to prospectively assess AEs of special interest (AESI). This can be accomplished in several ways. The best approach is to prespecify diagnostic criteria for the AESI, potentially including laboratory criteria or an adjudication committee. For example, a system for capturing abuse-related events for CNS-central nervous systemacting drugs identifies and classifies these AESI.[242] For dedicated safety studies, such an approach may become the primary endpoint of the trial, such as in studies focused on comparing the GI or cardiac safety of different NSAIDs.[177] A less satisfying approach is to simply prespecify the AE codes (eg, MedDRA Preferred Terms) that will comprise that category. Standardized structured MedDRA queries (SMQs) are available for certain AEs, although it is important to shop with care for MedDRA SMQs because some are more comprehensive and better validated than others.[238] For drug studies in which AEs may be related to the drugs pharmacokinetic profile, such as time to maximum concentration, standard narratives designed to evaluate such events should include fields for time and amount of the last dose preceding onset of the AE and related information.

8.5. Clinical outcome assessments for harms

Passive AE capture can miss differences between treatment groups in clinically important harms, even harms associated with mortality. Active capture of harms using structured interviews or questionnaires to assess specific symptoms or syndromes may discriminate differences in harms between treatment groups more effectively than passive capture.[7,67] Single AEs can be assessed with event counts or single-item intensity ratings, such as assessing postoperative nausea and vomiting with counts of vomiting events or a single-item NRS for nausea.[89] Syndromes (clusters of individual signs and/or symptoms) reflecting harm can also be assessed using questionnaires. A classic example is the opioid withdrawal syndrome, which has been measured for decades with several validated instruments.[263] A second example is opioid-related side effects, which represent another symptom cluster that includes nausea, dizziness, sedation, and constipation, and can also be assessed with several validated questionnaires.[29,30] Importantly, harms assessed with prospective outcome assessments will result in a higher incidence than passive AE capture, and therefore, values derived from prospective and passive measures cannot be directly compared.

8.6. Abuse-related events

Given the current opioid epidemic, extensive efforts are being directed towards the development of analgesics with low abuse potential. This requires evaluating abuse potential during development to inform labeling, approval, and clinical use, especially because many pain treatments come from classes of drugs with known abuse potential (eg, opioids, cannabinoids, and gabapentinoids) or are CNS active, which trigger abuse potential evaluations per regulatory guidelines.[82] Oddly, the premarketing assessment of abuse potential has historically focused on in vitro pharmacology, preclinical models, and human abuse liability studies in recreational drug abusers, without any systematic approaches to measuring abuse potential in phase 2 and 3 clinical trials enrolling patients with the target disease.[82] In 2013, IMMPACT recommended the systematic assessment of abuse-related events in clinical trials,[179] followed by recommended terminology and definitions of abuse-related events[226] and a call for the development of a standardized measurement approach after a systematic review found no appropriate measures.[228] In response, one such system has been developed and validated,[242] and a subsequent ACTTION review found that this system was the most suitable available tool for assessing abuse-related events in clinical trials.[229]

8.7. Adverse events for nondrug treatment trials

Although much of the above discussion has focused on AEs associated with drug therapies, AE capture in clinical trials of nondrug treatments is also important. First, patients in clinical trials of nondrug treatments, such as psychological therapies or invasive treatments, will also use drugs during the trial, and AEs associated

with rescue medication use should be captured because they may reflect a benefit (or harm) of the primary investigational treatment. Second, the study treatment itself may be associated with AESI. For example, spinal cord stimulation for chronic pain is associated with a well-defined set of complications that should be assessed prospectively.[140]

9. Dosing of study treatment

In principle, the optimal dosing and administration of study treatment is established before launching confirmatory studies. In practice, there are often unresolved issues related to dosing and administration of study treatments. In this section, we will frame dosing issues with respect to drug treatments; however, similar principles apply to physical modalities, devices that administer electromagnetic stimulation, psychological treatments, exercise, complementary treatments, invasive treatments, and other modalities.[39,67,140] It is difficult to evaluate the efficacy or safety of a treatment without knowing the right dose; to paraphrase Paracelsus, the only difference between a drug and a poison is the dose. Issues that may not be fully resolved by phase 3 include the minimum effective dose or frequency, optimal dose, optimal dose in specific subgroups, optimal frequency of administration, optimal titration rate if titration is needed, whether fixed or flexible dosing is most appropriate, and how to combine different elements of a complex treatment regimen.

The classical paradigm for dose finding in drug studies is the prospective, parallel, fixed-dose design. In analgesia, this is typified by NSAID development strategies, which provided the historical foundation for modern analgesic clinical research. This classic approach works well for drugs such as NSAIDs where there is little interindividual variability in the effective dose and a wide therapeutic index. Unfortunately, this method was assumed to be appropriate for all analgesics and has been applied unsuccessfully to analgesics with opposite pharmacology (ie, wide interpatient variability in the optimal dose and a relatively narrow therapeutic index). This became clear in the case of opioids where, despite clear evidence that this class of drugs cannot be used at fixed doses, investigators conducted fixed-dose, parallel studies and generally failed to demonstrate analgesia, contradicting thousands of years of known efficacy of the class. Other classes of analgesics that to some extent share these features include cannabinoids, gabapentinoids, and antidepressants. The issue of the fixed-dose paradigm is further complicated when studying pain syndromes with significant fluctuations such as osteoarthritis, where fixed doses of CNS-acting drugs may produce more side effects than benefit during times when patients have minimal pain. This is a good example of the principle that any drug can be made to look bad with the wrong study design. Drugs need to be studied the way they need to be used. Similarly, it is conceivable that different patients need different doses, or frequencies of administration, of nondrug treatments, such as cognitive–behavioral therapy[67] or neuromodulation techniques.[140] In summary, the fixed- vs flexible-dose paradigm, and how to measure the degree to which patients are using the assigned treatment, is an important issue for every trial.

Several alternatives to fixed dosing are available. A common option for CNS-active drugs is the "titration to a common fixed-dose" design,[130] where patients start at a low dose and titrate to a predetermined target for that treatment arm. This is the same as a fixed-dose design except with an initial titration period. Optimizing the titration method is critical and often gets short shrift during phase 2, which can lead to phase 3 failures or problems in the marketplace due to poor tolerability (eg, titration too fast), lack of efficacy (eg, peak dose too low), slow onset (eg, titration too slow), or lack of appreciation of interindividual differences (eg, needs customizable target doses). Some sponsors have attempted to remediate this issue by performing postmarketing studies to refine the dosage and administration regimen.[190]

The opposite of the fully fixed design is the fully flexible design, where patients can adjust their dose as needed to optimize efficacy and tolerability. This approach in a sense engages the patient as a partner in the drug development process because they are the best judge of their optimal dose. Flexible dosing is usually subject to constraints such as a maximum dose (which may relate to toxicology coverage), minimum dose, or frequency of allowed dose changes. In some cases, sponsors have endeavored to satisfy both patient needs for flexibility and drug development desires to evaluate dose–response by allowing flexibility within separate dose strata.[194] Flexible designs generally work better than fixed-dose designs for CNS drugs used to treat pain,[86,130,142,143,230] although this depends a great deal on the target dose in the fixed-dose group and how missing data are imputed. Fully flexible designs are not a panacea and can introduce new problems, especially when there is no obvious way for the patient to determine the optimal dose or when toxicities may be asymptomatic (eg, effects on liver function). Yet, another approach is to dose patients on an mg/kg basis or other method for individual tailoring based on factors that impact pharmacokinetics.

10. Improving reliability and decreasing failure risk of pain trials

Major sources of error in the results of pain trials are described below, along with evidence-based recommendations on how to remediate them (**Table 3**).

10.1. Outcome measure selection

In most clinical trials aiming to confirm that the treatment reduces pain, pain intensity is the primary endpoint. Pain intensity can be measured with single-item generic pain intensity scales, such as the 0 to 10 NRS or the 0 to 100 Visual Analog Scale, which are agnostic to the painful disorder being assessed (unless intentionally modified) and agnostic to any specific aspects of the pain experience that may be specific to that disorder. For some painful conditions, multiitem disease-specific measures of pain intensity are available. For example, the WOMAC Pain Scale is a 5-item measure of pain intensity in patients with osteoarthritis of the hip or knee.[15] Multiple studies have shown that this disease-specific pain intensity scale is a more responsive measure of pain in patients with knee OA than a single-item generic pain measure.[40] Unfortunately, multi-item pain intensity scales are not available for most chronic pain conditions; development of such scales represents an opportunity to reduce measurement error and improve assay sensitivity of chronic pain studies. In some cases, pain intensity is not the primary objective of the study, which may instead be pain interference with physical function, quality of life, or other domains that are important to patients or other stakeholders.[247] Selecting measures that accurately reflect these domains avoids measurement error because of a mismatch between primary objective and endpoint. Selection of outcome measures is further elaborated in the article by Patel in this series.[186]

10.2. Accurate pain and symptom reporting

A major factor contributing to the plethora of failed trials in chronic pain was the notion that because pain is subjective, there was no

Table 3

Sources of measurement error in clinical trials.

Source of error	Description	Mitigation options
Positive bias		
Allocation bias	Investigators choose which subjects go in which groups	Randomization
Expectation bias	Subjects report the response they expect (eg, pain relief)	Double blinding Placebo controls
Baseline imbalance in predictors of outcome*	Treatment groups differ by prognostic factors or treatment effect modifiers	Stratified randomization Adjusting for covariates
Observer bias	Whoever is observing the treatment effect reports the outcome they desire	Double blinding
Negative bias		
Errors related to patient selection		
Inaccurate diagnosis	Patient does not have the disease being studied	Central review of diagnostic assessment[161] Investigator training[161] Central review of diagnostic interviews[148]
Masquerading disorders	Patient has another disorder masquerading as the disorder being studied	Tools to identify masquerading disorders[255]
Inaccurate pain reporting	Patients differ in their ability to accurately report pain and other symptoms and can be trained to perform better Inaccurate pain reporters are also preferential placebo responders	Accurate pain reporting training Exclude patients with excess variability of clinical or experimental pain[63,64,239]
Placebo responders	Preferential placebo responders have higher than average responses to placebo, but not to active treatment	Select subjects whose attention is internally directed, eg, accurate pain reporters[240,241] Neutralize expectation across groups with expectation-based training of staff and subjects[64,73,269]
Baseline score inflation	Subjects/investigators may inflate baseline scores on measures when a minimum score is needed for enrollment; after randomization, scores decrease to true levels, mimicking a placebo response	Mask entry requirements Use different measures for the primary endpoint and for inclusion Statistical surveillance of baseline score inflation[63,148]
Unstable, resolving, and mild pain conditions	Enrolling patients with pain that is destined to resolve during the study, is highly variable, or is intermittent biases study to the null	Enroll patients with a history of at least 6–12 mo of moderate to severe chronic pain Do not worry about maximum pain duration Minimum pain intensity of 4–5/10 Ensure that the baseline period is long enough to establish a stable baseline
Psychological comorbidities and substance abuse	Patients with psychological comorbidities and substance abuse report pain less reliably, may be less compliant with study procedures, and compromise assay sensitivity	Exclude such patients based on established validated assessments, such as questionnaires and urine drug screens, unless specifically studying these populations[63,63]
Studying heterogeneous phenotypes	Treatments may not work on all pain mechanisms; studying a mixed group may result in failed studies when the drug is effective in a specific phenotype	Consider phenotyping all subjects at baseline and evaluate efficacy by phenotype[69]
Duplicate subjects	Patients often deceitfully enroll in the same study at multiple sites or in multiple studies at once, putting themselves and the study at risk	Use a duplicate subject detection service in every study[55,56,98,218]
Medical and treatment history	Patients are often unable to supply all relevant information about the past or current medical history and pharmacologic and nonpharmacologic treatments	Consider methods to import prescription monitoring data and electronic medical records data for enrolled subjects
Errors related to outcome assessment		
Insensitive outcome measures	Measures must not only be valid and reliable but also responsive to treatment differences; otherwise, differences will not be detected.	Choose the most responsive measure that is valid for the target concept. Prioritize disease-specific over generic measures. Do not use an instrument simply because it was used by a previous study. Consider developing a new measure if no suitable measures are available, or if there is reason to believe that a new measure would be substantially more responsive than available measures.[60,77,195]
E-diary noncompliance	E-diary compliance is poor in many studies despite assurances by vendors.	Insist on a system of automated reminders to subjects who miss entries, alerts to coordinators for all missed entries, calls from coordinators to subjects after every missed entry, real-time documentation of those calls, and real-time central monitoring of all elements of the system. Always have back-up in-clinic assessments of the primary endpoint.

(*continued on next page*)

Table 3 (*continued*)

Sources of measurement error in clinical trials.

Source of error	Description	Mitigation options
Errors related to dosage and administration of study medication		
Failure to reasonably establish safe and effective dose before phase 3	Some programs enter phase 3 without adequate dose-finding studies	Plan for enough phase 2 studies to characterize the dose–response relationship, determine whether fixed or flexible dosing is optimal, and decide upon the frequency of administration before phase 3
Failing to measure adherence accurately	Nonadherent subjects cause studies to fail. Pill counts are not valid as assessments of adherence.	Use computerized packaging or other electronic means of documenting adherence to each dose.
Poor adherence to study treatment	Little is done in most studies to encourage adherence, which may be the greatest opportunity for improving trial success.	Select patients who are adherent during a prerandomization period. Effective subject and staff training at screening and periodically thereafter. Coordinator calls to subjects who miss any doses; real-time central monitoring of compliance with those calls; periodic feedback to subjects about their adherence.[21,26,51]
Errors related to confounding subject activities during the study		
Concomitant and rescue medications	Use of concomitant and rescue treatments (pharmacologic and nonpharmacologic) can bias results	Train subjects and staff on standards for handling concomitant/rescue medications. Minimize concomitant/rescue treatment when feasible Track each dose of rescue medication as if it was a study drug. Prespecify in the protocol how concomitant/rescue treatment will be quantified and accounted for analytically.[63,212]
Failing to train subjects effectively	Subjects need to follow the protocol, particularly medication adherence, diary compliance, accurate symptom reporting, and steady physical activity.	Perform a data quality risk assessment followed by a subject training needs assessment. Follow principles of developing and deploying validated training.[241]
Physical and psychological treatments	No new interventions should begin during studies. Established programs should continue unchanged.	Provide structured guidance to subjects about physical activity; consider structured online support. Capture changes in physical and psychological activity/function using questionnaires; consider objective measures such as actigraphy.[244]
Errors related to site selection and management		
Overly heterogeneous sites or regions	Heterogeneity in healthcare systems, language, culture, availability of alternative treatments, and other variables introduces variability to the treatment effect.	Minimize the number of sites; invest in prestudy recruitment activities to maximize the number of patients per site. Minimize heterogeneity in sites and regions. Carefully consider differences among sites that may impact outcome and control these extraneous factors to the extent possible.
Selecting sites based on unverified patient databases	The basis for recruitment in most chronic pain studies is site databases; however, most are exaggerated and out of date.	Invest in prestudy patient identification activities such as patient registries, prescreening studies, and database verification before finalizing site selection.
Variability in study conduct by sites	Sites implement protocols in varying ways that may be difficult to predict, describe, or understand. The more sites, the more variability.	Minimize the number of sites. Invest in prestudy activities to maximize the number of patients per site. Perform a data quality risk assessment and a site training needs assessment. Develop and deploy validated training for sites. Central statistical monitoring and intervention.[81,121,139]
Errors related to study design		
Longer observation periods	Assay sensitivity degrades over time. There is a tension between duration of observation that is most clinically relevant (long) and cleanest from a measurement error perspective (short).	Prespecify primary endpoints at the earliest time point that is acceptable from a clinical, scientific, and regulatory point of view. Carefully implement all feasible measures to reduce types of measurement error that increase over time, such as adherence and nonprotocol treatments.[63,64]
"Hockey stick"	Studies are often positive up to the last week when the primary endpoint is determined then fail. Presumed cause is expectation aroused by the last treatment visit.[161]	Blind investigators and subjects to the time point of the primary endpoint. Choose a time point for the primary endpoint before the final visit, eg, a 14-wk trial that determines the primary endpoint at week 12. Accurate pain reporting training and placebo response reduction training.

(*continued on next page*)

Table 3 (*continued*)

Sources of measurement error in clinical trials.

Source of error	Description	Mitigation options
No. of arms, allocation ratios	Studies with a higher probability of assignment to active treatment have higher placebo responses and smaller differences between active treatment and placebo.	In 2-arm studies, use a 1:1 randomization ratio, unless there is a strong reason not to. Expectation-based training to neutralize expectation of benefit regardless of the allocation ratio.[64,73,269]
No. of visits	Some evidence suggests that higher numbers of visits lead to larger placebo responses and smaller between-group differences.	Some have recommended minimizing the number of study visits. Use visits to deliver standardized neutral expectation messages
Active comparator	The most definitive method to assess the integrity of a study is to measure the efficacy of an active comparator vs placebo.	Include an active comparator whenever possible. Consider allocating fewer subjects to the active comparator than study drug.
Noninferiority studies	Noninferiority studies are not scientifically valid without an internal demonstration of assay sensitivity and controls for nonspecific factors, particularly in medical device studies.	Avoid noninferiority studies except in highly specific circumstances; incorporate internal demonstrations of assay sensitivity.
Errors related to data quality control		
Failure to perform a data quality risk assessment	Data quality begins with identifying potential threats to data quality and implementing preventive actions, per regulations.	Perform a data quality risk assessment at the protocol synopsis stage.[139]
Failure to implement effective training	Clinical trials that allow participants to perform activities that influence the primary endpoint require training to calibrate these activities.	Perform a training needs assessment, design and deploy training for key activities where human performance may vary, and evaluate effectiveness and modify as needed.
Consistency and reliability of site and subject performance	Regulations require central monitoring of variables known to influence the reliability of study results and timely corrections. In pain studies, these include pain variability, consistency of different measures, diary compliance, medication adherence, and others.	Select and monitor variables that impact study results, not just regulatory compliance. Have a system in place for timely and systematic root cause analysis and intervention. Monitor and document outcomes of these activities.

Types of measurement error are roughly divided into those that inflate the true effect size of treatment (positive bias) and those that shrink it (negative bias). Note that some sources of error can produce either positive or negative effects.

* Can produce a positive or negative bias.

way to know whether the patient was reporting their pain accurately, or worse, that a patient's pain report was inherently accurate. Although the concept that patient symptoms should not be ignored is a foundation of compassionate care, subjectivity does not require that the report be accepted as accurate. A series of studies[165,224,239–241] has now emerged demonstrating that patients differ in their ability to report experimental pain accurately; these differences correlate with clinical pain reporting variability; variability in reporting either experimental pain or clinical pain predicts responses to placebo and ability to discriminate drug from placebo; and pain reporting accuracy can improve with training.

The first study to investigate the reliability of pain reporting used an assessment called the Focused Analgesia Selection Test, which consists of a battery of brief noxious thermal stimuli applied to the patient's forearm.[239] This study demonstrated that about one-third of patients with osteoarthritis of the knee reported experimental pain in a highly inconsistent manner.[239] The Focused Analgesia Selection Test was then used to exclude "poor pain reporters" in a study comparing mavatrep, an investigational analgesic, with naproxen and placebo in patients with osteoarthritis of the knee. The study went on to demonstrate significant superiority of both active treatments to placebo in 33 participants.[165] The largest differences were observed among participants who most accurately reported experimental pain. This was confirmed in a subsequent study: an RCT of an accurate pain reporting training program demonstrated that training not only improved the accuracy of experimental pain reporting but also reduced clinical pain variability and the response to placebo (but not to treatment), thus greatly improving assay sensitivity in the trained cohort.[241] Another trial of accurate pain reporting training performed by the ACTTION group reported improvements in some indices of pain reporting accuracy, but because the training was not performed in a drug vs placebo clinical trial, it is unclear whether the training would have improved effect sizes of treatment.[224]

Another factor to consider is timing of pain intensity reporting. In certain disorders, pain has a predictable diurnal pattern, although this may differ from patient to patient.[17,180] For studies focusing on "pain right now," the time of day pain is measured should be standardized. This also applies to studies with longer recall periods because patients' perceptions of their pain over recent periods are influenced by their current pain.

Current best practices for optimizing pain reporting are to measure daily pain intensity over a 1- to- 2-week baseline period and exclude patients with "excessive" variability and other abnormal patterns and provide an accurate pain reporting training program. These approaches appeared to decrease the placebo response in 2 studies, which will be discussed further below.

10.3. The placebo response

The placebo response has achieved notoriety as a leading culprit responsible for failed trials, beginning with Beecher's landmark paper, "The Powerful Placebo," in 1955.[13] The "placebo response" refers to the reduction in symptom intensity observed among patients assigned to placebo. Several factors contribute to this decline in symptom intensity (**Fig. 3**). Pain severity among patients with chronic pain goes up and down over time, better some months, and worse other months. Should such a patient enroll in a trial during a time when their pain is severe, just based its cyclic *natural history* it would be expected to decline whether they were in a trial or not. When such patients enroll in a clinical trial at the peak severity of their symptoms, it can be expected

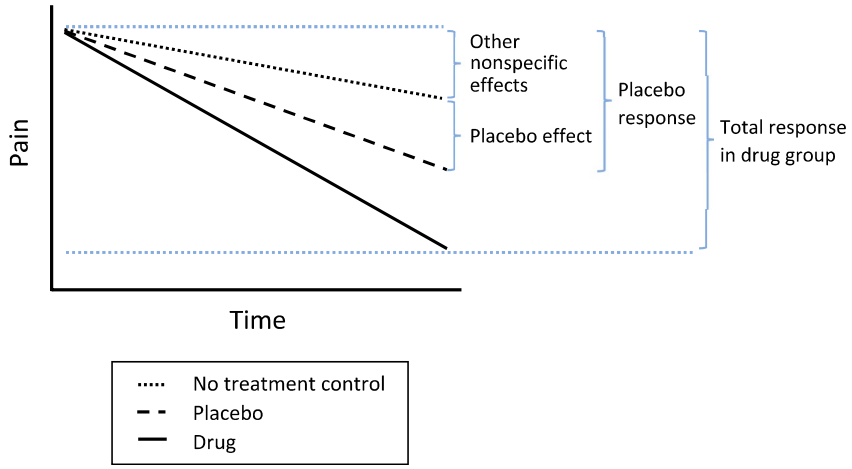

Figure 3. Anatomy of the placebo response: the additivity theory. The total response in the drug group is the sum of the effects of the sham treatment itself (the placebo effect) plus the effects of the treatment context (eg, attention from nurses) plus the specific pharmacologic effect of the drug. These effects have been explored using 3-arm studies (drug vs placebo vs no treatment).

their pain will decline (during the study), which manifests as a decline in pain in all groups (placebo and active treatment), and will mimic a placebo effect. One might refer to this phenomenon as "clinical regression to the mean"—such disease-based fluctuations occur in a number of chronic pain syndromes, such as osteoarthritis, rheumatoid arthritis, and chronic pancreatitis. A similar phenomenon is statistical regression to the mean. For example, a patient with chronic low back pain may have an average daily pain of 3/10, which fluctuates up and down around that mean, day by day. If such a patient enters a clinical trial on a day their pain happens to be 5/10, thus meeting minimum entry criterion for pain severity for that clinical trial, it can be expected that their pain intensity will decrease, although on average their pain intensity over time has remained stable. A third component, the treatment context, can have a variety of influences on patients' reported pain intensity, for example, as a result of attention from nurses, physical examinations, and physician body language, which in effect is part of the placebo response induced by the treatment context. Finally, there is the *placebo effect* attributed to the inert treatment itself, whether a sugar pill, fake injection, sham acupuncture needle, or other imitation of actual treatment. All the above-mentioned effects are collectively referred to as *nonspecific effects* because they are not produced by any specific pharmacologic or physiologic effect of the investigational treatment. The *additivity theory* posits that the total response in the treatment group is a simple sum of the specific effects of treatment plus the sum of these nonspecific effects (**Fig. 3**). The magnitude of the placebo effect itself has been documented by no-treatment control studies, where patients are randomized to active treatment, placebo, or no treatment at all (patients undergo all study procedures but do not get the sugar pill).[151]

Perhaps the most important concept to clarify is the "placebo responder" vs the "preferential placebo responder." For decades, researchers have searched for identifiable characteristics of the placebo responder, which is the person who will have a higher-than-normal response to placebo. This may be of interest to placebo researchers but is of little relevance to clinical trials: if a patient has a high response to placebo and an equally high response to the active treatment, there is no problem in terms of assay sensitivity. Instead, clinical trialists are interested in identifying characteristics of the preferential placebo responder—the patient who shows a higher response than normal to placebo but not the active treatment, thus failing to discriminate.

Studies with a high placebo response are less likely to show a difference between a truly effective treatment and placebo,[59,129] making the placebo response a source of measurement error on a trial level. Attempts to overcome the effects of placebo responses by inflating the sample size are generally in vain; even if there are enough patients to achieve a $P < 0.05$ threshold, the observed net treatment effect will still be small, which will undermine the position of the treatment in meta-analyses, treatment guidelines, and reimbursement decisions. The association between high placebo response and small effect size or trial failure has been demonstrated in multiple indications including chronic pain.[59,63,64,143,144] From a historical perspective, there appears to be an increase over time in placebo responses (but not active treatment responses) in studies of chronic pain.[248] Just as problematic is the variability of the placebo response, which can range from 0% to 100% across chronic pain studies,[198] making it virtually impossible to reliably plan the sample size in a clinical trial based on the placebo response observed in another trial. Instead, sample size calculations should be based on the minimum between-group difference that is important to detect.

The traditional explanation for the placebo effect is patient expectation: you get what you expect. According to this concept, aspects of the therapeutic context such as the demeanor, appearance, and reputation of the physician; the physical environment; the invasiveness of the treatment; information provided to the patient; and the body language of those in the environment all serve to create an expectation of symptom reduction in the mind of the patient, which in turn triggers various neural mechanisms that actually reduce symptom intensity. Many studies support this thesis. Comprehensive reviews of the placebo literature are available elsewhere.[18,251]

Considering the paradigm described above, it has been suggested that "neutralizing" expectation may reduce the placebo response without reducing the response to active treatment. A few studies have examined this hypothesis. The largest was an RCT of montelukast vs placebo in over 600 patients with asthma.[264] Patients were randomly assigned to either a "high expectation" group, where they were exposed to inspiring television advertisements, a professional-appearing

physician, and positive messages about the treatment, or a "neutral expectation" group, where they did not see the advertisement and had a casual-appearing clinician. Patients were subsequently randomized to receive either active drug or placebo. Patients in both expectation groups had similar responses to active treatment, but patients in the high expectation group had large responses to placebo (resulting in no discernable difference between active treatment and placebo), whereas patients in the neutral expectation group had much lower responses to placebo (resulting in a statistically significant difference between drug and placebo; **Fig. 4**). Although the drivers of the placebo response are likely to be more complex than accounted for by the expectation theory,[128] this study and others like it[127,234,241] have led to the development of placebo response reduction training programs and other interventions to keep expectation neutral to improve assay sensitivity. Available reports on their effectiveness suggest that they perform as expected.[73,241]

Current best practices to consider include identifying and excluding patients with high variability in daily pain scores (or experimental pain) because this predicts preferential placebo responsiveness; using an accurate pain reporting training program because this has been shown to decrease both pain variability and preferential placebo responsiveness; and creating a neutral expectation environment though placebo response reduction training and other controls to decrease the external cues that drive an expectation of benefit. Placebo run-in periods have been used extensively in other therapeutic areas but do not seem to improve postrandomization effect sizes.[59,144]

10.4. Treatment adherence

Measuring and improving adherence represents one of the most obvious and achievable opportunities for improving clinical trial performance. If patients do not use the study treatment, it will not work—and often they do not. It has been estimated that variable adherence vies with pharmacokinetics as the leading source of variation in drug response in ambulatory settings.[112] The purpose of confirmatory trials is to confirm efficacy and safety *at a specified dosage and administration regimen*, and this applies to any type of treatment. If patients do not use the treatment as prescribed, confirmatory trials do not confirm anything. Patients in general are poorly adherent in clinical trials (see further below), contributing to inaccurate estimates of both safety and efficacy of the nominal regimen, failures to transition from phase 1 to 2 or phase 2 to 3, bad go/no-go decisions, misleading labeling, and various postmarketing problems. Moreover, adherence in clinical trials is not usually measured using reliable methods, and methods to promote adherence are seldom incorporated into a clinical trial design.

A meta-analysis of adherence among more than 16,000 patients in 95 clinical trials[21] found that 4% of patients never started the study drug, and adherence declined steadily for the duration of follow-up. At day 100 (about the time the primary endpoint is captured in most pivotal trials), less than 70% of patients were taking the study medication as directed. In a study of cancer pain (where one would think that patients would be highly adherent), Caucasian patients were 73% adherent and African Americans 53% adherent using an electronic method for monitoring adherence.[167] In a phase 1b trial, 70% of completers were considered compliant, and only 39% took every dose of the study medication (as assessed by pharmacokinetics), although pill counts indicated 92% compliance.[42] In a phase 3 study of 634 patients with chronic low back pain, only half of the participants were compliant at week 12. The overall study failed to meet its primary endpoint but was positive at $P < 0.01$ in the compliant subgroup across multiple endpoints; these data indicate the possibility that patients who did not take their medication as directed did not experience full benefit, but these comparisons are not of randomized groups and are subject to potentially significant confounding.[42] Once again, pill counts indicated high compliance. Studies in other therapeutic areas have demonstrated similar patterns: failed primary study, apparent compliance by pill counts, objective evidence of poor compliance by other means, and efficacy in the compliant subgroup.[12,26] Although analysis of subgroups based on compliance is associated with a number of flaws, unless reasonable methods for causal inference are applied that attempt to deal with confounding, the fundamental point remains that treatments do not work in patients who do not use them.

Missed drug is not the only problem; another important issue is the pattern of missed doses. The same amount of missed doses can be due to delayed initiation, early discontinuation, sporadic missed doses, or a drug holiday, each with a different effect on the primary endpoint (and all impossible to ascertain without a reliable method for monitoring the intake of each dose) or on adverse events.[21] One might think that adherence would be excellent for drugs designed to treat life-threatening conditions or drugs unburdened by significant side effects; however, this is incorrect. In one trial, a substantial number of patients were poorly adherent to a relatively unobtrusive curative drug for an otherwise uniformly fatal form of leukemia, and adherence to antiviral treatments in clinical trials for HIV disease was poor in several studies.[21]

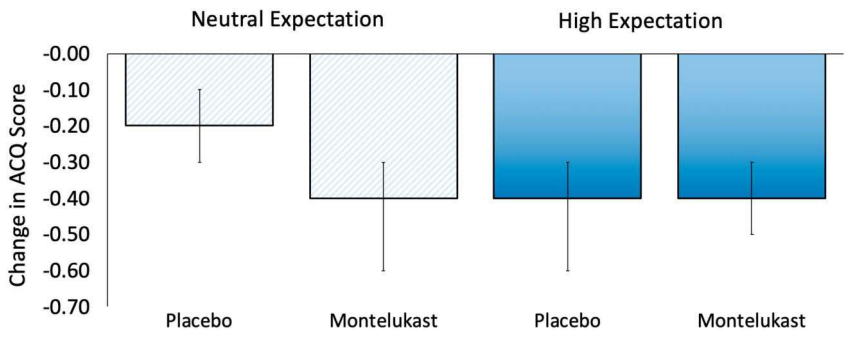

Figure 4. Influence of expectation on subjective outcome measures in asthma: Patients were randomized to "neutral" vs "high" expectation in a large randomized controlled trial of montelukast (Singulair) vs placebo in asthma. Drug vs placebo differences (subjective outcome measure, the Asthma Control Questionnaire [ACQ]) were largest in the neutral expectation condition and disappeared in the high expectation condition (adapted from Wise et al.[264]).

Variable adherence makes it difficult to estimate the safety and efficacy of the treatment if it were to be used as directed.[188] Poor adherence may lead to failure to demonstrate true efficacy, which may be a source of failure to translate from phase 1 (where all doses are supervised) to phase 2 (where patients generally take responsibility for taking their own medications) and to phase 3 (which generally involves more heterogenous patients). Risks may be underestimated as a result of nonadherence: if patients were taking full doses, there may have been more AEs. This may be one reason for the observation that 10% to 20% of approved drugs undergo postmarketing dose reductions due to safety issues. Poor adherence can also lead to increased AEs, for example, if patients experience withdrawal or rebound symptoms when they skip a few doses or stop taking the drug entirely, or when patients restart after a holiday without titration ("recurrent first-dose effects"). Poorly adherent patients may also fail to meet responder criteria during the enrichment phase of enriched studies or may initially respond when they are taking the medication and show less response as adherence wanes. Poor adherence can also muddy the dose–response relationship. Adherence is not only important for study medication; adherence to the prescribed rescue medication regimen and accurate documentation of rescue medication consumption are also important for study interpretation and at times for computation of the primary endpoint.[4,134]

Several methods for measuring adherence are available, each with advantages and disadvantages.[124,259] The 3 cardinal characteristics of any method for measuring adherence are *accuracy* (does it measure true dose intake?), *sampling density* (what proportion of doses are measured?), and *obtrusiveness* (how burdensome is it?). *Measuring drug concentrations in body fluid* is an accurate method for measuring medication use but does not inform the exact time of administration; multiple studies have demonstrated "white coat compliance" where patients self-administer their medication right before clinic visits but seldom at other times.[193] Monitoring drug levels at key visits is still invaluable because the primary endpoint is typically captured at the final visit, many analgesics can exert a therapeutic effect with just a few doses, and plasma concentrations have been used to identify patients who discriminate drug from placebo in otherwise failed trials.[42] Plasma concentrations can also indicate errors in randomization or drug supply that would not have been identified by other means. *Pill counts*, counting the number of dosage units returned by patients and calculating the deviation from expected, is an essential study procedure. Unfortunately, pill counts frequently overestimate adherence by as much as 40%.[192,196] Therefore, although counting pills is a regulatory expectation, and appreciable discrepancies between the number and the expected number of returned pills indicate an issue, counting pills is by itself not an accurate measure of adherence. Unsurprisingly, *patient self-report questionnaires* overestimate adherence and also cannot be relied on.[258] For example, in one study of antiretroviral medications for HIV disease, adherence measured by patient interview was 93%, whereas that measured by a computerized medication bottle in the same patients was 63%.[157]

Electronic packaging involves incorporating microcircuitry into packages of solid oral dosage forms or other formulations to detect, time-stamp, and record when patients remove a dose from the packaging. Although these methods measure "pill to hand," they are reasonably accurate measures of "pill to mouth."[257,259] When the packaging burden is modest, patient burden is minimal; however, bulky packaging can burden patients, leading them to remove and store multiple doses at a time, defeating the purpose of the packaging. Theoretically, every dose is monitored. Time and cost of packaging must be accounted for in study start-up activities, and these methods are not inexpensive. Electronic packaging can be considered the current gold standard for adherence monitoring in ambulatory trials.[26]

Electronic diaries (phone-based or hand-held devices) are commonly used in pain studies to record pain and other symptoms as well as medication consumption, either at the time of dosing or as a 24-hour summary. The daily use of a hand-held device is a substantial burden. If patients are already using such a device to enter clinical data, the additional burden of a single nightly entry for medication consumption is trivial. Requiring data entry at every dose is more burdensome. To the author's knowledge, no study has yet investigated the accuracy of e-diaries for tracking medication consumption. Accuracy has been demonstrated indirectly by separation between drug and placebo on the endpoint of rescue medication consumption in clinical trials. However, rescue medication consumption recorded in e-diaries often fails to separate drug from placebo, and thus, the accuracy of this method of measuring rescue medication use remains uncertain.

10.4.1. Smart ingestible sensor (pill)

A microcircuit can be integrated into a pill that is activated by gastric acid to generate a weak radio signal containing information about the timing of ingestion. This signal is then amplified and retransmitted to a more distant source, such as a smart phone, through a signal-detection patch worn on the patient's abdomen.[259] Several studies have demonstrated the usability and accuracy of such systems.[14,70] Patient burden is high because patients need to wear a skin patch that must be tracked, changed, and can cause skin reactions and carry a device to capture and retransmit signals. In addition, this method is only suited for solid oral dosage forms, and setup involves substantial effort. Further research is needed to determine whether this technology has advantages over simpler electronic techniques.

Photographic documentation of pill intake requires patients or caregivers to photograph the suitably identified medication sitting in the patient's mouth; some approaches even record the sound of swallowing. This system appears least prone to error because it is the only direct measure of "pill to mouth." Such systems impose a substantial burden on patients who must take and transmit a photograph of every dose, and it is unclear to what extent patients comply with these systems over time. Several clinical trials have been performed on at least 1 such system and suggest improvements in patient compliance.[9,152]

A long list of approaches to improve adherence has been subject to multiple systematic reviews.[51,115,150,257] Remarkably, most clinical trials of interventions to improve adherence have not used reliable methods to measure adherence[21]; nonetheless, it is possible to draw some conclusions from studies that did use reliable methods.[51] There are essentially 2 options: selecting patients who are already adherent under the assumption that they will continue to be adherent and improving the adherence of patients who are already enrolled.

The author is unaware of any studies comparing prerandomization with postrandomization adherence. Nonetheless, it has become increasingly common to measure adherence before randomization and exclude poorly adherent patients.[107,166] It is uncertain whether the typical 1-week baseline is sufficient to establish a "trait" of nonadherence; nonetheless, in the absence of more data, this approach is recommended.

In rigorous studies of methods to improve adherence, all interventions have worked to one degree or another.[51,176] The most effective interventions tended to provide patients with feedback on their actual adherence using electronically compiled dosing histories and cognitive educational interventions.[51] Other interventions also produced some effect, including treatment simplification, behavioral counseling interventions, social-psycho-affective interventions, reminder systems, physiologic feedback, and rewards. Multicomponent interventions did best.[176]

In summary, measuring and improving adherence may be the single greatest opportunity for improving the assay sensitivity of clinical trials. Adherence should be planned into the study design and must be monitored using technologies that are accurate, provide dense sampling, and are minimally burdensome. At present, electronic packaging appears to be the best available solution, although photographic techniques are promising. Multicomponent approaches to support adherence work best, including exclusion of poorly adherent patients before randomization, communication to patients by the investigator about the importance of treatment adherence, individualized adherence support programs, regular feedback, troubleshooting of adherence issues, automated reminders, and prompt and documented interactions with study coordinators after any lapses in adherence.

11. Experimental noise and confounding

Confounding, from the Latin word *confundere*, to confuse, occurs in clinical trials when the observed outcome is influenced by something other than the treatment (ie, a confounder or confounding variable). The term "confounding" has been used in the literature to describe several distinct concepts.[45,106,188,235,250] Strictly speaking, a variable should meet 3 criteria[250] to be a confounder: (1) it must be associated with the treatment; (2) it must be associated with the outcome; and (3) it cannot be a mediator of the treatment effect. An obvious source of confounding is baseline imbalances between treatment groups in *variables that are associated with the outcome*; if one group contains fewer participants with a clinical characteristic that is associated with a poor outcome (eg, disability in a back pain study), the outcome will be better in that group even in the absence of a treatment effect. Although randomization balances groups for known and unknown baseline prognostic factors as studies achieve sufficient size, this objective is not always achieved, especially in small studies.[88,106] Baseline covariates that may be associated with outcomes of analgesic treatments include age, sex, race, severity of illness, and comorbidities. For example, if low back pain has a better prognosis in patients with a short duration of prestudy pain, and the treatment group has substantially more patients with brief prestudy pain duration, this difference may be sufficient to explain any observed advantages of treatment over control. Reflecting back on the 3 criteria for a confounder: brief prestudy pain duration was associated with treatment (more patients with brief prestudy pain duration in the treatment group) and the outcome (brief prestudy duration was associated with favorable outcome), but prestudy pain duration could not have been a mediator of the treatment effect because it was already present before randomization.

In the above example, the confounding variable was present before randomization. Much of the literature on confounding insists that confounding variables must be present before treatment, and postexposure factors associated with both exposure and outcome cannot be considered confounders but are potential mediators.[253] For example, let us say that a drug that relieves pain also helps patients sleep better, and, as is well known, sleep improves pain. The drug thus may relieve pain in 2 ways: first by a direct analgesic effect and second by an indirect effect improving sleep (and subsequently pain). Sleep is then a partial mediator of the effect of the drug on pain. Reflecting back on the 3 criteria for a confounder, sleep is not a confounder: improved sleep was associated with treatment and outcome (pain) but is a partial mediator of the treatment effect. Therefore, it would not make sense to adjust for sleep in a regression model as a method to reduce bias in estimating the treatment effect. On the contrary, this would increase bias and inappropriately reduce the estimated treatment effect. However, researchers who are interested in the direct mechanism by which the drug relieves pain might perform such analyses (often called mediator or path analyses) to understand how the drug works.

Nonetheless, true postrandomization confounding has been described in the literature. An example comes from a hormone replacement therapy study: patients on placebo had a higher rate of statin initiation during the study, which confounded the interpretation of primary cardiovascular outcomes.[160] Reflecting back on the 3 criteria for a confounder, statin initiation was associated with treatment (more likely in placebo patients) and cardiovascular outcomes but was certainly not a mediator of the effect of hormone replacement. Although it was possible to demonstrate analytically that the primary interpretation of the studies would not have changed, the confounding was still problematic.

Several types of postrandomization confounding occur in pain studies. Rescue medication consumption is a ubiquitous example: patients on placebo generally have more pain and use more rescue medication, leading to a decrease in pain. Reflecting on the 3 criteria for confounding, rescue medication is associated with treatment and outcome but is not a mediator of treatment effect (ie, analgesics do not relieve pain by causing increased rescue medication consumption). This confounding can be partially mitigated by design, but residual confounding usually remains and is a source of bias in estimating treatment effects. It can be difficult to determine whether a third variable is a confounder, a mediator, both, or neither; such a determination depends on previous knowledge and on clinically informed causal models.[118,119,187,250]

Other variables that do not conform to the strict definition of a confounder can also influence the interpretation of study results; these influences will be referred to as "experimental noise." Numerous factors influence a patient's pain intensity, even when it is assumed to be measured without error. For example, a patient who slept poorly the night before a clinic visit or who had to walk a long way from the bus stop to the clinic entrance might report high pain intensity because the pain intensity is truly high. In this case, there is no measurement error at the level of the individual pain assessment, and neither variable meets the formal definition of a confounder. Nonetheless, these variables introduce error in measurement of the effect of a treatment compared with placebo at the study level, by adding variability to the measurement of treatment effect, with the potential of attenuating the observed between-group difference.

Control of confounding and experimental noise begins in the study design phase. Randomization already puts the trialist into a relatively secure position in relation to the epidemiologist. Stratification of randomization by site and covariates known to have an appreciable effect on study results helps to ensure that important baseline covariates are equally distributed between treatment groups and across sites. Unfortunately, it is impossible to stratify for every variable that could potentially influence an

outcome, and attempting to do so can do more harm than good.[106] Adjusting for important covariates is an important part of the statistical analysis plan and should be based on prespecified methods.

Adequate control over factors that influence pain intensity is a topic that draws substantial attention in study design. Inpatient studies are particularly suited for controlling these factors, as the investigator and staff have some control over the patient's diet, sleep, activity, and concomitant treatments. Yet, keeping patients in a ward for the typical duration of a confirmatory study is impractical, and exerting tight control over the activities of "free range humans" is unrealistic. Protocols often attempt to achieve this goal by attempting to exclude patients who cannot commit to keeping their physical activity constant or avoid extraneous treatments, but these criteria are often vague and rarely enforced or documented. It is surprising that there have been so few attempts to keep patients domiciled for critical portions of studies, eg, 48 hours at baseline and at the end of treatment.

12. Data quality and central statistical monitoring

12.1. Overview

Although there has been much talk about the importance of data quality, there is little clarity on what data quality is. A useful definition comes from an Institute of Medicine Roundtable[44] that was subsequently echoed by FDA contributors,[146] which defined high-quality data as "data strong enough to support conclusions and interpretations equivalent to those derived from error-free data." In other words, *data quality is the minimization of measurement error*, which has been the focus of this article. Although there is overlap between data quality and regulatory compliance, they are not the same.

Data quality begins with designing a protocol that combines scientific rigor with operational simplicity, followed by an analysis of the potential risks to data quality and how they will be prevented, detected, and resolved. Each data quality risk should prompt consideration of whether the protocol can be augmented to support quality checks and mitigation steps. For example, if application of a clinician-administered diagnostic assessment may be associated with reliability issues, then procedures should be added to the protocol, such as having 2 raters perform a sample of assessments or having a central verification process. Once the trial is underway, procedures designed to support data quality include audit checks in electronic data capture systems, queries for implausible or missing data, and surveillance of variables that have been shown to impact the ultimate study results. Vendors of specific services such as labs and radiology generally have their own data quality control procedures in place. Ironically, the one place where attention to measurement error is not routine is in the capture of clinical endpoints, which are almost always the primary endpoints of clinical studies.

Traditional procedures for monitoring the quality of clinical data consist of sending monitors to sites where they check on general conditions and perform source document verification (SDV). Source document verification addresses one source of measurement error: transcription errors from so-called source documents to electronic data capture systems. Complete SDV is not cost-effective, is prone to human error, and provides at best marginal improvements in data quality.[6,98,246] For this reason, regulators have encouraged the implementation of risk-based monitoring,[81,121] which was originally narrowly interpreted as a method for improving the efficiency of human monitoring (ie, saving money) by allocating monitoring visits and SDV to places where the highest risks to data quality were expected (eg, sites with high enrollment). This narrow approach to risk-based monitoring has largely failed because (1) anticipated cost savings have not been realized; (2) both pharmaceutical companies and regulators are often too conservative to sleep well at night without 100% SDV; and (3) SDV does little to assure data quality in the first place because the greatest sources of measurement error are not addressed by SDV.

A relatively new approach articulated in recent regulatory guidelines is Central Statistical Monitoring (CSM)—when actions to improve clinical data quality are added to this monitoring, terms such as Risk-Based Quality Management are used.[139] Central Statistical Monitoring can detect performance issues at the level of the site, participant, and study.[52,68] These approaches have been used to detect not only low-performing sites but also fraudulent sites.[91,145,181,195,236] Moreover, evidence has demonstrated the ability of CSM to identify many of the errors that are usually identified using on-site SDV[10]: in a comparative study, CSM performed as well as complete SDV.[27] However, the main value of CSM is not human resource allocation or SDV but timely detection and remediation of data quality problems that cannot be detected any other way.

Regulatory guidelines that are currently in force require the following components of a CSM program[81,121]:

(1) Identify the processes and data that are critical to ensure the reliability of trial results;
(2) Identify risks to critical trial processes and data, including organizational-level risks;
(3) Consider the likelihood of errors occurring, the extent to which such errors would be detectable, and the impact of such errors on the reliability of trial results;
(4) Establish predefined quality tolerance limits;
(5) Detect deviations from predefined quality tolerance limits and trigger an evaluation to determine whether action is needed;
(6) Assess the effectiveness of quality management activities;
(7) Document everything in the clinical study report.

12.2. Identification of critical processes and data

Selecting variables for CSM is the foundation of effective central monitoring. If one is interested in the defect rate of car engines from a manufacturing plant, it is most important to measure variables that directly impact the performance of the engine, such as cylinder bore. One could (and should) measure the rate at which cars are recalled for defects, but at that point the problem is too far gone to fix. In a hospital, the goal may be to minimize postoperative wound infections; while measuring rates of infection is important on its own, it is too late for prevention once the wound is infected. Instead, one could monitor *predictors* of wound infection, such as nonadherence to wound-shaving protocols and operating times. In clinical trials, waiting until the end of the trial provides information that is too far downstream to allow corrective action. Instead, we must monitor *predictors* of the reliability of trial results as early warnings about quality problems that ultimately place the entire study at risk.

What variables predict the reliability of study results? Although this science is in its infancy, examples have been cited throughout this article and include excess pain variability, poor medication adherence, duplicate patients, baseline score inflation, and high expectation of benefit. One group presented a comprehensive CSM method[139] which included a list of such variables that was developed by expert consensus and subjected to validation. The list included items such as extremely high or low symptom variability, e-diary compliance, and measure discordance. This

list is a good starting point when considered together with other variables relevant to a particular study.

12.3. Statistical process control methods for aberrancy detection

Once variables have been selected for monitoring, the next question is: what is the best way to monitor them? One useful approach, which has become ubiquitous in manufacturing quality control, is called "statistical process control (SPC)." Statistical process control was introduced by Walter Shewhart, an engineering statistician at Bell Laboratories in the United States, and his protégé W. Edwards Deming, who first achieved successful adoption of SPC in Toyota automotive manufacturing.[49,217,254] Statistical process control combines sequential, time-based analysis methods with graphical presentation of data in a "process control chart," which allows real-time determination of whether variation in an ongoing process is attributable to random fluctuations, or represents a systematic change over time, or difference from other units of assessment (eg, sites and participants). A systematic change would indicate a potential quality control problem that merits further attention or corrective action.[19,254]

There are different types of control charts to suit different purposes. **Figure 5** presents example control charts for e-diary compliance from a single site in a multicenter trial of a treatment for osteoarthritis of the knee. When the variable crosses the upper or lower "control limits," it is regarded as being "out of control," which should prompt an investigation. As illustrated in the figure, different control charts have different performance characteristics in terms of sensitivity, specificity, and time to detection of loss of control.

12.4. Root cause analysis of performance issues

The most effective treatment of a performance issue follows a correct diagnosis of the cause of that issue. Root cause analysis has been extensively used to diagnose performance problems in other industries, including health care.[262] Multiple resources and reviews are available on the root cause analysis process,[34,35,38,109,204] which is designed to identify not only *what* occurred and *how* it occurred but also *why* it occurred. For example, patient noncompliance with electronic diaries is

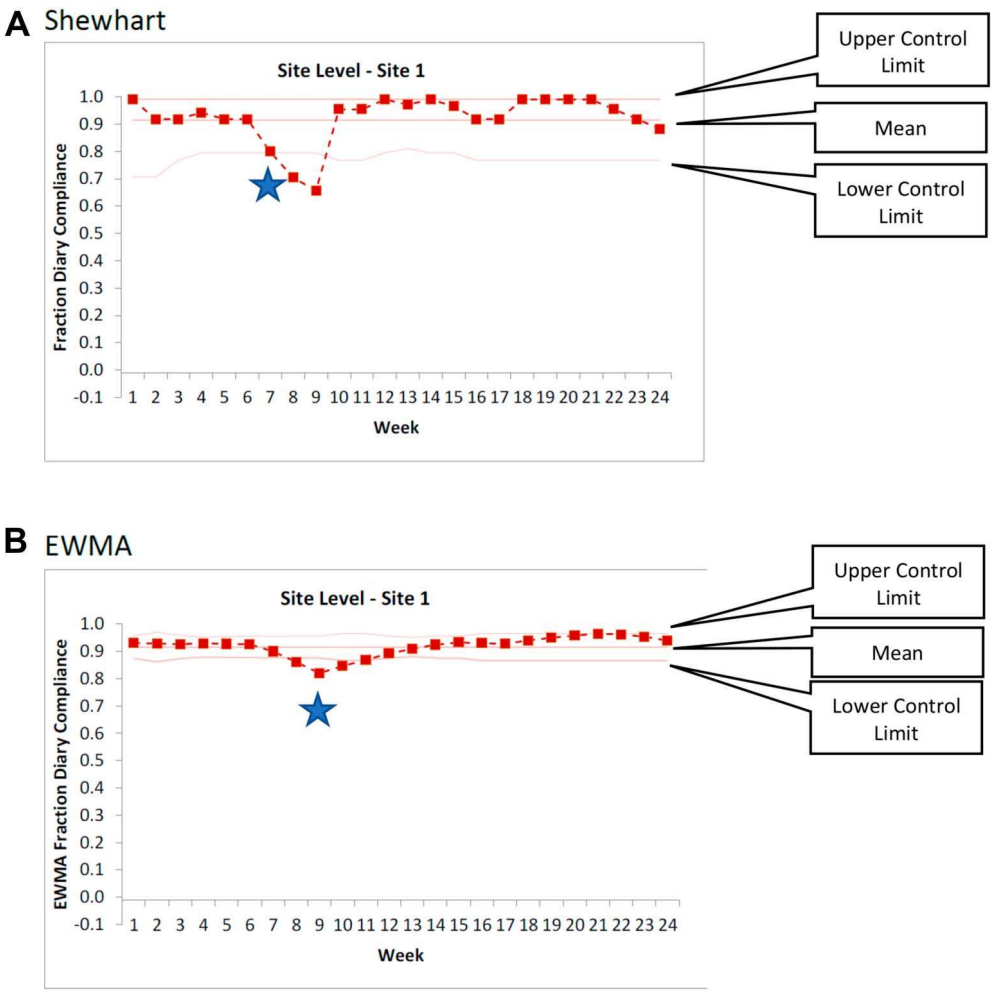

Figure 5. Statistical process control charts from a clinical trial: (A) A Shewhart control chart of e-diary compliance from a site in a multicenter clinical trial of a treatment for knee osteoarthritis. The red squares indicate the actual value of e-diary compliance, with 1.0 indicating 100% compliance. Upper, mean, and lower control limits are indicated. Note that the values for the control limits change as data accumulate. The blue star illustrates where e-diary compliance crossed the lower control limit, indicating a loss of process control. (B) An Exponentially Weighted Moving Average (EWMA) chart of the same data: This approach smooths out random fluctuations. The blue star illustrates where e-diary compliance crossed the lower control limit; detection occurred 2 weeks later compared with the Shewhart chart.

common in clinical trials. A superficial analysis might conclude that *what* happened was that Mr Jones forgot to enter his pain score last night. *How* it happened was "he forgot." Such an analysis might lead to weak recommendations for remediating the problem, such as retraining Mr Jones about compliance or retraining the site to remind Mr. Jones about his diaries. Instead, questions about *why* Mr Jones forgot might lead to underlying root causes such as lack of an effective reminder system, confusing data entry formats, dysfunctional log-ins, excessive data entry burden, lack of alarms, or failure of study coordinators to call patients who have missed entries. Abstract root causes, those requiring indefinite investigations, generic causes such as "human error" or "poor patient selection," or root causes leading to general recommendations such as "improve protocol compliance" mean that a useful root cause has not been identified, and effective prevention of recurrences is unlikely.[204] A recent unpublished study found that root cause analyses only took place in 4% of cases of performance issues in 2 clinical trials, leaving plenty of room for improvement.[141]

12.5. Interventions

Once one or more root causes have been identified, timely intervention should follow. In the case of clinical trials, there are multiple constraints on interventions because the integrity of the protocol and general research principles must be protected, including preservation of blinding, avoiding the introduction of new sources of measurement error or bias, and protection of patient safety. It can be tricky to decide when an intervention is correcting measurement error vs introducing bias. For this reason, a multidisciplinary team can be useful to make decisions about intervention policies. A clear and documented rationale for an intervention policy may serve a sponsor well with respect to regulatory interactions.

The Centers for Medicare & Medicaid Services has provided a useful approach by classifying interventions, in the general healthcare context, as strong, intermediate, or weak.[38] Weak interventions include double checks, warnings and labels, adding a new procedure/memorandum/policy, retraining, and additional analysis. Intermediate actions include increasing staffing/decreasing workload, modifying software, reducing distractions, implementing job aids, and enhancing communication. Strong actions include changing physical surroundings, usability testing, introducing engineering controls into a system, simplifying processes and removing unnecessary steps, and standardizing equipment or processes. Needless to say, after an intervention is implemented, continued surveillance should be performed to evaluate whether the intervention was effective in remediating the performance issue.

13. Site and country selection for multicenter trials

Ideally, the process of selecting clinical research sites should engage sites that can recruit large numbers of patients, care for these patients well during the study, and generate high-quality data. A study from the Tufts Center for the Study of Drug Development of 151 phase 2 to 3 studies found that, of more than 16,000 sites, 11% failed to enroll a single patient, and half of sites did not achieve enrollment targets.[92] Moreover, over half of studies were completed late, and 1 of every 6 studies took twice as long as planned. In a 2018 update, the same group reported an increase over time in the time needed to identify, select, and initiate sites, which averaged nearly 8 months. These metrics focus only on enrollment; site-specific surveys of data quality are not available.

Globalization of trials, which essentially means shifting studies away from the United States and Europe and towards the developing world, has increased in recent years. Pressures leading to globalization include decreased protected time and financial incentives for academic investigators; higher liability and clinical pressures for clinicians; protocol complexity; increased regulatory requirements by the Food & Drug Administration/European Medicines Agency; availability of low-cost global sites; and better enrollment from sites in developing regions.[210] Globalization has come at a significant price, however, including increased regional heterogeneity in enrolled populations, larger trial sizes, higher trial costs, issues with data reliability and security, and challenges in characterizing global study sites. In an interesting counterpoint, a survey regarding site selection experiences in Europe found that administrative complexity rather than cost was the main obstacle to using European sites.[90] This suggests that reducing study start-up time and effort by simplifying complex administrative obstacles would decrease pressure towards globalization. There has been virtually no published research on regional variability in study outcomes of pain (or any other) treatments. One study in acute migraine found that placebo responses were higher and treatment differences were smaller in Japan compared with the United States or European Union,[208] reinforcing concerns about regional differences.

Other important questions about the impact of site selection on trial results have also been subject to little systematic research. Some studies support the intuitive notion that minimizing the number of sites leads to better results.[130,169] In some studies, sites with low recruitment rates have compromised assay sensitivity, presumably because of challenges in reliably executing study procedures with long gaps between patients.[98,221]

The essential challenge is that sponsors and investigators must transition from a trial-by-trial mentality to a long-term infrastructure and collaboration mentality. Beginning to think about site selection during study startup while under high pressure to meet timelines can be expected to replicate previous results for site selection, data quality, and patient enrollment. Incentives for achieving recruitment timelines without incentives for clinical data quality often achieve that goal. Establishing a network of qualified and certified investigators and creating a patient registry or performing a survey study to characterize patients in the orbit of potential sites has been shown to save money and accelerate timelines in other therapeutic areas.[100] Prestudy patient outreach activities to promote a spirit of volunteerism are needed to produce a material shift in the current site selection and recruitment crisis.[92] Precompetitive cross-company collaborations have shown promise in related arenas. Ensuring that contracts with Contract Research Organizations and sites are aligned with incentives for meeting not only enrollment but also data quality targets requires a shift from established precedent.

14. Summary: best practices for conducting confirmatory pain studies

The most important step in creating a mindset of quality clinical research is to abandon the antiquated concept that clinical trials are somehow a method for capturing data from clinical practice, which is unfortunately enshrined in the very language of clinical research (eg, clinical research standards are referred to as "Good Clinical Practice"), and shifting to a concept of the clinical trial as a

measurement system, consisting of an interconnected set of processes, each of which must be in calibration in order for the trial as a whole to generate an accurate and reliable estimate of the efficacy (and safety) of a given treatment. This task can be framed as a search for sources of measurement error in clinical trials and validation of methods to minimize them. The status quo of inaccurate, unreliable, and protracted clinical trials is unacceptable and unsustainable. Only through leadership and collaboration will the existing paradigm of human experimentation shift to a new paradigm of high measurement quality that patients, investigators, sponsors, and other stakeholders deserve.

Disclosures

N. Katz is an employee of WCG Analgesic Solutions, a clinical trials services company with many clients in the pharmaceutical and medical device industries.

Acknowledgements

N. Katz would like to extend a special thanks to Robert Dworkin, Robert Kerns, Michael McDermott, Dennis Turk, and Christin Veasley for their expertise in the editing process of this paper.

Article history:
Received 27 May 2020
Received in revised form 7 August 2020
Accepted 19 August 2020
Available online 18 December 2020

References

[1] ACRP White Paper on future trends. Monitor 1997:15–25.
[2] Afilalo M, Etropolski MS, Kuperwasser B, Kelly K, Okamoto A, Van Hove I, Steup A, Lange B, Rauschkolb C, Haeussler J. Efficacy and safety of Tapentadol extended release compared with oxycodone controlled release for the management of moderate to severe chronic pain related to osteoarthritis of the knee: a randomized, double-blind, placebo- and active-controlled phase III study. Clin Drug Investig 2010;30:489–505.
[3] Agarwal D, Udoji MA, Trescot A. Genetic testing for opioid pain management: a primer. Pain Ther 2017;6:93–105.
[4] Altman R, Hochberg M, Gibofsky A, Jaros M, Young C. Efficacy and safety of low-dose SoluMatrix meloxicam in the treatment of osteoarthritis pain: a 12-week, phase 3 study. Curr Med Res Opin 2015;31:2331–43.
[5] Anchisi D, Zanon M. A Bayesian perspective on sensory and cognitive integration in pain perception and placebo analgesia. PLoS One 2015; 10:e0117270.
[6] Andersen JR, Byrjalsen I, Bihlet A, Kalakou F, Hoeck HC, Hansen G, Hansen HB, Karsdal MA, Riis BJ. Impact of source data verification on data quality in clinical trials: an empirical post hoc analysis of three phase 3 randomized clinical trials. Br J Clin Pharmacol 2015;79:660–8.
[7] Anderson RB, Hollenberg NK, Williams GH. Physical Symptoms Distress Index: a sensitive tool to evaluate the impact of pharmacological agents on quality of life. Arch Intern Med 1999;159: 693–700.
[8] Backonja M, Williams L, Miao X, Katz N, Chen C. Safety and efficacy of neublastin in painful lumbosacral radiculopathy: a randomized, double-blinded, placebo-controlled phase 2 trial using Bayesian adaptive design (the SPRINT trial). PAIN 2017;158:1802–12.
[9] Bain EE, Shafner L, Walling DP, Othman AA, Chuang-Stein C, Hinkle J, Hanina A. Use of a novel artificial intelligence platform on mobile devices to assess dosing compliance in a phase 2 clinical trial in subjects with schizophrenia. JMIR Mhealth Uhealth 2017;5:e18.
[10] Bakobaki JM, Rauchenberger M, Joffe N, McCormack S, Stenning S, Meredith S. The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi-centre clinical trial. Clin Trials 2012;9:257–64.
[11] Baron R, Maier C, Attal N, Binder A, Bouhassira D, Cruccu G, Finnerup NB, Haanpaa M, Hansson P, Hullemann P, Jensen TS, Freynhagen R, Kennedy JD, Magerl W, Mainka T, Reimer M, Rice AS, Segerdahl M, Serra J, Sindrup S, Sommer C, Tolle T, Vollert J, Treede RD. Peripheral neuropathic pain: a mechanism-related organizing principle based on sensory profiles. PAIN 2017;158:261–72.
[12] Baros AM, Latham PK, Moak DH, Voronin K, Anton RF. What role does measuring medication compliance play in evaluating the efficacy of naltrexone? Alcohol Clin Exp Res 2007;31:596–603.
[13] Beecher HK. The powerful placebo. J Am Med Assoc 1955;159: 1602–6.
[14] Belknap R, Weis S, Brookens A, Au-Yeung KY, Moon G, DiCarlo L, Reves R. Feasibility of an ingestible sensor-based system for monitoring adherence to tuberculosis therapy. PLoS One 2013;8:e53373.
[15] Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
[16] Bellamy N, Klestov A, Muirden K, Kuhnert P, Do KA, O'Gorman L, Martin N. Perceptual variation in categorizing individuals according to American College of Rheumatology classification criteria for hand, knee, and hip osteoarthritis (OA): observations based on an Australian Twin Registry study of *OA*. J Rheumatol 1999;26:2654–8.
[17] Bellamy N, Sothern RB, Campbell J. Rhythmic variations in pain perception in osteoarthritis of the knee. J Rheumatol 1990;17:364–72.
[18] Benedetti F. Placebo effects: Understanding the mechanisms in health and disease. Oxford: Oxford University Press, 2014.
[19] Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. Qual Saf Health Care 2003; 12:458–64.
[20] Bjordal JM, Klovning A, Ljunggren AE, Slordal L. Short-term efficacy of pharmacotherapeutic interventions in osteoarthritic knee pain: a meta-analysis of randomised placebo-controlled trials. Eur J Pain 2007;11: 125–38.
[21] Blaschke TF, Osterberg L, Vrijens B, Urquhart J. Adherence to medications: insights arising from studies on the unreliable link between prescribed and actual drug dosing histories. Annu Rev Pharmacol Toxicol 2012;52:275–301.
[22] Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. J Public Health Dent 2011;71(suppl 1):S52–63.
[23] Bothwell LE, Avorn J, Khan NF, Kesselheim AS. Adaptive design clinical trials: a review of the literature and ClinicalTrials.gov. BMJ Open 2018;8: e018320.
[24] Boutron I, Guittet L, Estellat C, Moher D, Hro´ bjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. PLoS Med 2007;4:e61.
[25] Bower P, Brueton V, Gamble C, Treweek S, Smith CT, Young B, Williamson P. Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities. Trials 2014;15:399.
[26] Breckenridge A, Aronson JK, Blaschke TF, Hartman D, Peck CC, Vrijens B. Poor medication adherence in clinical trials: consequences and solutions. Nat Rev Drug Discov 2017;16:149–50.
[27] Brosteanu O, Schwarz G, Houben P, Paulus U, Strenge-Hesse A, Zettelmeyer U, Schneider A, Hasenclever D. Risk-adapted monitoring is not inferior to extensive on-site monitoring: results of the ADAMON cluster-randomised study. Clin Trials 2017;14:584–96.
[28] Buchbinder R, Batterham R, Ciciriello S, Newman S, Horgan B, Ueffing E, Rader T, Tugwell PS, Osborne RH. Health literacy: what is it and why is it important to measure? J Rheumatol 2011;38:1791–7.
[29] Butler S, Katz N, Budman SH, Fernandez K. Development of an opioid side effects scale. J Pain 2005;6:S85.
[30] Butler SF, Black RA, Techner L, Fernandez KC, Brooks D, Wood M, Katz N. Development and validation of the post-operative recovery index for measuring quality of recovery after surgery. J Anesth Clin Res 2013; 3:1–8.
[31] Buzkova P. Measurement error and outcomes defined by exceeding a threshold: biased findings in comparative effectiveness trials. Pharm Stat 2012;11:429–41.
[32] Byas-Smith MG, Max MB, Muir J, Kingman A. Transdermal clonidine compared to placebo in painful diabetic neuropathy using a two-stage "enriched enrollment" design. PAIN 1995;60:267–74.
[33] Campbell CM, Gilron I, Doshi T, Raja S. Designing and conducting proof-of-concept chronic pain analgesic clinical trials. Pain Rep 2019;4: e697.
[34] Carroll JS, Rudolph JW, Hatakenaka S. Lessons learned from non-medical industries: root cause analysis as culture change at a chemical plant. Qual Saf Health Care 2002;11:266–9.

[35] Charles R, Hood B, Derosier JM, Gosbee JW, Li Y, Caird MS, Biermann JS, Hake ME. How to perform a root cause analysis for workup and future prevention of medical errors: a review. Patient Saf Surg 2016;10:20.
[36] Chen PA, Cheong JH, Jolly E, Elhence H, Wager TD, Chang LJ. Socially transmitted placebo effects. Nat Hum Behav 2019;3:1295–305.
[37] Chiron C, Dulac O, Gram L. Vigabatrin withdrawal randomized study in children. Epilepsy Res 1996;25:209–15.
[38] CMS. Guidance for performing root cause analysis (RCA) with performance improvement Projects (PIPs). Center for Medicare and Medicaid Services. Available at: https://www.cms.gov/medicare/provider-enrollment-and-certification/qapi/downloads/guidanceforrca.pdf.
[39] Cohen SP, Wallace M, Rauck RL, Stacey BR. Unique aspects of clinical trials of invasive therapies for chronic pain. Pain Rep 2018;4:e687.
[40] Conaghan PG, Hunter DJ, Cohen SB, Kraus VB, Berenbaum F, Lieberman JR, Jones DG, Spitzer AI, Jevsevar DS, Katz NP, Burgess DJ, Lufkin J, Johnson JR, Bodick N, Investigators FXP. Effects of a single intra-articular injection of a microsphere formulation of triamcinolone acetonide on knee osteoarthritis pain: a double-blinded, randomized, placebo-controlled, multinational study. J Bone Joint Surg Am 2018;100:666–77.
[41] Cruciani RA, Katz N, Portenoy RK. Dose equivalence of immediate-release hydromorphone and once-daily osmotic-controlled extended-release hydromorphone: a randomized, double-blind trial incorporating a measure of assay sensitivity. J Pain 2012;13:379–89.
[42] Czobor P, Skolnick P. The secrets of a successful clinical trial: compliance, compliance, and compliance. Mol Interv 2011;11:107–10.
[43] Earls E. Clinical trial delays: America's patient recruitment dilemma. Clinical Trials Arena, July 18, 2012. Available at: https://www.clinicaltrialsarena.com/analysis/featureclinical-trial-patient-recruitment/
[44] Davis JR, Nolan VP, Woodcock J, Estabrook RW; Institute of Medicine (US) Roundtable on Research and Development of Drugs, Biologics, and Medical Devices. Assuring data quality and validity in clinical trials for regulatory decision making: workshop report. Washington, DC: National Academies Press; 1999.
[45] Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Soc Sci Med 2018;210:2–21.
[46] Dellemijn PL, Vanneste JA. Randomised double-blind active-placebo-controlled crossover trial of intravenous fentanyl in neuropathic pain. Lancet 1997;349:753–8.
[47] Demant DT, Lund K, Finnerup NB, Vollert J, Maier C, Segerdahl MS, Jensen TS, Sindrup SH. Pain relief with lidocaine 5% patch in localized peripheral neuropathic pain in relation to pain phenotype: a randomised, double-blind, and placebo-controlled, phenotype panel study. PAIN 2015;156:2234–44.
[48] Demant DT, Lund K, Vollert J, Maier C, Segerdahl M, Finnerup NB, Jensen TS, Sindrup SH. The effect of oxcarbazepine in peripheral neuropathic pain depends on pain phenotype: a randomised, double-blind, placebo-controlled phenotype-stratified study. PAIN 2014;155:2263–73.
[49] Deming WE. Out of the crisis. Cambridge, MA: The MIT Press; 2000.
[50] Demitrack MA, Faries D, Herrera JM, DeBrota D, Potter WZ. The problem of measurement error in multisite clinical trials. Psychopharmacol Bull 1998;34:19–24.
[51] Demonceau J, Ruppar T, Kristanto P, Hughes DA, Fargher E, Kardas P, De Geest S, Dobbels F, Lewek P, Urquhart J, Vrijens B; team ABCp. Identification and assessment of adherence-enhancing interventions in studies assessing medication adherence through electronically compiled drug dosing histories: a systematic literature review and meta-analysis. Drugs 2013;73:545–62.
[52] Desmet L, Venet D, Doffagne E, Timmermans C, Burzykowski T, Legrand C, Buyse M. Linear mixed-effects models for central statistical monitoring of multicenter clinical trials. Stat Med 2014;33:5265–79.
[53] Detke MJ, Lu Y, Goldstein DJ, Hayes JR, Demitrack MA. Duloxetine, 60 mg once daily, for major depressive disorder: a randomized double-blind placebo-controlled trial. J Clin Psychiatry 2002;63:308–15.
[54] Detke MJ, Lu Y, Goldstein DJ, McNamara RK, Demitrack MA. Duloxetine 60 mg once daily dosing versus placebo in the acute treatment of major depression. J Psychiatr Res 2002;36:383–90.
[55] Devine EG, Peebles KR, Martini V. Strategies to exclude subjects who conceal and fabricate information when enrolling in clinical trials. Contemp Clin Trials Commun 2017;5:67–71.
[56] Devine EG, Waters ME, Putnam M, Surprise C, O'Malley K, Richambault C, Fishman RL, Knapp CM, Patterson EH, Sarid-Segal O, Streeter C, Colanari L, Ciraulo DA. Concealment and fabrication by experienced research subjects. Clin Trials 2013;10:935–48.
[57] Drennan KB. Patient recruitment: the costly and growing bottleneck in drug development. Drug Discov Today 2002;7:167–70.
[58] Durant RW, Wenzel JA, Scarinci IC, Paterniti DA, Fouad MN, Hurd TC, Martin MY. Perspectives on barriers and facilitators to minority recruitment for clinical trials among cancer center leaders, investigators, research staff, and referring clinicians: enhancing minority participation in clinical trials (EMPaCT). Cancer 2014;120(suppl 7):1097–105.
[59] Dworkin RH, Katz J, Gitlin MJ. Placebo response in clinical trials of depression and its implications for research on chronic neuropathic pain. Neurology 2005;65(12 suppl 4):S7–19.
[60] Dworkin RH, Peirce-Sandner S, Turk DC, McDermott MP, Gibofsky A, Simon LS, Farrar JT, Katz NP. Outcome measures in placebo-controlled trials of osteoarthritis: responsiveness to treatment effects in the REPORT database. Osteoarthritis Cartilage 2011;19:483–92.
[61] 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.
[62] Dworkin RH, Turk DC, Katz NP, Rowbotham MC, Peirce-Sandner S, Cerny I, Clingman CS, Eloff BC, Farrar JT, Kamp C, McDermott MP, Rappaport BA, Sanhai WR. Evidence-based clinical trial design for chronic pain pharmacotherapy: a blueprint for ACTION. PAIN 2011;152(3 suppl):S107–115.
[63] 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.
[64] 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.
[65] 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 ACTION meta-anlysis. Neurology 2013;81:67–75.
[66] Dworkin RH, Turk DC, Peirce-Sandner S, He H, McDermott MP, Hochberg MC, Jordan JM, Katz NP, Lin AH, Neogi T, Rappaport BA, Simon LS, Strand V. Meta-analysis of assay sensitivity and study features in clinical trials of pharmacologic treatments for osteoarthritis pain. Arthritis Rheumatol 2014;66:3327–36.
[67] Edwards JE, McQuay HJ, Moore RA, Collins SL. Reporting of adverse effects in clinical trials should be improved: lessons from acute postoperative pain. J Pain Symptom Manage 1999;18:427–37.
[68] Edwards P, Shakur H, Barnetson L, Prieto D, Evans S, Roberts I. Central and statistical data monitoring in the clinical randomisation of an antifibrinolytic in significant haemorrhage (CRASH-2) trial. Clin Trials 2014;11:336–43.
[69] Edwards RR, Dworkin RH, Turk DC, Angst MS, Dionne R, Freeman R, Hansson P, Haroutounian S, Arendt-Nielsen L, Attal N, Baron R, Brell J, Bujanover S, Burke LB, Carr D, Chappell AS, Cowan P, Etropolski M, Fillingim RB, Gewandter JS, Katz NP, Kopecky EA, Markman JD, Nomikos G, Porter L, Rappaport BA, Rice AS, Scavone JM, Scholz J, Simon LS, Smith SM, Tobias J, Tockarshewsky T, Veasley C, Versavel M, Wasan AD, Wen W, Yarnitsky D. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. PAIN 2016;157:1851–71.
[70] Eisenberger U, Wuthrich RP, Bock A, Ambuhl P, Steiger J, Intondi A, Kuranoff S, Maier T, Green D, DiCarlo L, Feutren G, De Geest S. Medication adherence assessment: high accuracy of the new Ingestible Sensor System in kidney transplants. Transplantation 2013;96:245–50.
[71] European Medicines Agency/Committee for Medicial Products for Human Use. Guideline on the clinical development of medicinal products intended for the treatment of pain: European Medicines

Agency/Committee for Medicial Products for Human Use. London, United Kingdom, 2016.
[72] Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical?. JAMA 2000;283:2701–11.
[73] Erpelding N, Evans K, Lanier R, Elder H, Katz N. Placebo response reduction and accurate pain reporting training reduces placebo responses in a clinical trial on chronic low back pain: results from a comparison to the literature. Clin J Pain 2020. doi: 10.1097/AJP.0000000000000873 [Epub ahead of print].
[74] European Medicines Agency. Note for guidance on clinical investigation of medicinal products for treatment of nociceptive pain. London, United Kingdom: EMA, 2003. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/note-guidance-clinical-investigation-medicinal-products-treatment-nociceptive-pain_en.pdf
[75] European Medicines Agency. Guideline on clinical and medicinal products intended for the treatment of neuropathic pain. London, United Kingdom: EMA,2007. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-medicinal-products-intended-treatment-neuropathic-pain_en.pdf
[76] Farrar JT, Troxel AB, Haynes K, Gilron I, Kerns RD, Katz NP, Rappaport BA, Rowbotham MC, Tierney AM, Turk DC, Dworkin RH. Effect of variability in the 7-day baseline pain diary on the assay sensitivity of neuropathic pain randomized clinical trials: an ACTTION study. PAIN 2014;155:1622–31.
[77] Food and Drug Administration. Patient-reported outcome measures: use in medical product development to support labeling claims. Rockville, MD: United States Food and Drug Administration, 2009.
[78] Food and Drug Administration. Non-inferiority clinical trials. Rockville, MD: United States Food and Drug Administration/Center for Drug Evaluation and Research, 2010.
[79] Food and Drug Administration. Enrichment strategies for clinical trials to support approval of human drugs and biological products. Rockville, MD: United States Food and Drug Administration/Center for Drug Evaluation and Research, 2012.
[80] Food and Drug Administration. Design considerations for pivotal clinical investigations for medical devices: guidance for industry, clinical investigators, institutional review boards and. Food And Drug Adm Staff 2013:31–2.
[81] Food and Drug Administration. Oversight of clinical investigations—a risk-based approach to monitoring. Rockville, MD: United States Food and Drug Administration/Center for Drug Evaluation and Research, 2013.
[82] Food and Drug Administration. Assessment of abuse potential of drugs. Rockville, MD: United States Food and Drug Administration/Center for Drug Evaluation and Research, 2017.
[83] Food and Drug Administration/National Institutes of Health. BEST (Biomarkers, EndpointS, and other Tools) resource. Silver Spring, MD: FDA-NIH Biomarker Working Group, 2016.
[84] Frank, JD, Frank, J. Persuasion and healing: a comparative study of psychotherapy. Baltimore, MD: Johns Hopkins University Press, 1991.
[85] Freeman R, Baron R, Bouhassira D, Cabrera J, Emir B. Sensory profiles of patients with neuropathic pain based on the neuropathic pain symptoms and signs. PAIN 2014;155:367–76.
[86] Freynhagen R, Strojek K, Griesing T, Whalen E, Balkenohl M. Efficacy of pregabalin in neuropathic pain evaluated in a 12-week, randomised, double-blind, multicentre, placebo-controlled trial of flexible- and fixed-dose regimens. PAIN 2005;115:254–63.
[87] Frisaldi E, Shaibani A, Benedetti F. Why we should assess patients' expectations in clinical trials. Pain Ther 2017;6:107–10.
[88] Fuller J. The confounding question of confounding causes in randomized trials. Br J Phil Sci 2019;70:901–26.
[89] Gan TJ, Kranke P, Minkowitz HS, Bergese SD, Motsch J, Eberhart L, Leiman DG, Melson TI, Chassard D, Kovac AL, Candiotti KA, Fox G, Diemunsch P. Intravenous amisulpride for the prevention of postoperative nausea and vomiting: two concurrent, randomized, double-blind, placebo-controlled trials. Anesthesiology 2017;126:268–75.
[90] Gehring M, Taylor RS, Mellody M, Casteels B, Piazzi A, Gensini G, Ambrosio G. Factors influencing clinical trial site selection in Europe: the survey of attitudes towards trial sites in Europe (the SAT-EU study). BMJ Open 2013;3:e002957.
[91] George SL, Buyse M. Data fraud in clinical trials. Clin Investig (Lond) 2015;5:161–73.
[92] Getz KA. Enrollment performance: weighing the "facts". Applied Clinical Trials 2012;21:5. Available at: https://www.appliedclinicaltrialsonline.com/view/enrollment-performance-weighing-facts
[93] Getz KA Campo RA. New benchmarks characterizing growth in protocol design complexity. Ther Innov Regul Sci 2018;52:22–8.
[94] Getz KA, Stergiopoulos S, Marlborough M, Whitehill J, Curran M, Kaitin KI. Quantifying the magnitude and cost of collecting extraneous protocol data. Am J Ther 2015;22:117–24.
[95] Getz KA, Zuckerman R, Cropp A, Hindle A, Krauss R, Kaitin K. Measuring the incidence, causes, and repercussions of protocol amendments. Drug Inf 2011;45:265–75.
[96] Getz K, Sethuraman V, Rine J, Peña Y, Ramanathan S, Stergiopoulos S. Assessing patient participation burden based on protocol design characteristics. Ther Innov Regul Sci 2019. doi: 10.1177/2168479019867284 [Epub ahead of print].
[97] Gewandter JS, Dworkin RH, Turk DC, McDermott MP, Baron R, Gastonguay MR, Gilron I, Katz NP, Mehta C, Raja SN, Senn S, Taylor C, Cowan P, Desjardins P, Dimitrova R, Dionne R, Farrar JT, Hewitt DJ, Iyengar S, Jay GW, Kalso E, Kerns RD, Leff R, Leong M, Petersen KL, Ravina BM, Rauschkolb C, Rice AS, Rowbotham MC, Sampaio C, Sindrup SH, Stauffer JW, Steigerwald I, Stewart J, Tobias J, Treede RD, Wallace M, White RE. Research designs for proof-of-concept chronic pain clinical trials: IMMPACT recommendations. PAIN 2014;155:1683–95.
[98] Gewandter JS, Dworkin RH, Turk DC, Devine E, Hewitt D, Jensen MP, Katz NP, Kirkwood A, Malamut R, Markman JD, Virijens B, Allen R, Burke L, Campbell JN, Carr D, Chen C, Cheung R, Conaghan PG, Cowan P, Doyle MK, Edwards R, Evans S, Farrar JT, Freeman R, Gilron I, Jacobs D, Judge D, Kopecky EA, Kerns RD, McDermott MP, Mullay S, Niebler G, Patel KV, Rauck R, Rice A, Rowbotham M, Sessler N, Simon LS, Singla N, Skljarevski V, Tockarshewsky T, Upmalis D, Vanhove TF, Wasan AD, Witter J. Ensuring data quality in clinical trials of pain treatments: Considerations for study execution and conduct. J Pain 2019;13:S1526–5900(19)30880-6.
[99] Gewandter JS, McDermott MP, McKeown A, Hoang K, Iwan K, Kralovic S, Rothstein D, Gilron I, Katz NP, Raja SN, Senn S, Smith SM, Turk DC, Dworkin RH. Reporting of cross-over clinical trials of analgesic treatments for chronic pain: analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks systematic review and recommendations. PAIN 2016;157:2544–51.
[100] Gheorghiade M, Vaduganathan M, Greene SJ, Mentz RJ, Adams KF Jr, Anker SD, Arnold M, Baschiera F, Cleland JG, Cotter G, Fonarow GC, Giordano C, Metra M, Misselwitz F, Muhlhofer E, Nodari S, Frank Peacock W, Pieske BM, Sabbah HN, Sato N, Shah MR, Stockbridge NL, Teerlink JR, van Veldhuisen DJ, Zalewski A, Zannad F, Butler J. Site selection in global clinical trials in patients hospitalized for heart failure: perceived problems and potential solutions. Heart Fail Rev 2014;19:135–52.
[101] Gilron I, Bailey JM, Tu D, Holden RR, Weaver DF, Houlden RL. Morphine, gabapentin, or their combination for neuropathic pain. N Engl J Med 2005;352:1324–34.
[102] Gilron I, Bailey JM, Tu D, Holden RR, Jackson AC, Houlden RL. Nortriptyline and gabapentin, alone and in combination for neuropathic pain: a double-blind, randomised controlled crossover trial. Lancet 2009;374:1252–61.
[103] Gimbel JS, Kivitz AJ, Bramson C, Nemeth MA, Keller DS, Brown MT, West CR, Verburg KM. Long-term safety and effectiveness of tanezumab as treatment for chronic low back pain. PAIN 2014;155:1793–801.
[104] Goldstein DJ, Lu Y, Detke MJ, Wiltse C, Mallinckrodt C, Demitrack MA. Duloxetine in the treatment of depression: a double-blind placebo-controlled comparison with paroxetine. J Clin Psychopharmacol 2004;24:389–99.
[105] Gracely RH, Dubner R, Wolskee PJ, Deeter WR. Placebo and naloxone can alter post-surgical pain by separate mechanisms. Nature 1983;306:264–5.
[106] Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health 2001;22:189–212.
[107] Gurrell R, Dua P, Feng G, Sudworth M, Whitlock M, Reynolds DS, Butt RP. A randomised, placebo-controlled clinical trial with the α2/3/5 subunit selective GABAA positive allosteric modulator PF-06372865 in patients with chronic low back pain. PAIN 2018;159:1742–51.
[108] Hale ME, Dvergsten C, Gimbel J. Efficacy and safety of oxymorphone extended release in chronic low back pain: results of a randomized, double-blind, placebo- and active-controlled phase III study. J Pain 2005;6:21–8.
[109] Harel Z, Silver SA, McQuillan RF, Weizman AV, Thomas A, Chertow GM, Nesrallah G, Chan CT, Bell CM. How to diagnose solutions to a quality of care problem. Clin J Am Soc Nephrol 2016;11:901–7.
[110] Hargreaves, B. Clinical trials and their patients: The rising costs and how to stem the loss. Pharmafile 2016. Available at: http://www.pharmafile.com/news/511225/clinical-trials-and-their-patients-rising-costs-and-how-stem-loss.

[111] Harris RE, Williams DA, McLean SA, Sen A, Hufford M, Gendreau RM, Gracely RH, Clauw DJ. Characterization and consequences of pain variability in individuals with fibromyalgia. Arthritis Rheum 2005;52:3670–4.
[112] Harter JG, Peck CC. Chronobiology. Suggestions for integrating it into drug development. Ann N Y Acad Sci 1991;618:563–71.
[113] Hassett AL, Pierce J, Goesling J, Fritsch L, Bakshi RR, Kohns DJ, Brummett CM. Initial validation of the electronic form of the Michigan Body Map. Reg Anesth Pain Med 2019. doi: 10.1136/rapm-2019-101084 [Epub ahead of print].
[114] Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol 2014;32:40–51.
[115] Haynes RB, Ackloo E, Sahota N, McDonald HP, Yao X. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2008;2:CD000011.
[116] Herman WH, Pop-Busui R, Braffett BH, Martin CL, Cleary PA, Albers JW, Feldman EL; DCCT/EDIC Research Group. Use of the Michigan Neuropathy Screening Instrument as a measure of distal symmetrical peripheral neuropathy in type 1 diabetes: results from the diabetes control and complications trial/epidemiology of diabetes interventions and complications. Diabet Med 2012;29:937–44.
[117] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Cochrane, 2020. Available at: www.training.cochrane.org/handbook.
[118] Howards PP. An overview of confounding. Part 1: the concept and how to address it. Acta Obstet Gynecol Scand 2018;97:394–9.
[119] Howards PP. An overview of confounding. Part 2: how to identify it and special situations. Acta Obstet Gynecol Scand 2018;97:400–6.
[120] Hughes-Morley A, Young B, Waheed W, Small N, Bower P. Factors affecting recruitment into depression trials: systematic review, meta-synthesis and conceptual framework. J Affect Disord 2015;172:274–90.
[121] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Integrated addendum to ICH E6(R1): Guideline for good clinical practice E6(R2). 2016.
[122] International Organization for Standardization. Accuracy (trueness and precision) of measurement methods and results—part 1: general principles and definitions (ISO Standard No. 5725-1:1994). 1994. Available at: https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en.
[123] 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.
[124] Jeffery RA, Navarro T, Wilczynski NL, Iserman EC, Keepanasseril A, Sivaramalingam B, Agoritsas T, Haynes RB. Adherence measurement and patient recruitment methods are poor in intervention trials to improve patient adherence. J Clin Epidemiol 2014;67:1076–82.
[125] Jeong S, Sohn M, Kim JH, Ko M, Seo HW, Song YK, Choi B, Han N, Na HS, Lee JG, Kim IW, Oh JM, Lee E. Current globalization of drug interventional clinical trials: characteristics and associated factors, 2011–2013. Trials 2017;18:288.
[126] Joint Committee for Guides in Metrology.. International vocabulary of metrology—basic and general concepts and associated terms. VIM 3rd ed. 2012.
[127] Kam-Hansen S, Jakubowski M, Kelley JM, Kirsch I, Hoaglin DC, Kaptchuk TJ, Burstein R. Altered placebo and drug labeling changes the outcome of episodic migraine attacks. Sci Transl Med 2014;6:218ra5.
[128] Kaptchuk TJ. Open-label placebo: reflections on a research agenda. Perspect Biol Med 2018;61:311–34.
[129] Katz J, Finnerup NB, Dworkin RH. Clinical trial outcome in neuropathic pain: relationship to study characteristics. Neurology 2008;70:263–72.
[130] Katz N. Methodological issues in clinical trials of opioids for chronic pain. Neurology 2005;65(12 suppl 4):S32–49.
[131] Katz N. Enriched enrollment randomized withdrawal trial designs of analgesics: focus on methodology. Clin J Pain 2009;25:797–807.
[132] Katz N, Borenstein DG, Birbara C, Bramson C, Nemeth MA, Smith MD, Brown MT. Efficacy and safety of tanezumab in the treatment of chronic low back pain. PAIN 2011;152:2248–58.
[133] Katz N, Fanciullo GJ. Role of urine toxicology testing in the management of chronic opioid therapy. Clin J Pain 2002;18(4 suppl):S76–82.
[134] Katz N, Kopecky EA, O'Connor M, Brown RH, Fleming AB. A phase 3, multicenter, randomized, double-blind, placebo-controlled, safety, tolerability, and efficacy study of Xtampza ER in patients with moderate-to-severe chronic low back pain. PAIN 2015;156:2458–67.
[135] Katz NP. The measurement of symptoms and side effects in clinical trials of chronic pain. Contemp Clin Trials 2012;33:903–11.
[136] Katz NP, Mou J, Paillard FC, Turnbull B, Trudeau J, Stoker M. Predictors of response in patients with postherpetic neuralgia and HIV-associated neuropathy treated with the 8% capsaicin patch (Qutenza). Clin J Pain 2015;31:859–66.
[137] Katz NP, Mou J, Trudeau J, Xiang J, Vorsanger G, Orman C, Kim M. Development and preliminary validation of an integrated efficacy-tolerability composite measure for the evaluation of analgesics. PAIN 2015;156:1357–65.
[138] Katz NP, Sherburne S, Beach M, Rose RJ, Vielguth J, Bradley J, Fanciullo GJ. Behavioral monitoring and urine toxicology testing in patients receiving long-term opioid therapy. Anesth Analg 2003;97:1097–102, table of contents.
[139] Katz NP, Paillard F, Benneyan J, Kacena K, Frye S, Lucius D, Treister R. Development and validation of a clinical trial data surveillance method to improve assay sensitivity of pain clinical trials. Poster presented at: The American Pain Society Annual Meeting; May 2015; Palm Springs, CA.
[140] Katz NP, Dworkin R, North R, Thomson S, Eldabe S, Hayek S, Kopell B, Markman J, Rezai A, Taylor R, Turk D, Buchser E, Fields H, Fiore G, Furguson M, Gewandter J, Hilker C, Jain R, Leitner A, Loeser J, McNicol E, Nurmikko T, Pena C, Scott P, Shipley J, Trescott A, van Dongen R, Venkatesan L. Research design considerations for randomized controlled trials of spinal cord stimulation for pain: IMMPACT/ION/neuromodulation foundation recommendations. 2020 (submitted).
[141] Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. J Clin Epidemiol 1999;52:19–26.
[142] Khan A, Khan SR, Walens G, Kolts R, Giller EL. Frequency of positive studies among fixed and flexible dose antidepressant clinical trials: an analysis of the food and drug administration summary basis of approval reports. Neuropsychopharmacology 2003;28:552–7.
[143] Khan A, Kolts RL, Thase ME, Krishnan KR, Brown W. Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. Am J Psychiatry 2004;161:2045–9.
[144] Khan A, Redding N, Brown WA. The persistence of the placebo response in antidepressant clinical trials. J Psychiatr Res 2008;42:791–6.
[145] Kirkwood AA, Cox T, Hackshaw A. Application of methods for central statistical monitoring in clinical trials. Clin Trials 2013;10:783–806.
[146] Kleppinger CF, Ball LK. Building quality in clinical trials with use of a quality systems approach. Clin Infect Dis 2010;51(suppl 1):S111–116.
[147] Kobak KA, Kane JM, Thase ME, Nierenberg AA. Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? J Clin Psychopharmacol 2007;27:1–5.
[148] Kobak KA, Leuchter A, DeBrota D, Engelhardt N, Williams JB, Cook IA, Leon AC, Alpert J. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. J Clin Psychopharmacol 2010;30:193–7.
[149] Kolahi J, Bang H, Park J. Towards a proposal for assessment of blinding success in clinical trials: up-to-date review. Commun Dent Oral Epidemiol 2009;37:477–84.
[150] Kripalani S, Yao X, Haynes RB. Interventions to enhance medication adherence in chronic medical conditions: a systematic review. Arch Intern Med 2007;167:540–50.
[151] Krogsboll LT, Hrobjartsson A, Gotzsche PC. Spontaneous improvement in randomised clinical trials: meta-analysis of three-armed trials comparing no treatment, placebo and active intervention. BMC Med Res Methodol 2009;9:1.
[152] Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. Stroke 2017;48:1416–19.
[153] Lachin JM. The role of measurement reliability in clinical trials. Clin Trials 2004;1:553–66.
[154] Landin R, DeBrota DJ, DeVries TA, Potter WZ, Demitrack MA. The impact of restrictive entry criterion during the placebo lead-in period. Biometrics 2000;56:271–8.
[155] Lane NE, Schnitzer TJ, Birbara CA, Mokhtarani M, Shelton DL, Smith MD, Brown MT. Tanezumab for the treatment of pain from osteoarthritis of the knee. N Engl J Med 2010;363:1521–31.
[156] Lee DJ, Avulova S, Conwill R, Barocas DA. Patient engagement in the design and execution of urologic oncology research. Urol Oncol 2017;35:552–8.
[157] Liu H, Golin CE, Miller LG, Hays RD, Beck CK, Sanandaji S, Christian J, Maldonado T, Duran D, Kaplan AH, Wenger NS. A comparison study of multiple measures of adherence to HIV protease inhibitors. Ann Intern Med 2001;134:968–77.

[158] Mallinckrodt CH, Tamura RN, Tanaka Y. Recent developments in improving signal detection and reducing placebo response in psychiatric clinical trials. J Psychiatr Res 2011;45:1202–7.
[159] Manchin J. Manchin urges HHS secretary to reverse approval of Zohydro. 2014. Available at: https://www.manchin.senate.gov/newsroom/press-releases/manchin-urges-hhs-secretary-to-reverse-approval-of-zohydro
[160] Manson JE, Shufelt CL, Robins JM. The potential for postrandomization confounding in randomized clinical trials. JAMA 2016;315:2273–4.
[161] Markman J, Resnick M, Greenberg S, Katz N, Yang R, Scavone J, Whalen E, Gregorian G, Parsons B, Knapp L. Efficacy of pregabalin in post-traumatic peripheral neuropathic pain: a randomized, double-blind, placebo-controlled phase 3 trial. J Neurol 2018;265:2815–24.
[162] Mauri L, D'Agonstino RB. Challenges in the design and interpretation of non-inferiority trials. N Engl J Med 2017;377:1357–67.
[163] Max MB. Divergent traditions in analgesic clinical trials. Clin Pharmacol Ther 1994;56:237–41.
[164] Max MB, Portenoy RK, Laska EM. The design of analgesic clinical trials. Advances in cancer research and therapy. New York: Raven Press, 1991.
[165] Mayorga AJ, Flores CM, Trudeau JJ, Moyer JA, Shalayda K, Dale M, Frustaci ME, Katz N, Manitpisitkul P, Treister R, Ratcliffe S, Romano G. A randomized study to evaluate the analgesic efficacy of a single dose of the TRPV1 antagonist mavatrep in patients with osteoarthritis. Scand J Pain 2017;17:134–43.
[166] McDonnell A, Collins S, Ali Z, Iavarone L, Surujbally R, Kirby S, Butt RP. Efficacy of the Nav1.7 blocker PF-05089771 in a randomised, placebo-controlled, double-blind clinical study in subjects with painful diabetic peripheral neuropathy. PAIN 2018;159:1465–76.
[167] Meghani SH, Thompson AM, Chittams J, Bruner DW, Riegel B. Adherence to analgesics for cancer pain: a comparative study of African Americans and Whites using an electronic monitoring device. J Pain 2015;16:825–35.
[168] Meldrum ML. A brief history of the randomized controlled trial. From oranges and lemons to the gold standard. Hematol Oncol Clin North Am 2000;14:745–60, vii.
[169] Meske DS, Vaugn B, Kopecky E, Katz N. Number of clinical trial study sites impacts observed treatment effect size: an analysis of randomized controlled trials of opioids for chronic pain. J Pain Res 2019;12:3161–5.
[170] Moore A, Derry S, Eccleston C, Kalso E. Expect analgesic failure; pursue analgesic success. BMJ 2013;346:f2690.
[171] 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.
[172] Morlock R, Braunstein GD. Pharmacoeconomics of genotyping-based treatment decisions in patients with chronic pain. Pain Rep 2017;2:e615.
[173] Mou J, Paillard F, Turnbull B, Trudeau J, Stoker M, Katz NP. Efficacy of Qutenza(R) (capsaicin) 8% patch for neuropathic pain: a meta-analysis of the Qutenza clinical trials database. PAIN 2013;154:1632–9.
[174] Muller MJ, Szegedi A. Effects of interrater reliability of psychopathologic assessment on power and sample size calculations in clinical trials. J Clin Psychopharmacol 2002;22:318–25.
[175] Nathan RA. How important is patient recruitment in performing clinical trials? J Asthma 1999;36:213–16.
[176] Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, Agoritsas T, Mistry N, Iorio A, Jack S, Sivaramalingam B, Iserman E, Mustafa RA, Jedraszewski D, Cotoi C, Haynes RB. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014;11:CD000011.
[177] Nissen SE, Yeomans ND, Solomon DH, Luscher TF, Libby P, Husni ME, Graham DY, Borer JS, Wisniewski LM, Wolski KE, Wang Q, Menon V, Ruschitzka F, Gaffney M, Beckerman B, Berger MF, Bao W, Lincoff AM, Investigators PT. Cardiovascular safety of celecoxib, naproxen, or ibuprofen for arthritis. N Engl J Med 2016;375:2519–29.
[178] North RB, Kidd DH, Farrokhi F, Piantadosi SA. Spinal cord stimulation versus repeated lumbosacral spine surgery for chronic pain: a randomized, controlled trial. Neurosurgery 2005;56:98–106.
[179] O'Connor AB, Turk DC, Dworkin RH, Katz NP, Colucci R, Haythornthwaite JA, Klein M, O'Brien C, Posner K, Rappaport BA, Reisfield G, Adams EH, Balster RL, Bigelow GE, Burke LB, Comer SD, Cone E, Cowan P, Denisco RA, Farrar JT, Foltin RW, Haddox JD, Hertz S, Jay GW, Junor R, Kopecky EA, Leiderman DB, McDermott MP, Palmer PP, Raja SN, Rauschkolb C, Rowbotham MC, Sampaio C, Setnik B, Smith SM, Sokolowska M, Stauffer JW, Walsh SL, Zacny JP. Abuse liability measures for use in analgesic clinical trials in patients with pain: IMMPACT recommendations. PAIN 2013;154:2324–34.
[180] Odrcich M, Bailey JM, Cahill CM, Gilron I. Chronobiological characteristics of painful diabetic neuropathy and postherpetic neuralgia: diurnal pain variation and effects of analgesic therapy. PAIN 2006;120:207–12.
[181] O'Kelly M. Using statistical techniques to detect fraud: a test case. Pharm Stat 2004;3:237–46.
[182] Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schunemann H, Briel M, Nordmann AJ, Pregno S, Oxman AD. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2011;4:MR000012.
[183] Osgood E, Trudeau JJ, Eaton TA, Jensen MP, Gammaitoni A, Simon LS, Katz N. Development of a bedside pain assessment kit for the classification of patients with osteoarthritis. Rheumatol Int 2015;35:1005–13.
[184] Papakostas GI, Fava M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. Eur Neuropsychopharmacol 2009;19:34–40.
[185] Park JJ, Thorlund K, Mills EJ. Critical concepts in adaptive clinical trials. Clin Epidemiol 2018;10:343–51.
[186] Patel KV, Amtmann D, Jensen MP, Smith SM, Veasley C, Turk DC. Clinical outcome assessment in clinical trials of chronic pain treatments. PAIN Rep 2020 :e784.
[187] Pearl J Why there is no statistical test for confounding, why many think there is, and why they are almost right. In: UoCL Angeles, editor. Technical Report R-256 1998.
[188] Peck CC, Temple R, Collins JM. Understanding consequences of concurrent therapies. JAMA 1993;269:1550–2.
[189] Perkins DO, Wyatt RJ, Bartko JJ. Penny-wise and pound-foolish: the impact of measurement error on sample size requirements in clinical trials. Biol Psychiatry 2000;47:762–6.
[190] Petrone D, Kamin M, Olson W. Slowing the titration rate of tramadol HCl reduces the incidence of discontinuation due to nausea and/or vomiting: a double-blind randomized trial. J Clin Pharm Ther 1999;24:115–23.
[191] Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA 2012;308:2594–604.
[192] Pocock SJ, Abdalla M. The hope and the hazards of using compliance data in randomized controlled trials. Stat Med 1998;17:303–17.
[193] Podsadecki TJ, Vrijens BC, Tousset EP, Rode RA, Hanna GJ. "White coat compliance" limits the reliability of therapeutic drug monitoring in HIV-1-infected patients. HIV Clin Trials 2008;9:238–46.
[194] Portenoy RK, Ganae-Motan ED, Allende S, Yanagihara R, Shaiova L, Weinstein S, McQuade R, Wright S, Fallon MT. Nabiximols for opioid-treated cancer patients with poorly-controlled chronic pain: a randomized, placebo-controlled, graded-dose trial. J Pain 2012;13:438–49.
[195] Powers JH III, Patrick DL, Walton MK, Marquis P, Cano S, Hobart J, Isaac M, Vamvakas S, Slagle A, Molsen E, Burke LB. Clinician-reported outcome assessments of treatment benefit: report of the ISPOR clinical outcome assessment emerging good practices task force. Value Health 2017;20:2–14.
[196] Pullar T, Kumar S, Tindall H, Feely M. Time to stop counting the tablets? Clin Pharmacol Ther 1989;46:163–8.
[197] Quessy SN. Two-stage enriched enrolment pain trials: a brief review of designs and opportunities for broader application. PAIN 2010;148:8–13.
[198] Quessy SN, Rowbotham MC. Placebo response in neuropathic pain trials. PAIN 2008;138:479–83.
[199] Raja SN, Haythornthwaite JA, Pappagallo M, Clark MR, Travison TG, Sabeen S, Royall RM, Max MB. Opioids versus antidepressants in postherpetic neuralgia: a randomized, placebo-controlled trial. Neurology 2002;59:1015–21.
[200] Reimer M, Forstenpointner J, Hartmann A, JC Otto, J Vollert, J Gierthmühlen, T Klein, P Hüllemann, R Baron. Sensory bedside testing: a simple stratification approach for sensory phenotyping. PAIN Rep 2020;5:e820.
[201] Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995;123:A12–13.
[202] Robinson KA, Dinglas VD, Sukrithan V, Yalamanchilli R, Mendez-Tellez PA, Dennison-Himmelfarb C, Needham DM. Updated systematic review identifies substantial number of retention strategies: using more strategies retains more study participants. J Clin Epidemiol 2015;68:1481–7.

[203] Rolke R, Baron R, Maier C, Tolle TR, Treede RD, Beyer A, Binder A, Birbaumer N, Birklein F, Botefur IC, Braune S, Flor H, Huge V, Klug R, Landwehrmeyer GB, Magerl W, Maihofner C, Rolko C, Schaub C, Scherens A, Sprenger T, Valet M, Wasserka B. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. PAIN 2006;123:231–43.
[204] Rooney JJ, VandenHeuvel LN. Root cause analysis for beginners. Qual Proc 2004;37:45–53.
[205] Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioral therapies research: getting started and moving on from stage I. Clin Psychol Sci Prac 2001;8:133–42.
[206] Rowbotham MC, Twilling L, Davies PS, Reisner L, Taylor K, Mohr D. Oral opioid therapy for chronic peripheral and central neuropathic pain. N Engl J Med 2003;348:1223–32.
[207] Ruehlman LS, Karoly P, Enders C. A randomized controlled evaluation of an online chronic pain self management program. PAIN 2012;153: 319–30.
[208] Sakai F, Diener HC, Ryan R, Poole P. Eletriptan for the acute treatment of migraine: results of bridging a Japanese study to Western clinical trials. Curr Med Res Opin 2004;20:269–77.
[209] Sang CN, Booher S, Gilron I, Parada S, Max MB. Dextromethorphan and memantine in painful diabetic neuropathy and postherpetic neuralgia: efficacy and dose-response trials. Anesthesiology 2002;96: 1053–61.
[210] Sarwar CMS, Vaduganathan M, Butler J. Impact of site selection and study conduct on outcomes in global clinical trials. Curr Heart Fail Rep 2017;14:203–9.
[211] Savovic J, Jones H, Altman D, Harris R, Juni P, Pildal J, Als-Nielsen B, Balk E, Gluud C, Gluud L, Ioannidis J, Schulz K, Beynon R, Welton N, Wood L, Moher D, Deeks J, Sterne J. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. Health Technol Assess 2012;16:1–82.
[212] Schnitzer TJ, Easton R, Pang S, Levinson DJ, Pixton G, Viktrup L, Davignon I, Brown MT, West CR, Verburg KM. Effect of tanezumab on Joint pain, physical function, and patient global assessment of osteoarthritis among patients with osteoarthritis of the hip or knee: a randomized clinical trial. JAMA 2019;322:37–48.
[213] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408–12.
[214] Senn SS. Cross-over trials in clinical research. Hoboken: Wiley, 2002.
[215] Shaya FT, Gbarayor CM, Huiwen Keri Yang, Agyeman-Duah M, Saunders E. A perspective on African American participation in clinical trials. Contemp Clin Trials 2007;28:213–7. Review.
[216] Sheiner LB. Learning versus confirming in clinical drug development. Clin Pharmacol Ther 1997;61:275–91.
[217] Shewhart WA. The economic control of quality of manufactured product. New York: D Van Nostrand, 1931.
[218] Shiovitz TM, Wilcox CS, Gevorgyan L, Shawkat A. CNS sites cooperate to detect duplicate subjects with a clinical trial subject registry. Innov Clin Neurosci 2013;10:17–21.
[219] Simpson DM, Robinson-Papp J, Van J, Stoker M, Jacobs H, Snijder RJ, Schregardus DS, Long SK, Lambourg B, Katz N. Capsaicin 8% patch in painful diabetic peripheral neuropathy: a randomized, double-blind, placebo-controlled study. J Pain 2017;18:42–53.
[220] Simsek I, Yazici Y. Incomplete reporting of recruitment information in clinical trials of biologic agents for the treatment of rheumatoid arthritis: a review. Arthritis Care Res (Hoboken) 2012;64:1611–16.
[221] Singla NK, Chelly JE, Lionberger DR, Gimbel J, Sanin L, Sporn J, Yang R, Cheung R, Knapp L, Parsons B. Pregabalin for the treatment of postoperative pain: results from three controlled trials using different surgical models. J Pain Res 2015;8:9–20.
[222] Sinyor M, Levitt AJ, Cheung AH, Schaffer A, Kiss A, Dowlati Y, Lanctot KL. Does inclusion of a placebo arm influence response to active antidepressant treatment in randomized controlled trials? Results from pooled and meta-analyses. J Clin Psychiatry 2010;71:270–9.
[223] Sjoding MW, Cooke CR, Iwashyna TJ, Hofer TP. Acute respiratory distress syndrome measurement error. Potential effect on clinical study results. Ann Am Thorac Soc 2016;13:1123–8.
[224] Smith SM, Amtmann D, Askew RL, Gewandter JS, Hunsinger M, Jensen MP, McDermott MP, Patel KV, Williams M, Bacci ED, Burke LB, Chambers CT, Cooper SA, Cowan P, Desjardins P, Etropolski M, Farrar JT, Gilron I, Huang IZ, Katz M, Kerns RD, Kopecky EA, Rappaport BA, Resnick M, Strand V, Vanhove GF, Veasley C, Versavel M, Wasan AD, Turk DC, Dworkin RH. Pain intensity rating training: results from an exploratory study of the ACTTION PROTECCT system. PAIN 2016;157: 1056–64.
[225] Smith SM, Chang RD, Pereira A, Shah N, Gilron I, Katz NP, Lin AH, McDermott MP, Rappaport BA, Rowbotham MC, Sampaio C, Turk DC, Dworkin RH. Adherence to CONSORT harms-reporting recommendations in publications of recent analgesic clinical trials: an ACTTION systematic review. PAIN 2012;153:2415–21.
[226] Smith SM, Dart RC, Katz NP, Paillard F, Adams EH, Comer SD, Degroot A, Edwards RR, Haddox JD, Jaffe JH, Jones CM, Kleber HD, Kopecky EA, Markman JD, Montoya ID, O'Brien C, Roland CL, Stanton M, Strain EC, Vorsanger G, Wasan AD, Weiss RD, Turk DC, Dworkin RH; Analgesic A, Addiction Clinical Trials TIO, Networks public-private p. Classification and definition of misuse, abuse, and related events in clinical trials: ACTTION systematic review and recommendations. PAIN 2013;154:2287–96.
[227] Smith SM, Gewandter JS, Kitt RA, Markman JD, Vaughan JA, Cowan P, Kopecky EA, Malamut R, Sadosky A, Tive L, Turk DC, Dworkin RH. Participant preferences for pharmacologic chronic pain treatment trial characteristics: an ACTTION adaptive choice-based conjoint study. J Pain 2016;17:1198–206.
[228] Smith SM, Jones JK, Katz NP, Roland CL, Setnik B, Trudeau JJ, Wright S, Burke LB, Comer SD, Dart RC, Dionne R, Haddox JD, Jaffe JH, Kopecky EA, Martell BA, Montoya ID, Stanton M, Wasan AD, Turk DC, Dworkin RH. Measures that identify prescription medication misuse, abuse, and related events in clinical trials: ACTTION critique and recommended considerations. J Pain 2017;18:1287–94.
[229] Smith SM, Paillard F, McKeown A, Burke LB, Edwards RR, Katz NP, Papadopoulos EJ, Rappaport BA, Slagle A, Strain EC, Wasan AD, Turk DC, Dworkin RH. Instruments to identify prescription medication misuse, abuse, and related events in clinical trials: an ACTTION systematic review. J Pain 2015;16:389–411.
[230] Stacey BR, Barrett JA, Whalen E, Phillips KF, Rowbotham MC. Pregabalin for postherpetic neuralgia: placebo-controlled trial of fixed and flexible dosing regimens on allodynia and time to onset of pain relief. J Pain 2008;9:1006–17.
[231] Stamer UM, Stuber F. Genetic factors in pain and its treatment. Curr Opin Anaesthesiol 2007;20:478–84.
[232] Stamer UM, Zhang L, Stuber F. Personalized therapy in pain management: where do we stand? Pharmacogenomics 2010;11: 843–64.
[233] Straube S, Derry S, McQuay HJ, Moore RA. Enriched enrollment: definition and effects of enrichment and dose in trials of pregabalin and gabapentin in neuropathic pain. A systematic review. Br J Clin Pharmacol 2008;66:266–75.
[234] Suarez-Almazor ME, Looney C, Liu Y, Cox V, Pietz K, Marcus DM, Street RL Jr. A randomized controlled trial of acupuncture for osteoarthritis of the knee: effects of patient-provider communication. Arthritis Care Res (Hoboken) 2010;62:1229–36.
[235] Suzuki E, Mitsuhashi T, Tsuda T, Yamamoto E. A typology of four notions of confounding in epidemiology. J Epidemiol 2017;27:49–55.
[236] Taylor RN, McEntegart DJ, Stillman EC. Statistical techniques to detect fraud and other data irregularities in clinical questionnaire data. Drug Inf J 2002;36:115–25.
[237] Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. Ann Intern Med 2000;133:455–63.
[238] Tieu C, Breder CD. A critical evaluation of safety signal analysis using algorithmic standardised MedDRA queries. Drug Saf 2018;41:1375–85.
[239] Treister R, Eaton TA, Trudeau JJ, Elder H, Katz NP. Development and preliminary validation of the focused analgesia selection test to identify accurate pain reporters. J Pain Res 2017;10:319–26.
[240] Treister R, Honigman L, Lawal OD, Lanier RK, Katz NP. A deeper look at pain variability and its relationship with the placebo response: results from a randomized, double-blind, placebo-controlled clinical trial of naproxen in osteoarthritis of the knee. PAIN 2019;160:1522–28.
[241] Treister R, Lawal OD, Shecter JD, Khurana N, Bothmer J, Field M, Harte SE, Kruger GH, Katz NP. Accurate pain reporting training diminishes the placebo response: results from a randomised, double-blind, crossover trial. PLoS One 2018;13:e0197844.
[242] Treister R, Trudeau JJ, Van Inwegen R, Jones JK, Katz NP. Development and feasibility of the misuse, abuse, and diversion drug event reporting system (MADDERS(R)). Am J Addict 2016;25:641–51.
[243] Trijau S, Avouac J, Escalas C, Gossec L, Dougados M. Influence of flare design on symptomatic efficacy of non-steroidal anti-inflammatory drugs in osteoarthritis: a meta-analysis of randomized placebo-controlled trials. Osteoarthritis Cartilage 2010;18:1012–18.
[244] Trudeau J, Van Inwegen R, Eaton T, Bhat G, Paillard F, Ng D, Tan K, Katz NP. Assessment of pain and activity using an electronic pain diary and actigraphy device in a randomized, placebo-controlled crossover trial of celecoxib in osteoarthritis of the knee. Pain Pract 2015;15:247–55.

[245] Tso AR, Goadsby PJ. Anti-CGRP monoclonal antibodies: the next era of migraine prevention?. Curr Treat Options Neurol 2017;19:27.
[246] Tudur Smith C, Stocken DD, Dunn J, Cox T, Ghaneh P, Cunningham D, Neoptolemos JP. The value of source data verification in a cancer clinical trial. PLoS One 2012;7:e51623.
[247] Turk DC, Dworkin RH, Revicki D, Harding G, Burke LB, Cella D, Cleeland CS, Cowan P, Farrar JT, Hertz S, Max MB, Rappaport BA. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. PAIN 2008;137:276–85.
[248] 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.
[249] van Seventer R, Bach FW, Toth CC, Serpell M, Temple J, Murphy TK, Nimour M. Pregabalin in the treatment of post-traumatic peripheral neuropathic pain: a randomized double-blind trial. Eur J Neurol 2010; 17:1082–9.
[250] van Stralen KJ, Dekker FW, Zoccali C, Jager KJ. Confounding. Nephron Clin Pract 2010;116:c143–147.
[251] Vase L, Vollert J, Finnerup NB, Miao X, Atkinson G, Marshall S, Nemeth R, Lange B, Liss C, Price DD, Maier C, Jensen TS, Segerdahl M. Predictors of the placebo analgesia response in randomized controlled trials of chronic pain: a meta-analysis of the individual data from nine industrially sponsored trials. PAIN 2015;156:1795–802.
[252] Verbeurgt P, Mamiya T, Oesterheld J. How common are drug and gene interactions? Prevalence in a sample of 1143 patients with CYP2C9, CYP2C19 and CYP2D6 genotyping. Pharmacogenomics 2014;15: 655–65.
[253] Vetter TR, Mascha EJ. Bias, confounding, and interaction: lions and tigers, and bears, oh my!. Anesth Analg 2017;125:1042–8.
[254] Vetter TR, Morrice D. Statistical process control: No hits, No runs, No errors? Anesth Analg 2019;128:374–82.
[255] Vinik AI, Perrot S, Vinik EJ, Pazdera L, Jacobs H, Stoker M, Long SK, Snijder RJ, van der Stoep M, Ortega E, Katz N. Capsaicin 8% patch repeat treatment plus standard of care (SOC) versus SOC alone in painful diabetic peripheral neuropathy: a randomised, 52-week, open-label, safety study. BMC Neurol 2016;16:251.
[256] Vinik AI, Tuchman M, Safirstein B, Corder C, Kirby L, Wilks K, Quessy S, Blum D, Grainger J, White J, Silver M. Lamotrigine for treatment of pain associated with diabetic neuropathy: results of two randomized, double-blind, placebo-controlled studies. PAIN 2007;128:169–79.
[257] Vrijens B, Belmans A, Matthys K, de Klerk E, Lesaffre E. Effect of intervention through a pharmaceutical care program on patient adherence with prescribed once-daily atorvastatin. Pharmacoepidemiol Drug Saf 2006;15:115–21.
[258] Vrijens B, Urquhart J. Patient adherence to prescribed antimicrobial drug dosing regimens. J Antimicrob Chemother 2005;55:616–27.
[259] Vrijens B, Urquhart J. Methods for measuring, enhancing, and accounting for medication adherence in clinical trials. Clin Pharmacol Ther 2014;95:617–26.
[260] Walton MK, Powers JH III, Hobart J, Patrick D, Marquis P, Vamvakas S, Isaac M, Molsen E, Cano S, Burke LB; International Society for P, Outcomes Research Task Force for Clinical Outcomes A. Clinical outcome assessments: conceptual foundation-report of the ISPOR clinical outcomes assessment—emerging good practices for outcomes research task force. Value Health 2015;18:741–52.
[261] Wasan AD, Davar G, Jamison R. The association between negative affect and opioid analgesia in patients with discogenic low back pain. PAIN 2005;117:450–61.
[262] Wawrzyniak KM, Sabo A, McDonald A, Trudeau JJ, Poulose M, Brown M, Katz NP. Root cause analysis of prescription opioid overdoses. J Opioid Manag 2015;11:127–37.
[263] Wesson DR, Ling W. The clinical opiate withdrawal scale (COWS). J Psychoactive Drugs 2003;35:253–9.
[264] Wise RA, Bartlett SJ, Brown ED, Castro M, Cohen R, Holbrook JT, Irvin CG, Rand CS, Sockrider MM, Sugar EA, American Lung Association Asthma Clinical Research C. Randomized trial of the effect of drug presentation on asthma outcomes: the American Lung Association Asthma Clinical Research Centers. J Allergy Clin Immunol 2009;124: 436–44, 444e431-438.
[265] Woolf CJ, Max MB. Mechanism-based pain diagnosis: issues for analgesic drug development. Anesthesiology 2001;95:241–9.
[266] Rauck R, Makumi CW, Schwartz S, Graff O, Meno-Tetang G, Bell CF, Kavanagh ST, McClung CL. A randomized, controlled trial of gabapentin enacarbil in subjects with neuropathic pain associated with diabetic peripheral neuropathy. Pain Pract 2013;13:485–96.
[267] Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, Dougados M, Hochberg M, Hunter DJ, Kwoh K, Lohmander LS, Tugwell P. OARSI recommendations for the management of hip and knee osteoarthritis: part III: changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage 2010;18:476–99.
[268] Ziegler D, Pritchett YL, Wang F, Desaiah D, Robinson MJ, Hall JA, Chappell AS. Impact of disease characteristics on the efficacy of duloxetine in diabetic peripheral neuropathic pain. Diabetes Care 2007; 30:664–9.
[269] Zimbroff DL. Patient and rater education of expectations in clinical trials (PREECT). J Clin Psychopharmacol 2001;21:251–2.