REVIEW

FOCUS ON PAIN

nature medicine

Advances in clinical research methodology for pain clinical trials

John T Farrar

Pain is a ubiquitous phenomenon, but the experience of pain varies considerably from person to person. Advances in understanding of the growing number of pathophysiologic mechanisms that underlie the generation of pain and the influence of the brain on the experience of pain led to the investigation of numerous compounds for treating pain. Improved knowledge of the subjective nature of pain, the variations in the measurement of pain, the mind-body placebo effect and the impact of differences in the conduct of a clinical trial on the outcome have changed approaches to design and implement studies. Careful consideration of how these concepts affect the choice of study population, the randomization and blinding process, the measurement and collection of data, and the analysis and interpretation of results should improve the quality of clinical trials for potential pain therapies.

Pain is experienced in some form by all human beings on a daily basis and serves a protective role by warning us of potential injury and helping us prevent further damage once an injury has occurred. When pain is severe or becomes chronic, it often has a detrimental effect on a person's quality of life, and treatment is needed. Despite several centuries of human experience with potent opioid analgesics, pain often still remains undertreated. In part, this results from the ongoing controversy about the appropriate use of opioids. Although this class of drugs can substantially reduce pain, they do not completely relieve pain without major side effects and a considerable risk of opioid abuse or misuse in the pursuit of total pain relief. As a consequence, there is still a large unmet need for pain relief beyond what opioids and nonsteroidal anti-inflammatories (the other oral analgesics) can achieve.

Advancement in the treatment of pain was relatively slow until the serendipitous discovery of the efficacy of gabapentin for post-herpetic neuralgia and painful diabetic neuropathy in the late 1990s, resulting in US Food and Drug Administration approval for post-herpetic neuralgia in 2004. Over the past ten years, the number of potential pain therapies that have reached the stage of being tested in humans has grown dramatically, driving an interest in the development of more efficient and standardized study methods.

Building on the study designs originally developed for acute pain trials in the 1950s, methodologic enhancements were specifically developed for the clinical trials of gabapentin. As other products have been tested for efficacy, the methodology has been further refined. Here we focus on the innovations in clinical trial designs that have been necessary to integrate rapidly expanding scientific knowledge about the functioning of the nervous system with an understanding of its implications for the study of pain. Driven by the growing focus on evidence-based medicine, the usefulness of clinical trials requires the data to be collected, analyzed and interpreted using approaches that facilitate the application of research findings into clinical practice.

Definition of pain

Although pain is a ubiquitous experience, it has been hard to formally define. A generally accepted definition is an 'unpleasant sensory and emotional experience of actual or potential tissue damage or an experience expressed in such terms'[1]. The most common type of pain, often referred to as normal or nociceptive pain, is generated by stimulation of specialized nerve receptors, predominantly in the skin and the covering of other organs, by heat, cold, pressure, toxins or tissue damage. Unless severe, this type of pain is generally tolerable and self-limited. For severe pain, treatment with nonsteroidal or opioid analgesics are usually effective. Neuropathic pain, the other major type of pain, is generated directly from nervous system structures proximal to specialized pain receptors. A recent definition is 'pain arising as a direct consequence of a lesion or disease affecting the somatosensory system'[2]. A concern about this definition is that it does not clearly include pain resulting from anatomical or biochemical changes in the nervous system (neuroplasticity) that seem to underlie some forms of chronic neuropathic pain. In contrast, the term 'dysfunction of the nervous system'[1] is considered too broad. This lack of both a concise definition and specificity in making the diagnosis complicate the task of defining the appropriate population of individuals to be included in studies of specific pain syndromes.

The history of pain studies

As long as there has been an awareness of pain, there have been attempts by humans to learn how to control it. Some of the effects linked to pain, such as death related to the shock of un-anesthetized surgery, have been easily measured, and the efficacy of adequate treatments, such as chloroform and ether, have been so dramatic that carefully controlled clinical studies have not been necessary. The ability of opioids to ease severe pain has also been generally recognized since our earliest recorded

Associate Professor of Epidemiology and Anesthesia, Clinical Associate in Neurology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. Correspondence should be addressed to J.T.F. (jfarrar@upenn.edu).

Published online 14 October 2010; doi:10.1038/nm.2249

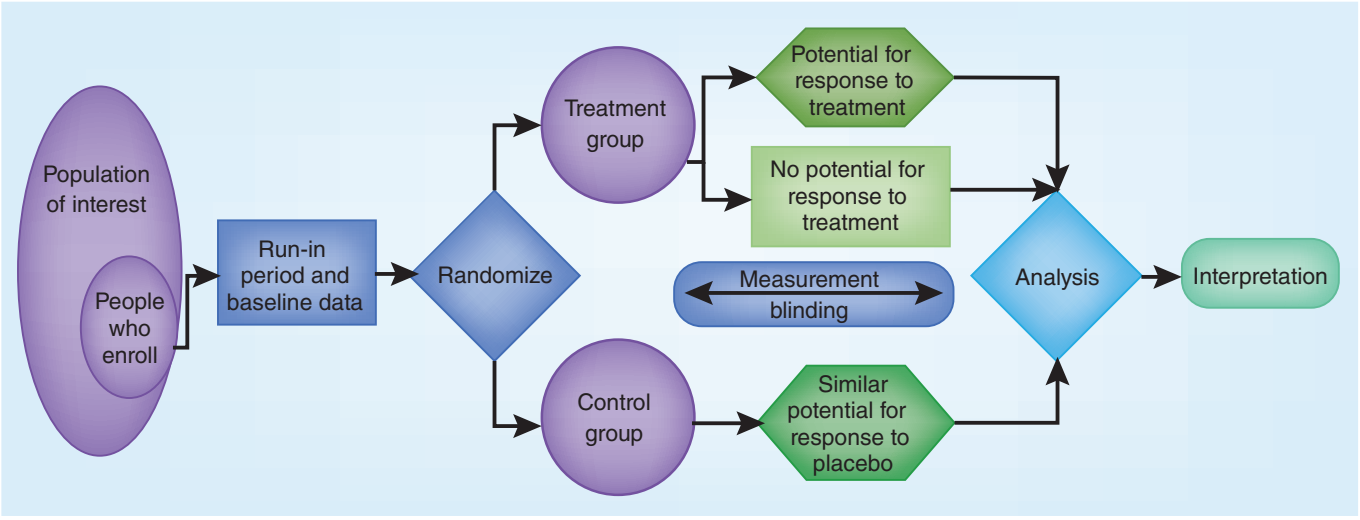

Figure 1 Schematic diagram of the crucial components of a prospective randomized, double blind, placebo controlled clinical trial.

history. Nevertheless, many other therapies have been promoted without providing much pain relief to the consumer. The need for a scientific approach to determining true cause and effect led to the development of clinical trials.

The importance of clinical trials in medicine was recognized after World War II. In the late 1940s, Austin Bradford Hill described the basics of a well-designed clinical trial[3]. The primary underlying principles codified during that period, including random allocation, blinding, outcome measures and *a priori* hypotheses, continue to be the basis for valid assessments of efficacy. Clinical trials of pain therapeutics started in the mid-1950s with the pioneering work of Louis Lasagna, Henry Beecher[4] and Ray Houde[5], who conducted simple, short-term, randomized two-group parallel trials focused on the major analgesic groups of opioids and nonsteroidal anti-inflammatories. In a seminal 1959 book[6], Henry Beecher identified difficulties associated with the measurement of pain, barriers in identifying populations of patients with specific pain syndromes, the magnitude and prevalence of the placebo effect, and the lack of a complete definition of a response to therapy. After more than 15,000 analgesic clinical trials, scientists are still struggling with many of the same problems. But expanded understanding of the physiologic processes that underlie the human experience of pain, the application of clinical trial principles to pain studies and the appropriate use of subjective measures have had important impacts on all aspects of the design and implementation of clinical trials, and the analysis and interpretation of their results.

Physiological processes and their implications for clinical trials

The expansion of knowledge about the physiology underlying pain, including the descending pain modulation systems and the processes underlying neuropathic pain, has substantially influenced the design of pain clinical trials.

The descending pain modulation system was formally proposed in 1965 by Melzack and Wall[7]. Since then, understanding of the mechanisms for these mind-body interactions has evolved, providing physiologic explanations for the effects of placebo treatment and of behavioral and psychological interventions that have affected how clinical trials are conducted.

In the placebo-instigated mind-body effect, individuals treated with the placebo experience a reduction in pain due to changes in the function of the central nervous system. These changes lead to production of endogenous opioids, descending nerve impulses that suppress the transmission of pain impulses, altered processing of the pain input by the brain that reduces the perception of pain, or other less well-described changes in the interplay between central nervous system components. The initiation of these processes depends on the belief of the study subject that the treatment received has a high probability of being effective. In addition, different patient populations, different diseases and subtle differences in study design are all thought to influence the strength of the placebo effect[8].

The capability of the human brain to exert substantial control over the perception and reporting of pain complicates the measurement of this subjective experience. Whereas researchers can try to observe the effect that pain has on a person's actions and ability to function, only the person feeling pain can know what he or she is experiencing. The principal method of measurement is to ask this person to report the experience on a scale, using words, diagrams, colors, or numbers. The subjectivity of the response has changed the measurement of pain, the rigor necessary in the conduct of pain clinical trials and the types of analyses used to assess the efficacy of the results.

The physiologic mechanisms of receptors[9], transmitters, nerve cell function and growth, and the role of glial cells have been shown to contribute to the production and modulation of pain. An important consequence has been the discovery of a number of mechanisms by which pain results from changes in functioning of the nervous system, generally referred to as neuropathic pain. The physiologic rationale for the persistence of pain, and the changes in the nervous system that enable it to spread and change in character, have only recently become known.

In recent population-based surveys, approximately 20% of chronic pain can be categorized as neuropathic, which is generally more severe than other types of pain[10,11]. The causes of neuropathic pain include cell death, abnormal cell growth and cell firing, atypical irritation of nerve fibers and a number of defined abnormal transmitter-receptor interactions. The discovery of these myriad pathophysiologic mechanisms has led to the progression of pain studies from a strictly symptom-based therapeutic approach to attempts at mechanism-based therapeutic approaches. Understanding that a large part of the chronic-pain problem is neuropathic pain has opened new avenues of research but also has demanded a reevaluation of approaches designed to identify patients with a specific underlying pain process. This is made more difficult

Table 1 The measurement of pain

Pain measure question format	Definition	Primary use	Potential problems	Refs.
Pain intensity	Measure of pain strength	Primary outcome measure of most pain studies (often modified by worst, least or average pain over a specific period)	Only one characteristic of pain; subjective perception of individual	46
Pain characteristics	Separate questions for describing multiple qualities of pain	Differentiates components and different types of pain	People's ability to differentiate pain types may be limited	47,48
Pain interference or bother	Measure of the effect of pain on life activities	Supports pain intensity; measures the importance of the pain to the individual	Does not rank importance of the activity for the individual	49,50
Pain relief	Single question about the reduction in pain intensity compared with a pretreatment time point	Measures individual's perception of efficacy of treatment over time	Does not always correlate with change in pain intensity; meaning may not be clear if used alone	51–53
Global change	Single question about the perception of change compared with a specific earlier time point	Measures change in pain or overall status over time	Reason for change may not be clear if no other questions are included	36,54,55
Pain type	Multi-question instrument measuring pain characteristics	Used to differentiate neuropathic from nociceptive pain	Lack of accepted description of neuropathic pain	27–30

Scalar assessment tool	Definition	Advantages	Potential problems	Refs.
VRS	Level of pain is indicated with descriptive words; usually four to ten words	Commonly used measure in clinical assessment of pain and acute pain research models	People interpret words differently, especially across cultures	56–58
Verbal transition scales	Change is indicated with words; offers choices of better or worse using a balanced scale	Direct measure of change; good psychometric properties	Perception of change may be influenced by other symptoms or characteristics of the individual's life	59–61
NRS	Numbers (0–10) with anchor descriptors are used to rate pain level or other symptoms	Good psychometric properties in measurement of change; ease of use	No intrinsic meaning to numbers	36,62
Percentage scale	Numbers expressed as 0–100%	Used to indicate percentage change over time	No intrinsic value to specific percentages	63,64
Graphical scales	Graphic presentation of response gradation (e.g., faces or colors)	Measure of pain in nonverbal adults or children	May not translate well across cultures	65–67
VAS (special instance of graphical scale)	Fixed-length line (often 10 cm) with delimiting descriptors; a mark is placed to indicate the level of pain	Commonly used; good psychometric properties	Requires visual presentation; no intrinsic meaning	68,69
Mechanical devices	Recording of response with mechanical or electronic devices	Can provide reminder, leading to a more consistent response	Requires a machine; not always appropriate for older population	70,71

Wide variation in the experience of chronic and neuropathic pain has led to the development of a broad range of pain-measurement instruments. In general, pain measures can be classified into four groups: measures of pain intensity, pain characteristics, pain interference, and pain relief or global change. In addition, there are multi-question measures to differentiate between types of pain. The primary outcome measure in most pain clinical trials is the change in pain over time and consists of two components, the pain question format and the scale used for measurement. Although multiple measures of intensity averaged over time may be more exact, subjects can accurately average their symptoms over a week[72].

because syndromes such as back pain are often made up of a mixture of both nociceptive and neuropathic pain.

Changes in the nervous system in response to the experience of pain are known as neuroplasticity. They can be transient if the inciting stimulus is reversible, or permanent if the inciting stimulus is irreversible or the adaptation of the nervous system function remains after the stimulus is gone. The previous dogma that the nervous system does not substantially change in adult life has been proven wrong by a large body of evidence, including recent work revealing the association of growth of stem cells with new memories in the hippocampus[12] and growth of nerve endings in the dorsal horn of the spinal cord in response to nerve injury[13]. The conditions necessary to alter the function of nerve fibers, as well as the constraints on timing after injury, remain unknown. Therefore, the duration of symptoms and the underlying pathophysiologic changes are crucial criteria in selecting a population of individuals for a pain clinical trial. Treatments to preempt pain or to prevent chronic changes must be tested in individuals who have the potential to respond to such a therapy (that is, those in whom damage is not yet irreversible). Preemptive analgesia has been demonstrated in animals[14] but so far not in humans[15].

The identification of the specific population of patients to study for possible response to a therapy that is thought to act on a specific pain mechanism is made more difficult because the perception and reporting of pain often does not adequately differentiate the pain etiology. Despite the advances in the tools available to examine brain and nervous system function, researchers are not yet able to differentiate the various abnormal processes in the individual. In a population with a mixed set of underlying pain etiology, it will be hard to detect any benefit from a treatment that effectively modulates only one or a few of these mechanisms.

Applying the principles of clinical trials to pain studies

The primary components of a clinical trial (**Fig. 1**) include selection of the study population, random allocation into two or more groups, blinding of the treatment, measurement of the outcome and analysis (**Box 1**).

Selection of the study population. Well-designed clinical trials must grapple with issues related to participant selection and the conflict between testing a treatment in the population that is most likely to respond and testing it in the larger population that is likely to be treated in the future. As a proof of concept, it is reasonable to focus on the most restricted and responsive group. Testing the treatment in a larger and more diverse group, however, may be prerequisite before a medication can be approved. The lower likelihood of response in a more diverse group means a larger sample size will be needed to test the same hypothesis, which is certainly possible but will be more expensive both in cost and in time.

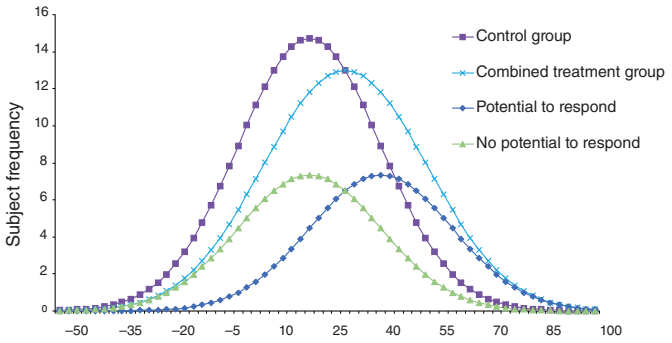

Figure 2 Normal density curves of simulated data for a single normal distribution control group and a combined (bimodal) treatment group. The combined treatment group is created by adding the values of the two normally distributed treatment subgroups each with 50% of the subjects. One subgroup includes subjects with the potential to respond to the treatment, resulting in a mean value that is larger than the control group. The other subgroup includes subjects without the potential to respond, which mimics the control group in mean change and distribution. The values on the horizontal axes represent the change in pain intensity (in percentage) from baseline to study endpoint. The vertical axis is the subject frequency, which indicates the number of individuals who achieved each level of change. The cumulative distribution function curves shown in **Figure 3** were constructed from the normal distribution curves shown here.

The group of people willing to volunteer for a clinical trial will always be only a subset of the total population of interest (**Fig. 1**). In general, the subject pool will be larger for studies of diseases for which there are few effective treatments. Once there is an effective therapy, clinical trials tend to enroll a more treatment-resistant group, as people who get relief from another treatment are less likely to want to try a new therapy. The difference between groups may be harder to detect, requiring a larger sample size. Recruiting and enrolling people who have not undergone any treatment is also a possibility but can be logistically difficult.

It is also crucial to properly identify a population of people with the specific disorder the study aims to treat. For pain studies, this means finding people with pain of the appropriate type, onset, duration of symptoms and etiology. A controversial issue currently is identifying individuals with primarily neuropathic pain and determining the specific underlying etiology of that pain. One consequence of this difficulty in diagnosis has been the focus of pain therapeutics over the last ten years on post-herpetic neuralgia and diabetic neuropathy pain, as these result from diseases known to consistently cause neuropathic pain syndromes and are relatively easy to diagnose. In contrast, common chronic back pain, for example, is known to have many causes, only some of which are neuropathic. As a result, the primary cause of pain is much more difficult to diagnose and is often intermixed with psychological and legal issues making it considerably harder to study.

Despite increasing interest in designing a rational approach to mechanism-based therapeutic interventions[16], the selection of study populations for a clinical trial is still based primarily on their symptoms that can result from multiple underlying etiologies. This will probably produce a group with several different potential mechanisms for their pain. If the therapy to be tested is specific for one or a small number of mechanisms that are not present in all individuals, then the study population will be comprised of individuals with the potential to respond to treatment and others without it (**Fig. 1**). This will result in a bimodal distribution in the study treatment group (**Fig. 2**), which may make nonparametric analyses more appropriate than the standard parametric Student's *t*-test, as described below.

Run-in period. Once subjects have been recruited, there is usually a baseline and/or run-in period before randomization. Purposes for this period include determining the stability of the baseline pain, testing each person's compliance with the clinical trial process and data collection requirements, collecting baseline data and sometimes tapering people off their current pain-related medications.

Much more controversial is the use of a placebo run-in period to exclude placebo responders[17]. In studies of depression treatments, a lower rate of response in the placebo-treated group usually correlates with a higher probability of achieving a statistically significant result[18]. Initial evaluations of pain studies have mostly corroborated these findings[8]. Nevertheless, when depression studies were designed to drop the placebo responders after a run-in period, no difference in statistical significance was found[18]. Although the removal of true placebo responders may seem to make sense, excluding people who experience less pain during a single short period of exposure to a placebo will also remove people who have normal random variation in their pain or who improve because they had worse pain at baseline. The result could be that the subjects remaining in a study after a run-in placebo exclusion are the individuals more difficult to treat, resulting in a negative trial outcome. At present, there is not enough information to determine whether such exclusion would have a beneficial or detrimental effect in the long run.

Randomization. A key scientific advantage of a clinical trial is the random allocation of appropriate and adequately sized populations into treatment and control groups for an *a priori* hypothesis to be tested. The balancing of known and unknown confounders with randomization is crucial to increase the probability that the differences observed between the groups at the end of a study are attributable to the effect of the experimental treatment. Stratified randomization can be used to assure a similar blinded assignment to treatment and control groups when participants are recruited at multiple study sites or treated by different practitioners[19].

Choice of the experimental and control treatment. The choice of treatment and the degree of blinding are crucial considerations. The therapeutic intervention under study is defined by the difference in treatment that is applied to the study versus the control group. For example, in drug trials, the active and placebo pills are made to be identical so that the only difference between the two groups is the physiologic activity of the medication. In a study of acupuncture, if the control group has acupuncture but with the needles placed in the wrong locations, the study question is the location of the needles.

Choice of the control group. The use of a control group is a requirement of all comparative studies and is necessary to differentiate the effect of the treatment from changes caused by four other nonrandom factors: the natural history of the process causing the pain, regression to the mean (for example, in pain studies patients tend to enroll when they are not doing well, and some will get better regardless of treatment), the mind-body effect from the additional clinical care provided during the study, and the mind-body effect due to subjects' expectation that they will improve with the treatment

Two types of control groups are most commonly used in clinical trials. A standard-of-care group, also called a no-treatment group, controls for natural-history phenomena and regression to the mean. A placebo-treatment group with blinding can also control for both of the potential mind-body effects. Notably, the response of a group treated with a blinded placebo controls for all four factors mentioned above. The difference in response between an unblinded no-treatment group and a group receiving a blinded placebo is, therefore, the change attributed to the two nonspecific mind-body effects, which is sometimes of interest in a clinical trial.

Box 1 Glossary of terms

Clinical trial terms	
Active control group or active comparator	A control group in a clinical trial given a treatment that is know to work and should produce improvement, to ensure that the study design can detect a positive response.
Blinding	Procedure to prevent the study participants, care providers, data collectors and statistical analysts from being aware of treatment assignments.
Outcome measures	Measures used to gauge the extent to which treatment affects the subjects' outcomes and to answer the scientific or clinical question of the study.
Placebo-controlled study	A clinical trial in which the control group is treated with a placebo designed to no effect, as a comparison group for the study treatment.
Placebo response	The response measured in a control group treated with a placebo. This is caused by a combination of regression to the mean, the natural history of the disease, the mind-body effect of belief in the treatment, and the mind-body effect of additional attention subjects receive when included in a treatment trial.
Placebo run-in periods	A period of time before randomization when all subjects are administered a placebo to assess procedure compliance and response to the placebo treatment.
Psychometric properties	Properties that link a measure to the underlying process it is used to evaluate. These are principally reliability, validity and responsiveness.
Stratified randomization	A procedure in which groups (called strata) are randomized separately to ensure treatment balance within that group.
Statistics terms	
Cumulative distribution curve	A graphical representation of the number of people who reach a value at or above each level of response in a clinical trial.
Normal distribution curve	A graphical representation of a distribution of any continuous variable that clusters around a midpoint in such a way that it can be completely described by a mean and standard deviation.
Ordinal regression	A statistical technique used to predict the behavior of ordered categorical dependent variables with a set of independent variables.
Population distribution	The distribution of scores or values across a population of interest.
Proportional odds statistics	A regression model for ordinal dependent variables.
Regression to the mean	A phenomenon in which subjects initially measured at extreme values will tend to be less extreme in a later measurement.
Statistical probability	The likelihood that an event or observation will occur over a specified period of time.
Statistical significance	The probability that the results observed in a clinical trial could have occurred by chance. The standard test evaluates whether the observed result could have occurred by chance at least 1 out of 20 times (producing $P \geq 0.05$).
Wilcoxon rank sum	A nonparametric alternative to the two-sample *t*-test, based solely on the order in which the observations from two groups occur.

There are, however, issues that may arise with the use of a placebo group. One is the ethics of using placebo groups for pain studies are frequently questioned, given that potentially effective therapies are available[20]. This problem might be overcome by keeping the period of placebo exposure to a minimum, by providing access to a rescue medication if the pain cannot be tolerated or by designing studies where the test treatment is an add-on to subjects' current chronic pain therapy. A second issue in the choice of a placebo-treated comparison group occurs in studies of pain-related procedures or instrumentation, where blinding may be difficult or unethical. As a result, surgical procedures are frequently not blinded for the patient or the investigator. The lack of blinding of the study groups must be carefully considered in the interpretation of any results from such trials[21].

In equivalency or comparative-effectiveness clinical trials, one treatment is compared to a second treatment that is known to have efficacy against the type of pain to be studied, also called an active comparator control. A major problem with equivalency studies is that a poorly executed study or one that is too small may not show a statistically significant difference even if a difference is present. In a placebo-controlled clinical trial, an active comparator control group can be added to provide a measurement standard to test whether the clinical trial has been designed and conducted correctly. For example, if the analysis of a clinical trial does not demonstrate that the active comparator control is better than the placebo group, then the trial procedure probably was flawed. In that case, a lack of efficacy of the new study treatment may be due to problems with the study design rather than the treatment. This is sometimes referred to as a 'failed trial'.

Just as there can be multiple control groups, there can be multiple treatment groups receiving different doses of the pain therapy. Demonstration of the lowest effective dose is important in minimizing the unnecessary exposure to higher doses. Showing increased efficacy with an increased dose (dose-effect) adds confidence to the findings of the trial.

Study implementation and blinding. One well-known component of the mind-body placebo response is the expectation of benefit[22], or the degree to which the trial subjects expect to improve with the active treatment. To prevent differences in expectation based on a variability in study implementation, clinical trials should have standardized study procedures, including scripting of the enrollment process and as many other processes as is practicable. Creation and distribution of a manual of procedures should be a routine practice. Training of all associated investigators and their study managers and regular review of procedures must be conducted. Study-site–specific factors are known to influence the degree of placebo response[23].

Once the control and treatment groups have been established, a process for blinding of the treatments must be designed, especially for clinical trials of pain therapy, given the strength of the mind-body process. The purpose of blinding is to ensure that the study groups, both treatment and control, are treated in exactly the same way throughout the trial. Any component of the treatment procedure that is not blinded must be considered potentially responsible for any change in the measured

response. Conversely, only the blinded part of the treatment is being tested[24]. In comparing results across clinical trials, careful consideration must be given to the nature and effectiveness of the blinding procedures (**Fig. 1**). Problems with blinding, such as medication side effects or the taste of the pill, can significantly affect the results of a study.

Measuring the pain. The usefulness of the information obtained from a clinical trial depends on the primary and secondary outcome measures selected. In pain studies, it would be helpful to directly measure the amount of painful input and the activation of associated central nervous system processes. With the advent of improved technologies to explore brain function, researchers are closer to objectively and efficiently quantifying functional changes in specific regions of the brain and understanding how those changes relate to the experience of pain in any individual. Functional brain images and, to some degree, electrical images have produced remarkable results regarding the components of the brain involved in pain[25,26]. So far, however, none of these findings has provided a measureable outcome that quantifies people's perception of pain and that is superior to asking how much pain they are experiencing, also known as patient-reported measures.

The questions used to interrogate people about their pain vary widely depending on the component of the pain being studied and its consequences. The clarity and specificity of the questions used will affect the reliability of the answers and the degree to which the questions relate to the process being studied—that is, their validity. Expanded understanding of the psychometric properties of patient-reported measures has led to careful testing of these measures in appropriate populations, and their usefulness has been demonstrated in a growing number of clinical trials (**Box 2**). To achieve consistent and interpretable collection of data, a standardized approach for the administration of patient-reported measures must be planned for in the design, execution, analysis and interpretation of the study.

Questionnaires are developed for one of two purposes: to identify people with a specific condition within a population (discriminative), or to evaluate the change in a condition over time (evaluative). Although some scales aim to do both, the features that distinguish a specific condition are often not the best ones to follow over time. For instance, the SF-36 is a widely used questionnaire that provides an accurate estimate of an individual's quality of life across a wide range of conditions compared to a well-validated global population standard. As it is designed to cover a broad range of the population, it is not useful as a measure of change in people who comprise only a small portion of that range. Similarly, signs or symptoms that may be useful in establishing the diagnosis of neuropathic pain (for example, loss of sensation or mild muscle weakness in a dermatome) may not be appropriate for following improvement over time. Scales such as the Neuropathic Pain Diagnostic Questionnaire[27] and the Leeds Assessment of Neuropathic Symptoms and Signs Pain Scale[28], which are based on analysis of sensory description and bedside examination of sensory findings, are useful to identify individuals with neuropathic pain syndromes but are not generally used to assess changes in pain over time. The Neuropathic Pain Scale[29] and the Neuropathic

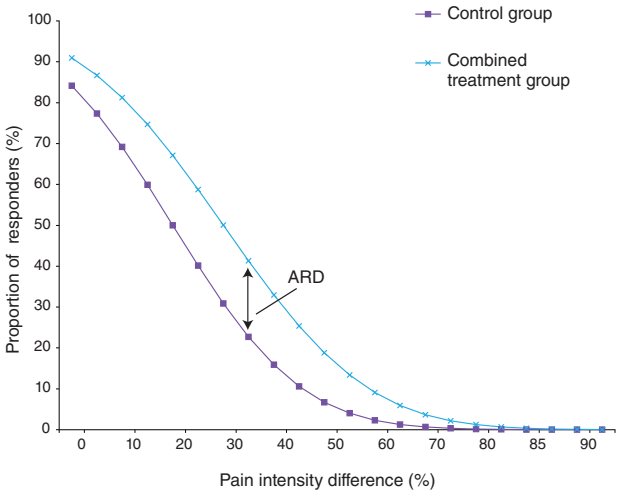

Figure 3 Cumulative distribution function curve for a simulated analysis of two groups, placebo and combined treatment created from data used for the graph in **Figure 2**. The *x* axis represents cut-off points in percentage of change in pain intensity, and the *y* axis shows the cumulative proportion of subjects that have achieved that level of response or higher. For example, for the combined treatment group line at the 50% response mark, approximately 19% of subjects achieved that level or more. The distance between the two lines (placebo and combined treatment) is the absolute risk difference (ARD) between the two groups, and 1 / ARD provides an estimate of the number needed to treat (NNT).

Box 2 Structure of self-report measures in pain clinical trials

Every item of a self-reported measure of health consists of a question and a scale used to report the answer. The format of each question contains several common elements, such as the condition being measured, the time frame, and modifiers of each, as needed.

The reporting scales can be yes or no, descriptive words, numbers, a line or diagram, a range of representative pictures or colors, or other indicators of change. VRSs consist of ordered categorical words, where the value of the amount of change in pain needed to move between adjacent pairs of words varies depending on the word pair, so such scales are best treated as ordinal data. Verbal scales convey meaning but are harder to translate across cultures and languages. Number (for example, 0–10 NRS) and diagram scales (for example, the 100-mm VAS) are usually labeled with upper and lower bounds but have no intrinsic meaning at any given level in between. When pain is measured on the same numeric scale at different time points, it provides a clear indication of change, which is usually best represented as a proportional change[36,69]. The VAS requires a diagram, whereas the NRS can be either written or verbally administered. Numbers and diagrams tend to translate more consistently across cultures. Serial pictures or color scales also do not have intrinsic meaning for a given level and require a display. They may be more appropriate for young children or nonverbal adults, but can have different meanings in different cultures.

Psychometric testing for the reliability, validity and responsiveness of single questions or multi-item questionnaires enhances confidence in the value of the data. When appropriately constructed, patient-reported measures consistently produce accurate answers to the research questions they are designed to address. Although a number of the most common questionnaires are robust and can be used across a broad range of studies, careful consideration should be given to each particular use, as subtle changes in the questions used or the population of interest can affect the results.

Table 2 Pain-associated symptoms used to compare results of chronic pain clinical trials[39,40]

Primary outcome domains	Description
Pain	Measures include pain intensity (11-point (0–10) numerical rating scale or categorical verbal rating scale, if necessary) and usage of rescue analgesic medications (for example, opioids)
Physical functioning	Subscale component of interference scales from the Multidimensional Pain Inventory Interference Scale or the Brief Pain Inventory interference items; if relevant, a more complete measure or activity test can be used
Emotional functioning	Test of emotional state; Beck Depression Inventory and Profile of Mood States suggested, but other scales acceptable
Participant ratings of global improvement	Predominant measure used is the Patient Global Impression of Change; other measures also possible
Symptoms and adverse events	Passive capture of spontaneously reported adverse events is acceptable, but collection of symptoms via a specific questionnaire and use of open-ended prompts are encouraged
Supplementary outcome domains	Measures of pain coping, social functioning and personality may be useful

In addition to pain, it is crucial to measure other signs and symptoms known to be associated with the experience of pain. To evaluate the impact of any treatment on chronic pain, multiple outcome domains need to be assessed.

Pain Questionnaire[30] measure only symptoms, are easier to administer but may not be as good at differentiating neuropathic pain from nociceptive pain. However, simpler measures of pain intensity or pain relief are commonly used to follow individuals over time once their pain diagnosis is established.

Evaluative measures for pain include a broad range of questions and scales used to measure the response (**Table 1**). Pain intensity is the primary measure for many clinical trials of pain therapies. Even the relatively simple concept of pain intensity can be broken down into three components by asking about pain on average, pain at its worst and pain at its least, as is done in the commonly used Brief Pain Inventory[31]. Single- or multiple-item pain measures can also assess other pain features, such as frequency and distress. In research and clinical settings, the predominant scales used to evaluate pain intensity include verbal rating scales (VRSs), numerical rating scales (NRS) and the visual analog scale (VAS) (**Box 2**). All three types of scale produce interchangeable results as long as they are properly administered[32,33], and all are highly correlated, although the numerical and visual analog scales correlate most strongly[34]. These pain scales are reliable for multiple measurements within a single individual, but there is high variability among individuals with chronic pain. This variability makes it difficult to interpret the clinical importance of single pain measurements, but the change over time is a reliable and valid outcome for clinical trials. The improvement in pain within an individual, expressed as a percent change to control for different baseline pain intensities, is highly correlated with his or her overall condition[35,36].

It is also common to measure the effect of pain on various aspects of a person's life, such as mood, enjoyment of life, ability to function, social interactions and sleep. These pain interference scales are also part of the 12-item Brief Pain Inventory and the 50-item Multidimensional Pain Inventory[37]. Health-related quality-of-life measures are also used to assess the effect of pain. Quality-of-life questionnaires usually measure the level of physical, psychological, social and spiritual function, along with the health status of the individual, including pain. Pain can be a crucial contributor to quality of life, but it is not innately more important than other components. In light of this, most studies of pain should consider including measures of other components to improve the interpretation of the results. Groups such as the Outcome Measures in Rheumatoid Arthritis Clinical Trials[38] and the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials[39,40] have applied the best current evidence to define a core group of measures that should be included in each pain clinical trial to allow easier comparison of results across trials, in addition to other specialized measures needed (**Table 2**).

Analysis and interpretation of pain clinical trials. A data analysis comprises a summary of the size of the effect measured in the treated group compared to the control group, along with a test of the statistical probability that the results occurred by chance, with an arbitrary cutoff of $P \geq 0.05$ (that is, a 1 in 20 chance or more). For even small differences between groups, a positive $P < 0.05$ can usually be achieved by increasing the sample size of the study. Although the probability is necessary to promote confidence in the findings, the statistical test has no direct applicability to the clinical relevance of the trial (**Box 1**). Reporting statistical significance without the size of the effect renders the reported results of a clinical trial uninterpretable, as the P value is is often highly dependent on the number of subjects included in the trial.

The size of the effect can be represented in one of two forms: a central point, such as a mean or median, with s.d.; or a proportion of subjects who have responses of a specific size (also called a responder analysis). Researchers can produce variations of these two types of numbers by manipulating their form (for example, converting them to ratios, percentages or person-time measurements) or by applying a variety of units (for example, time, events, people or area under a curve). Information collected as ordered categorical data with more than one level, such as the pain VRS, or as continuous data, such as blood pressure, can be summarized as a central point or as a proportion of subjects who have a specified level of response. To present continuous data as a proportion, researchers must define a cutoff point representing a scientifically or clinically relevant response (**Box 3**). A proportion of responders is generally thought to be easier to apply in a clinical setting.

These two data summary forms present different characteristics of the same clinical trial data. For example, if the central point of a normally distributed data set is presented as a median, then by definition 50% of the subjects achieved a level of response at or above the median (**Fig. 2**). The number of subjects above or below any other point along the curve can be calculated using the mean and s.d. and the standard normal distribution formula. Similarly, choosing a single cutoff point for a responder analysis will provide the proportion of subjects above and below a given value. The proportion of responders can also be presented for a set of cutoff points or for the range of all possible cutoff points on the measurement scale used to collect the data. A display of the cumulative number of subjects at or above all possible cutoff points is called a cumulative distribution function curve (**Fig. 3**). Each point on the cumulative distribution curve represents the proportion of subjects at that point or to the right in the normal distribution curve—for example, the proportion of subjects with that much change in pain score or more. In a graph of the proportion of responders over the full range of possible cutoff points, users of the data can find the proportion of people at any cutoff point that may be appropriate for their needs.

With current advanced statistics, nonparametric analyses applied to the cumulative distribution function will have nearly the same statistical power to detect a difference without assuming a normal distribution. Nonparametric statistics are more appropriate in the analysis of popu-

Box 3 Data-driven methods to determine a clinically important level of response to a pain therapy

The clinician's goal is to predict whether a patient will feel better with a given treatment. To answer this question using the data collected from clinical trials, information is needed about the probability of study subjects experiencing improvement with the treatment. Pain investigators should show what constitutes a clinically important improvement in a subject receiving a pain therapy, defined as the change in pain intensity that correlates with subjects reporting that they are 'better'. Initially, pain experts chose a 50% change as a 'reasonable' estimate[41].

Subsequently, two data-driven methods of analysis were investigated: distribution-based and anchor-based methods[73]. Distribution-based methods determine the minimal change perceived by the individual that would exceed the physiologic and measurement-derived error inherent in the study. This is called the minimally important difference and should be considered the minimal detectable difference for the measures used, rather than a reflection of how subjects perceive the importance of an effect. Before values are compared across studies, the definitions of CID in each study should be carefully reviewed. The individual anchor-based methods, in contrast, establish the CID from the individual's perspective, which is a more clinically useful number. Application of these two methods to data from acute and chronic pain studies using various subject-oriented global outcome measures has established that approximately a 30% change in pain intensity is consistently associated with individuals reporting notable improvement[35,36]. A 20% change in pain is associated with individuals reporting that they were at least slightly better (sometimes referred to as the minimal clinically important difference, or MCID), and a 50% change in pain is best associated with individuals reporting substantial improvement (also called a clinically substantial difference). These values have been shown to be remarkably consistent across pain syndromes, patient populations and treatment options[54,63].

lations of people, some of whom have the potential to respond to the study therapy and some who do not. The potential for response can be based on differences in genetic profiles, previous and current environmental exposures, and underlying mechanisms of pain. For example, 10–25% of the US population is unable to properly metabolize the pain-killer codeine to morphine. As another example, consider a clinical trial in which the hypothetical active treatment is given to a group consisting of a 50/50 mix of subjects who have the inherent potential to respond and subjects who do not. In such a case, it would be reasonable to expect the subjects without the potential to respond to mimic the response that occurs in the control group, producing a data curve similar to the smaller curve on the left of **Figure 2**, which has the same mean and distribution as the control-group curve. If the treatment is effective, subjects with the potential to respond will have a larger response than the control group, producing a data curve with a median higher than the control group (smaller curve on the right in **Fig. 2**) and creating a bimodal distribution. Although the combined treatment curve in **Figure 2** is probably close enough to a normal distribution to allow the application of a parametric analysis, the more appropriate analysis of a bimodal distribution would be a nonparametric statistical test such as a Wilcoxon rank sum, or ordinal regression, neither of which make any parametric assumptions.

To better understand the how to apply clinical trial data to the clinical care of patients, pain investigators still strive to determine what constitutes a clinically important level of response to a pain therapy. A 50% reduction in pain intensity has been selected by experts as a 'reasonable' cutoff point for a clinically important response, allowing the systematic comparison of trials using meta-analytic techniques[41]. Data-derived methods have identified a 30% change as the amount of response on the 0–10 NRS scale that is best associated with subjects reporting that they are "much better" on a global-impression-of-change scale[36], which is the clinical response that is considered to be a clinically important difference (CID; **Box 3**). Analyses using "very much better" on the global-impression-of-change scale have found a cutoff point of approximately 50% change, validating the standard previously selected by expert opinion. Both cutoff points have become crucial in the complete reporting of clinical trial results.

Although detecting CIDs in individuals is crucial for presenting the number of subjects who have a clinically important response, studies must, in addition, compare the numbers of responders in the treatment and placebo groups (**Fig. 3**). Determining what constitutes a clinically relevant difference between groups involves a broader decision about the overall benefit that is necessary for the treatment to be considered worth the risk. This can be influenced by the availability of other treatments, the safety of the treatment, whether there is a subgroup of individuals who achieve substantial relief, and who is making the decision. The difference in response between the treated and placebo groups is the appropriate value to consider in a risk-benefit analysis. Yet with regard to treatment of individuals, the proportion of subjects in the treated group who respond is the most reasonable estimate for the proportion of people who may benefit from the treatment. In clinical practice, the treatment response also includes the placebo response.

One summary statistic that has been promoted for evaluating the difference between groups in a clinical trial is the number needed to treat (NNT)[42], which estimates the number of people who need to be treated for one person to benefit from the treatment. The difference between the number of responders in the treated group and that in the placebo group is known as the absolute risk difference (ARD; **Fig. 3**); the NNT is calculated by dividing 1 by the ARD. In early work with the NNT, a response level of 50% was used[41], but often the response cutoff has not been defined or stated. The difference between groups varies substantially depending on the response cutoff point used (**Fig. 3**). Recent reports have annotated the NNT with the level of response used, so that an NNT of the 50% response, for example, would be the NNT-50 (ref. 43). Given the variability in the NNT, a more appropriate estimate may be an average calculated as a difference in area under the curve, over a therapeutically important range of response, such as 20% to 70%. In addition, questions have been raised about the usefulness of the NNT as a summary measure for comparison across studies, given that it can vary depending on the trial design, especially the measures and analyses used[43]. Although clinically useful, researchers must exercise substantial caution when comparing NNTs across studies.

Large amounts of missing data also reduce confidence in study results. The causes of missing data vary widely and include inefficient data collection procedures, poor study implementation and the loss of subjects who drop out of the study because of side effects, lack of efficacy or procedural issues. An example of a design that minimizes the impact of occasional missing data points is the collection of more data than is required for the analysis. An additional method is to continue to collect data on subjects who need to stop taking the study medication, which is not synonymous with dropping out of the study[44].

There are a broad range of analytic methods for handling missing data, but all makeup data that was not recorded. Two commonly used methods that are relatively simple but less precise are last observation carry forward and baseline observation carry forward. More complex but precise techniques include the generalized linear mixed model, maximum likelihood ratios, Bayesian methods, generalized estimating equations and multiple imputation methods. In general, these methods attempt to use the information collected from the individuals who dropped out compared to those who did not, to better estimate values for the missing data. All such methods share two common features. First, they generate a best guess for the data that is missing. Second, each is based on a set of assumptions about the behavior of study participants. It is crucial to specify *a priori* how missing data will be handled for the primary analysis. A reasonable approach could involve conducting a set of sensitivity analyses using different approaches to examine the consequences of specific assumptions. If the different analyses generally agree, there can be more confidence that the result is an accurate representation of the data. If not, the results then must be considered with caution[44], especially in a regulatory environment.

Alternative study designs

With researchers turning toward *a priori* identification of populations with the potential to respond, enrichment designs have become more frequent. In an enrichment design, every potential subject is started on the study treatment, and those who can tolerate the treatment and demonstrate any degree of improvement are enrolled in the pain clinical trial. Two study designs are possible. First, subjects who improve can undergo a pharmacologically appropriate washout period before being randomly assigned to either the study drug or placebo group. The second approach is a randomized withdrawal trial in which responders are randomly chosen either to continue taking the active drug or to switch to a placebo. The difference in efficacy is then measured over time. In each of these designs, the individuals who stop their medication because of lack of efficacy can be considered nonresponders for the analysis.

A commonly used design that maximizes the likelihood of demonstrating some benefit from a novel pain therapy is a crossover study. Because each subject is exposed to both the treatment and the control therapy, each individual also serves as a control, the design has substantially more statistical power and fewer subjects are needed. These types of studies, however, are very sensitive to changes in the subject over time, carryover effects and patient dropout in the first period. For proof-of-concept studies, the greater efficiency of crossover studies may outweigh these drawbacks, but the inherent problems with this design limit its use in definitive studies of efficacy.

Conclusions

A randomized, double-blind clinical trial to detect the specific effect of a therapy for the treatment of pain is the primary method used to assess efficacy. The subjective nature of the experience and reporting of pain, along with the potential influence on the pain process of various factors specific to each individual, justify carefully designing and conducting a blinded and randomized placebo-controlled study.

This review has presented several crucial issues pertaining to the design, measurement, analysis and interpretation of data in clinical studies of pain and pain therapies, emphasizing problems that affect the quality of the results obtained and the methods available to overcome these barriers. The aim was to present provocative concepts and challenge researchers to consider how best to study pain. Ideally, as scientists strive for an evidence-based approach to medicine, the same approach will be brought to designing pain clinical trials, using knowledge gained from ongoing clinical studies[18,45]. Further understanding of how to study the treatment of pain may rapidly expand the availability of treatments based on comprehension of the mechanisms underlying pain.

ACKNOWLEDGMENTS

I would like to acknowledge R.H. Dworkin, N. Katz, B.A. Rappaport, D.C. Turk and the late M.B. Max for their roles in educating me on the subject matter of this article over many years, and E. Weisberg for his editorial help.

COMPETING FINANCIAL INTERESTS

The author declares no competing financial interests.

Published online at http://www.nature.com/naturemedicine/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Merskey, H. & Bogduk, N. Part III: Pain terms—a current list with definitions and notes on usage. in *Classification of Chronic Pain, Second Edition, IASP Task Force on Taxonomy* 209–214 (IASP, Seattle, 1994).
2. Treede, R.D. *et al.* Neuropathic pain: redefinition and a grading system for clinical and research purposes. *Neurology* **70**, 1630–1635 (2008).
3. Hill, A.B. The Clinical Trial. *N. Engl. J. Med.* **247**, 113–119 (1952).
4. Lasagna, L. & Beecher, H.K. Analgesic effectiveness of nalorphine and nalorphine–morphine combinations in man. *J. Pharmacol. Exp. Ther.* **112**, 356–363 (1954).
5. Houde, R.W. & Wallenstein, S.L. Clinical studies of morphine–nalorphine combinations. *Fed. Proc.* **15**, 440–441 (1956).
6. Beecher, H.K. Measurement of Subjective Responses: Quantitative Effects of Drugs (Oxford University Press, Oxford, 1959).
7. Melzack, R. & Wall, P.D. Pain mechanisms: a new theory. *Science* **150**, 971–979 (1965).
8. Dworkin, R.H. *et al.* Placebo and treatment group responses in postherpetic neuralgia vs. painful diabetic peripheral neuropathy clinical trials in the REPORT database. *Pain* **150**, 12–16 (2010).
9. Basbaum, A.I. *et al.* Cellular and molecular mechanisms of pain. *Cell* **139**, 267–284 (2009).
10. Bouhassira, D. *et al.* Prevalence of chronic pain with neuropathic characteristics in the general population. *Pain* **136**, 380–387 (2008).
11. Torrance, N. *et al.* The epidemiology of chronic pain of predominantly neuropathic origin. Results from a general population survey. *J. Pain* **7**, 281–289 (2006).
12. Deng, W., Aimone, J.B. & Gage, F.H. New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory? *Nat. Rev. Neurosci.* **11**, 339–350 (2010).
13. Woolf, C.J. & Mannion, R.J. Neuropathic pain: aetiology, symptoms, mechanisms, and management. *Lancet* **353**, 1959–1964 (1999).
14. Shavit, Y. *et al.* Postoperative pain management and proinflammatory cytokines: animal and human studies. *J. Neuroimmune Pharmacol.* **1**, 443–451 (2006).
15. Moiniche, S. *et al.* A qualitative and quantitative systematic review of preemptive analgesia for postoperative pain relief: the role of timing of analgesia. *Anesthesiology* **96**, 725–741 (2002).
16. Woolf, C.J., American College of Physicians & American Physiological Society. Pain: moving from symptom control toward mechanism-specific pharmacologic management. *Ann. Intern. Med.* **140**, 441–451 (2004).
17. Lee, S., Walker, J.R., Jakul, L. & Sexton, K. Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation. *Depress. Anxiety* **19**, 10–19 (2004).
18. Dworkin, R.H. *et al.* Placebo response in clinical trials of depression and its implications for research on chronic neuropathic pain. *Neurology* **65**, S7–S19 (2005).
19. Kundt, G. & Kundt, G. Comparative evaluation of balancing properties of stratified randomization procedures. *Methods Inf. Med.* **48**, 129–134 (2009).
20. Temple, R.J., Meyer, R., Temple, R.J. & Meyer, R. Continued need for placebo in many cases, even when there is effective therapy. *Arch. Intern. Med.* **163**, 371–373 (2003).
21. Boutron, I. *et al.* Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med.* **4**, e61 (2007).
22. Turner, J.A. *et al.* Blinding effectiveness and association of pretreatment expectations with pain improvement in a double-blind randomized controlled trial. *Pain* **99**, 91–99 (2002).
23. Irizarry, M.C. *et al.* Predictors of placebo response in pooled lamotrigine neuropathic pain clinical trials. *Clin. J. Pain* **25**, 469–476 (2009).
24. Boutron, I. *et al.* Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med.* **3**, e425 (2006).
25. Bingel, U., Tracey, I., Bingel, U. & Tracey, I. Imaging CNS modulation of pain in humans. *Physiology (Bethesda)* **23**, 371–380 (2008).
26. Tracey, I. Imaging pain. *Br. J. Anaesth.* **101**, 32–39 (2008).
27. Bouhassira, D. *et al.* Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). *Pain* **114**, 29–36 (2005).
28. Bennett, M. The LANSS Pain Scale: the Leeds assessment of neuropathic symptoms and signs. *Pain* **92**, 147–157 (2001).

29. Galer, B.S. & Jensen, M.P. Development and preliminary validation of a pain measure specific to neuropathic pain: the Neuropathic Pain Scale. *Neurology* **48**, 332–338 (1997).
30. Krause, S.J. & Backonja, M.M. Development of a neuropathic pain questionnaire. *Clin. J. Pain* **19**, 306–314 (2003).
31. Cleeland, C.S. & Ryan, K.M. Pain assessment: global use of the Brief Pain Inventory. *Ann. Acad. Med. Singapore* **23**, 129–138 (1994).
32. Hurny, C.B.J. *et al.* Responsiveness of a single-item indicator versus a multi-item scale: assessment of emotional well-being in an international adjuvant breast cancer trial. *Med. Care* **34**, 234–248 (1996).
33. Breivik, E.K., Bjornsson, G.A. & Skovlund, E. A comparison of pain rating scales by sampling from clinical trial data. *Clin. J. Pain* **16**, 22–28 (2000).
34. McDowell, I. & Newell, C. Pain measurements. in *Measuring Health: A Guide to Rating Scales and Questionnaires* Ch. 8 (Oxford University Press, Oxford, 1996).
35. Farrar, J.T., Portenoy, R.K., Berlin, J.A., Kinman, J. & Strom, B.L. Defining the Clinically Important Difference in Pain Outcome Measures. *Pain* **88**, 287–294 (2000).
36. Farrar, J.T., Young, J.P. Jr., LaMoreaux, L., Werth, J.L. & Poole, R.M. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* **94**, 149–158 (2001).
37. Kerns, R.D., Turk, D.C. & Rudy, T.E. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain* **23**, 345–356 (1985).
38. Bellamy, N. *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J. Rheumatol.* **24**, 799–802 (1997).
39. Dworkin, R.H. *et al.* Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* **113**, 9–19 (2005).
40. Turk, D.C. *et al.* Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain* **106**, 337–345 (2003).
41. Moore, A., McQuay, H. & Gavaghan, D. Deriving dichotomous outcome measures from continuous data in randomised controlled trials of analgesics. *Pain* **66**, 229–237 (1996).
42. Cook, R.J. & Sackett, D.L. The number needed to treat: a clinically useful measure of treatment effect. *Br. Med. J.* **310**, 452–454 (1995).
43. Edelsberg, J., Oster, G., Edelsberg, J. & Oster, G. Summary measures of number needed to treat: how much clinical guidance do they provide in neuropathic pain? *Eur. J. Pain* **13**, 11–16 (2009).
44. Committee on National Statistics. The prevention and treatment of missing data. in *Clinical Trials: Panel on Handling Missing Data in Clinical Trials* Ch. 2 and Ch. 5 (The National Academies Press, Washington, DC, 2010).
45. Katz, N. Methodological issues in clinical trials of opioids for chronic pain. *Neurology* **65**, S32–S49 (2005).
46. Keller, S. *et al.* Validity of the brief pain inventory for use in documenting the outcomes of patients with noncancer pain. *Clin. J. Pain* **20**, 309–318 (2004).
47. Melzack, R. The short-form McGill Pain Questionnaire. *Pain* **30**, 191–197 (1987).
48. Heft, M.W., Gracely, R.H., Dubner, R. & McGrath, P.A. A validation model for verbal description scaling of human clinical pain. *Pain* **9**, 363–373 (1980).
49. Portenoy, R.K., Thaler, H.T. & Kornblith, A.B. The Memorial Symptom Assessment Scale: An instrument for the evaluation of symptom prevalence, characteristics, and distress. *Eur. J. Cancer* **30A**, 1326–1336 (1994).
50. Serlin, R.C., Mendoza, T.R., Nakamura, Y., Edwards, K.R. & Cleeland, C.S. When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. *Pain* **61**, 277–284 (1995).
51. Farrar, J.T. *et al.* A comparison of change in the 0–10 numeric rating scale to a pain relief scale and global medication performance scale in a short-term clinical trial of breakthrough pain intensity. *Anesthesiology* **112**, 1464–1472 (2010).
52. Fishman, B., Houde, R.W., Wallenstein, S. & Foley, K.M. Appraisal of pain relief mediates the relationship between pain intensity and emotional distress in cancer pain. *Pain* **30**, S415 (1987).
53. Sloan, J.A. *et al.* Assessing the clinical significance of single items relative to summated scores. *Mayo Clin. Proc.* **77**, 479–487 (2002).
54. Farrar, J.T. *et al.* The clinical importance of changes in the 0 to 10 numeric rating scale for worst, least, and average pain intensity: analyses of data from clinical trials of duloxetine in pain disorders. *J. Pain* **11**, 109–118 (2010).
55. Guy, W. Clinical Global Impressions. in *ECDEU assessment manual for psychopharmacology.* 217–222 (US Department of Health and Human Services, Public Health Service, Alcohol Drug Abuse and Mental Health Administration, NIMH Psychopharmacology Research Branch, Rockville, Maryland, 1976).
56. Bergh, I., Sjostrom, B., Oden, A. & Steen, B. An application of pain rating scales in geriatric patients. *Aging Clin. Exp. Res.* **12**, 380–387 (2000).
57. Flaherty, S.A. Pain measurement tools for clinical practice and research. *AANA J.* **64**, 133–140 (1996).
58. Gracely, R.H., Dubner, R., McGrath, P. & Heft, H. New methods of pain measurement and their application to pain control. *Int. Dent. J.* **28**, 52–65 (1978).
59. Baker, R. Measuring change in health in general practice: a comparison of a simple transition question with the Nottingham Health Profile. *Int. J. Qual. Health Care* **10**, 207–212 (1998).
60. Fitzpatrick, R., Ziebland, S., Jenkinson, C. & Mowat, A. Transition questions to assess outcomes in rheumatoid arthritis. *Br. J. Rheumatol.* **32**, 807–811 (1993).
61. Jenkinson, C. & Ziebland, S. Interpretation of data from health status measures: What do the numbers mean? in *Individual quality of life: Approaches to conceptualisation and assessment.* (eds. Joyce, C.R.B. & McGee, H.M.) 75–86 (Harwood Academic Publishers, Amsterdam, 1999).
62. Jensen, M.P., Turner, J.A. & Romano, J.M. What is the maximum number of levels needed in pain intensity measurement? *Pain* **58**, 387–392 (1994).
63. Cepeda, M.S., Africano, J.M., Polo, R., Alcala, R. & Carr, D.B. Agreement between percentage pain reductions calculated from numeric rating scores of pain intensity and those reported by patients with acute or cancer pain. *Pain* **106**, 439–442 (2003).
64. Angst, M.S., Brose, W.G. & Dyck, J.B. The relationship between the visual analog pain intensity and pain relief scale changes during analgesic drug studies in chronic pain patients. *Anesthesiology* **91**, 34–41 (1999).
65. de Tovar, C. *et al.* Postoperative self-report of pain in children: interscale agreement, response to analgesic, and preference for a faces scale and a visual analogue scale. *Pain Res. Manag.* **15**, 163–168 (2010).
66. Decruynaere, C. *et al.* How many response levels do children distinguish on faces scales for pain assessment? *Eur. J. Pain* **13**, 641–648 (2009).
67. Tesler, M.D. *et al.* The word-graphic rating scale as a measure of children's and adolescents' pain intensity. *Res. Nurs. Health* **14**, 361–371 (1991).
68. Huskisson, E.C. Measurement of pain. *Lancet* **2**, 1127–1131 (1974).
69. Price, D.D., McGrath, P.A., Rafii, A. & Buckingham, B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain* **17**, 45–56 (1983).
70. Jamison, R.N. *et al.* Comparative study of electronic vs. paper VAS ratings: a randomized, crossover trial using healthy volunteers. *Pain* **99**, 341–347 (2002).
71. Marceau, L.D. *et al.* Electronic diaries as a tool to improve pain management: is there any evidence? *Pain Med.* **8** Suppl 3, S101–S109 (2007).
72. Jensen, M.P. & McFarland, C.A. Increasing the reliability and validity of pain intensity measurement in chronic pain patients. *Pain* [see comments] **55**, 195–203 (1993).
73. Guyatt, G.H. *et al.* Methods to explain the clinical significance of health status measures. *Mayo Clin. Proc.* **77**, 371–383 (2002).