Pain Physician 2008; 11:717-773 • ISSN 1533-3159

Health Policy Review

Evidence-Based Medicine, Systematic Reviews, and Guidelines in Interventional Pain Management: Part 2: Randomized Controlled Trials

Laxmaiah Manchikanti, MD[1], Joshua A. Hirsch, MD[2], and Howard S. Smith[3], MD

From:
[1]Pain Management Center of Paducah, Paducah, KY;
[2]Massachusetts General Hospital And Harvard Medical School, Boston, MA; and
[3]Albany Medical College Albany, NY

Dr. Manchikanti is Medical Director of the Pain Management Center of Paducah, Paducah, KY, and Associate Clinical Professor of Anesthesiology and Perioperative Medicine, University of Louisville, KY. Dr. Hirsch is Chief of Minimally Invasive Spine Surgery, Depts. of Radiology and Neurosurgery, Massachusetts General Hospital and Assistant Professor of Radiology, Harvard Medical School, Boston, MA. Dr. Smith is Associate Professor and Academic Director of Pain Management for Albany Medical College Department of Anesthesiology, Albany, NY.

Address correspondence:
Laxmaiah Manchikanti, MD
2831 Lone Oak Road
Paducah, KY 42003
E-mail: drlm@thepainmd.com

Disclaimer: There was no external funding in the preparation of this manuscript.
Conflict of interest: None.

Manuscript received: 10/7/2008
Accepted for publication: 10/15/2008

Free full manuscript:
www.painphysicianjournal.com

Evidence-based medicine (EBM) is a shift in medical paradigms and about solving clinical problems, acknowledging that intuition, unsystematic clinical experience, and pathophysiologic rationale are insufficient grounds for clinical decision-making. The importance of randomized trials has been created by the concept of the hierarchy of evidence in guiding therapy. Even though the concept of hierarchy of evidence is not absolute, in modern medicine, most researchers synthesizing the evidence may or may not follow the principles of EBM, which requires that a formal set of rules must complement medical training and common sense for clinicians to interpret the results of clinical research. N of 1 randomized controlled trials (RCTs) has been positioned as the top of the hierarchy followed by systematic reviews of randomized trials, single randomized trial, systematic review of observational studies, single observational study, physiologic studies, and unsystematic clinical observations. However, some have criticized that the hierarchy of evidence has done nothing more than glorify the results of imperfect experimental designs on unrepresentative populations in controlled research environments above all other sources of evidence that may be equally valid or far more applicable in given clinical circumstances.

Design, implementation, and reporting of randomized trials is crucial. The biased interpretation of results from randomized trials, either in favor of or opposed to a treatment, and lack of proper understanding of randomized trials, leads to a poor appraisal of the quality.

Multiple types of controlled trials include placebo-controlled and pragmatic trials. Placebo-controlled RCTs have multiple shortcomings such as cost and length, which limit the availability for studying certain outcomes, and may suffer from problems of faulty implementation or poor generalizability, despite the study design which ultimately may not be the prime consideration when weighing evidence for treatment alternatives. However, in practical clinical trials, interventions compared in the trial are clinically relevant alternatives, participants reflect the underlying affected population with the disease, participants come from a heterogeneous group of practice settings and geographic locations, and endpoints of the trial reflect a broad range of meaningful clinical outcomes.

Key words: Randomized controlled trial (RCT), placebo-controlled trial, pragmatic controlled trial, randomization, allocation concealment, sample size, blinding, consolidated standards of reporting trials (CONSORT) statement, minimal clinically important change (MCIC), minimal clinical important difference (MCID)

Pain Physician 2008; 11:6:717-773

Guyatt and Drummond (1) in their introduction to the philosophy of evidence-based medicine (EBM) begin with the assertion that EBM is a shift in medical paradigms and about solving clinical problems (2,3). They further elaborate that in contrast to the traditional paradigm of medical practice, EBM acknowledges that intuition, unsystematic clinical experience, and pathophysiologic rationale are insufficient grounds for clinical decision-making, and stresses the examination of evidence from clinical research. Further, EBM suggests that a formal set of rules must complement medical training and common sense for clinicians to interpret the results of clinical research effectively.

A hierarchy of strength of evidence for treatment decisions provided by Guyatt and Drummond (1) is as follows:

- N of 1 randomized controlled trial
- Systematic reviews of randomized trials
- Single randomized trial
- Systematic review of observational studies addressing patient-important outcomes
- Single observational study addressing patient-important outcomes
- Physiologic studies (studies of blood pressure, cardiac output, exercise capacity, bone density, and so forth)
- Unsystematic clinical observations

The N of 1 randomized controlled trial (RCT) which is at the top of Guyatt and Drummond's (1) hierarchy of strength of evidence for treatment decisions, is one in which patients undertake pairs of treatment periods receiving a target treatment during one period of each pair, and a placebo or alternative during the other. Patients and clinicians are blind to allocation, the order of the target and control is randomized, and patients make quantitative ratings of their symptoms during each period. The N of 1 RCT continues until both the patient and clinician conclude that the patient is, or is not, obtaining benefit from the target intervention. N of 1 RCT is often considered to be feasible (4,5), can provide definitive evidence of treatment effectiveness in individual patients, and may lead to long-term differences in treatment administration (6,7).

Knowing the tools of evidence-based practice is necessary but not sufficient for delivering the highest quality of patient care. In addition to clinical expertise, the clinician requires compassion, sensitive listening skills, and broad perspectives from the humanities and social sciences. Thus, a continuing challenge for EBM, and for medicine in general, will be to better integrate the new science of clinical medicine with the time-honored craft of caring for the sick (1).

The philosophy of a hierarchy of evidence in guiding therapy, though it is not absolute, has created emphasis on the importance of randomized trials. However, in modern medicine, most researchers synthesizing the evidence may or may not follow the rules of EBM, which require that a formal set of rules must complement medical training and common sense for clinicians to interpret the results of clinical research. In general, systematic clinical observations are limited by small sample size, and, more importantly, by deficiencies in the human processes of making inferences (8). At the same time, predictions about intervention effects on clinically important outcomes based on physiologic experiments while they are usually right, occasionally are disastrously wrong (1,9). Further, guideline developers and systematic reviewers seem to ignore that very different hierarchies are necessary for issues of diagnosis or prognosis (9).

Essentially, EBM has been characterized as a stick by which policy-makers and academicians beat clinicians (10-14). Further, it has been claimed that the research performed to test new treatments has often been of poor quality or has asked the wrong questions. Thus, clinicians criticize the research establishment, justifiably, for failing to provide answers to relevant clinical problems of everyday practice (15).

Miles et al (15) state that, "the hierarchy of evidence (of which RCT has the highest priority) has done nothing more than glorify the results of imperfect experimental designs on unrepresentative populations in controlled research environments above all other sources of evidence which may be equally valid or far more applicable in given clinical circumstances."

1. An Introduction to Randomized Trials

Randomized trials are considered as the evidence of progress in medicine. However, randomized trials work by first assuming there is no difference between a new and an old or placebo treatment – the null hypothesis. Basically, one may contend that the standard RCTs are in fact set up to show that treatments do not work, rather than to demonstrate that treatments do work (16). RCTs were designed to stop therapeutic bandwagons in their tracks and also quacks pedaling worthless treatments to patients made vulnerable and desperate by their illness.

Historically, randomized trials originated with epidemiology in the nineteenth century attempting to establish causation of infectious disorders (16). Introduced by Fisher in 1926 in an agricultural study (17), randomization was designed to reduce the required population to be tested from tens of thousands of patients to controllable levels. Subsequently, the two RCTs were published, one in 1931 (18) and the other in 1948 (19).

While RCTs are considered to provide the most internally valid evidence for medical decision-making, in the specialty of interventional pain management, results from clinical trials, both randomized and observational, with substantial impact on patient care, have been ruled ineffective based on flawed methodology of evidence synthesis (20-39).

In an attempted meta-analysis, Smith and Pell (40) reviewed the available randomized trials supporting the use of parachutes to prevent injuries caused by jumping out of an airplane. There were no trials available which had been done and they concluded that there was insufficient evidence to recommend the use of parachutes. Realizing that very few interventions in medicine work quite as definitely as parachutes, this attempted meta-analysis reminds us that some interventions are of such intuitive value that they do not require RCTs. The traditional view is that the most reliable evidence in medicine comes from blinded RCTs. This is the only design which is considered to reliably control for unobserved differences between treated and untreated patients (41).

2. Why Randomized Trials?

With increasing initiatives to improve the effectiveness and safety of patient care, there is a growing emphasis on evidence-based interventional pain management and incorporation of high-quality evidence into clinical practice. The World Health Organization (WHO) defines a clinical trial as, "any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes" (42). Very few studies in interventional pain management are RCTs and treatments even in surgery are only half as likely to be based on RCTs as treatments in internal medicine (20-39,43-46).

Many surgical and medical interventions recommended based on observational studies (46-51) have later been demonstrated to be ineffective or even harmful, even though there is contradictory evidence for RCTs also (6). The major advantage is that rigorously conducted RCTs minimize bias by controlling for known and unknown factors (confounders) that affect outcomes and distort the apparent treatment effect. However, not all questions can be addressed in an RCT. Evidence shows that only 40% of treatment questions involving surgical procedures are amenable to evaluation by an RCT, even in an ideal clinical setting (52-55). In fact, among the 4 trial objectives including measurement of the effect size, existence of effect, dose-response relationship, and comparison of therapies, placebo-controlled trials measure only the first 2 (56).

Different outcomes can also be observed in trial participants because of either the Hawthorne or placebo effect, both of which can distort the apparent treatment effect and threaten the validity of the trial (41). The Hawthorne effect is described as changes in clinicians' or patients' behavior because of being observed, improving the results. In contrast, the placebo effect occurs from patients' expectations for benefit (57-62). In the evaluation of interventional techniques, in most instances, researchers and practitioners are not aware of the effects of solutions injected into closed spaces, joints, and over the nerve roots (63-72). Many authors also have considered local anesthetic injection as placebo (35,65,66,73-75). However, the evidence has been to the contrary (65,67-80).

RCTs have become the gold standard for assessing the effectiveness of therapeutic agents (81-83). Sacks et al (84) compared published RCTs with those that used observational designs. This evaluation showed that the agent being tested was considered effective in 44 of 56 trials (79%) in observational studies utilizing historic controls, whereas the agent was considered positive in only 10 of 50 (20%) RCTs. This led to the conclusion that bias in patient selection may irretrievably weigh the outcome of historically controlled trials in favor of new therapies in observational studies. However, the concept that assignment of the subjects randomly to either experimental or controlled groups as the perfect science has been questioned (85).

It was also reported that in comparing effects between RCTs and observational studies in digestive surgery, one-fourth of the observational studies gave different results than randomized trials and between-study heterogenicity was more common in observational studies in the field of digestive surgery (86). A potential for confounding bias was reported in 98% of the studies in one systematic review due to poor qual-

ity of reporting in observational intervention studies (87). In fact, another systematic review (88) concluded that tools for assessing quality and susceptibility to bias in observational studies in epidemiology should be rigorously developed, evidence-based, valid, reliable, and easy to use.

Hartz et al (89), in a 2005 publication assessing observational studies of medical treatments, concluded that reporting was often inadequate to compare study designs or allow other meaningful interpretation of results. However, Benson and Hartz (90), in a 2000 publication comparing observational studies and RCTs, found little evidence that estimates of treatment effects in observational studies reported after 1984 were either consistently larger than or qualitatively different from those obtained in RCTs. Hartz et al (91), in assessing observational studies of spinal fusion and chemonucleolysis in a 2003 publication, concluded that the results suggested that review of several comparable observational studies may help evaluate treatment, identify patient types most likely to benefit from a given treatment, and provide information about study features that can improve the design of subsequent observational or RCTs. They also cautioned that the potential of comparative observational studies has not been realized because of concurrent inadequacies in their design, analysis, and reporting. In contrast, Concato et al (92), in a 2000 publication evaluating published articles in 5 major medical journals from 1991 to 1995 concluded that the results of well designed observational studies (with either a cohort or a case control design) do not systematically overestimate the magnitude of the effects of treatment as compared with those in RCTs on the same topic. Further, Shrier et al (93) found that the advantages of including both observational studies and randomized studies in a meta-analysis could outweigh the disadvantages in many situations and that observational studies should not be excluded a priori.

Deeks et al (94) compared the results of randomized and non-randomized studies across multiple interventions using meta-epidemiological techniques. They concluded that the results of non-randomized studies sometimes, but not always, differ from results of randomized studies of the same intervention. However, non-randomized studies may still give seriously misleading results when treated and control groups appear similar in key prognostic factors.

3. What Is Randomization?

Randomization is the process of assigning participants to treatment and control groups, assuming that each participant has an equal chance of being assigned to any group (41,95,96). Thus, randomization is considered as a fundamental aspect of scientific research methodology. Fisher (17) was the first to introduce the idea of randomization in a 1926 agricultural study. Since then, the academic community has deemed randomization an essential tool for unbiased comparisons of treatment groups, resulting in the publication of the first RCT involving tuberculosis (18). This study included a total of 24 patients randomized by the flip of a coin. Even though randomization may be accomplished by a simple coin toss, sophisticated methodology has been developed, and is demanded in clinical trials.

While researchers believe that randomization ensures that 2 groups will differ only by chance, it does not guarantee that the balance will actually be achieved through randomization (37,66,97). In fact, Manchikanti and Pampati (97) in an evaluation of the research designs in interventional pain management undertook to evaluate if randomization does provide the protective statistical shield that some think it provides. The results of this evaluation showed that there was only one significant difference when patients were allocated by means of non-randomization among the groups or compared to the total sample. In contrast, randomization showed significant differences in 7 parameters indicating that a randomized design may not be the best in interventional pain management settings.

Carragee et al (37) in a Task Force report criticized Lord et al (66) for a lack of similarities in both groups even though they were randomized. The Task Force (37) stated that the randomization process resulted in an unequal distribution of potentially confounding baseline variables. They cite that 10 of 12 sham vs 4 of 12 active subjects were involved in litigation; however supporters of the procedure (98) contend that unequal numbers is evidence of the honesty of the randomization procedure. Further the implied criticism that litigation biases outcomes from radiofrequency neurotomy was not supported by the difference in the active group between those patients involved in litigation and those not involved in litigation. Additionally, 4 subsequent radiofrequency neurotomy studies (99-102) have consistently shown that litigation does not significantly affect outcomes statistically.

Some researchers have inaccurately claimed that diagnostic studies require randomization (20,21). However, it has long been shown that the quality assessment of diagnostic studies always involves observational studies rather than randomized, double-blind trials (22,103-107).

4. How To Report Randomized Trials

The revised Consolidated Standards of Reporting Trials (CONSORT) statement (108) for reporting randomized trials and the extension of the CONSORT statement of reporting of non-inferiority and equivalence randomized trials (109) acknowledge that well-designed and properly executed RCTs provide the best evidence on the efficacy of health care interventions, but trials with inadequate methodologic approaches are associated with exaggerated treatment effects (110-114). In an evaluation of standards of reporting randomized trails in general surgery (54), of the 69 RCTs analyzed, only 37.7% had a Jadad score of greater than or equal to 3, and only 13% of the trials clearly explained allocation concealment. They concluded that the overall quality of reporting surgical RCTs was suboptimal with a need for improving awareness of the CONSORT statement among authors, reviewers, and editors of surgical journals and better quality control measures for trial reporting and methodology.

A controlled trial of arthroscopic surgery for osteoarthritis of the knee (115) in 2002 reported a lack of benefit. However, the methodology was questioned (116-121). Further, the authors' conclusions (115) that arthroscopy is ineffective for the treatment of moderate to severe arthritis of the knee has not been generally accepted (122-125). The same issue was revisited in another randomized trial of arthroscopic surgery for osteoarthritis of the knee (126) published in 2008 and this study also concluded that arthroscopic surgery provides no additional benefit to optimized physical and medical therapy. Even though the authors conceded that bias is possible because of the lack of sham surgery control, they contend that such a bias would be expected to favor surgery and would not be expected to explain the present results. However, arthroscopic surgery continues to be widely used for osteoarthritis of the knee even though scientific evidence to support its efficacy has been lacking (127).

The biased interpretation of the results from randomized trials, either in favor or not of a treatment, and lack of proper understanding of randomized trials, lead to a poor appraisal of the quality of clinical trials (108,113,128-140). It has been shown that the reporting of RCTs is often incomplete (113,128-130), compounding problems arising from poor methodology (108,109,131-136).

Apart from criteria developed for the reporting of randomized trials by means of the CONSORT statement (108,109), numerous criteria have been developed to assess the quality of randomized trials (103). These guidelines assess the methodologic quality of each trial. The CONSORT statement for reporting randomized trials (108) provides a checklist of items to include when reporting a randomized trial as shown in Table 1. The extension of the CONSORT statement for reporting of non-inferiority and equivalence randomized trials (109) expands the items on the CONSORT checklist. Of the 22 items on the CONSORT checklist, 11 required expansion (108,109). Systems to rate the strength of scientific evidence, a comprehensive document developed by the Agency for Healthcare Research and Quality (AHRQ) by West et al (103), evaluated numerous systems concerned with RCTs including 20 scales, 11 checklists, one component evaluation, and 7 guidance documents, along with review of 10 rating systems used by AHRQ's Evidence-Based Practice Centers. Subsequently, they designed a set of high-performing scales or checklists pertaining to RCTs by assessing their coverage of the 7 domains which included study question, study population, randomization, blinding, interventions, outcomes, statistical analysis, results, discussion, and funding or sponsorship. They concluded that 8 systems for RCTs represent acceptable approaches that could be used today without major modifications (140-147). Ten rating systems used by AHRQ's Evidence-Based Practice Centers are also considered critically developed and reviewed (148-158). Yet, the researchers tend to use modified systems to meet their needs or use outdated systems (20-22,25,26,29-39).

5. What Are Placebo-controlled Trials Versus Pragmatic Trials?

Controlled clinical trials of health care interventions are either explanatory or pragmatic. Explanatory trials test whether an intervention is efficacious; that is whether it can have a beneficial effect in an ideal situation. In contrast, pragmatic trials measure effectiveness; that is they measure the degree of beneficial effect in

Table 1. *Checklist of items to include when reporting a randomized trial.*

Paper Section and Topic	Item Number	Descriptor
Title and abstract	1	How participants were allocated to interventions (e.g., "random allocation," "randomized," or "randomly assigned").
Introduction		
Background	2	Scientific background and explanation of rationale.
Methods		
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.
Objectives	5	Specific objectives and hypotheses.
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.
Randomization Sequence generation	8	Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification).
Allocation concealment	9	Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.
Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. If done, how the success of blinding was evaluated.
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); methods for additional analyses, such as subgroup analyses and adjusted analyses.
Results		
Participant flow	13	Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.
Recruitment	14	Dates defining the periods of recruitment and follow-up.
Baseline data	15	Baseline demographic and clinical characteristics of each group.
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention to treat." State the results in absolute numbers when feasible (e.g., 10 of 20, not 50%).
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group and the estimated effect size and its precision (e.g., 95% confidence interval).
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those prespecified and those exploratory.
Adverse events	19	All important adverse events or side effects in each intervention group.
Discussion		
Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with multiplicity of analyses and outcomes.
Generalizability	21	Generalizability (external validity) of the trial findings.
Overall evidence	22	General interpretation of the results in the context of current evidence.

Source: Altman DG et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001; 134:663-694 (108).

real clinical practice. The explanatory trial seeks to maximize the internal validity by issuing rigorous control of all variables other than the intervention, whereas the pragmatic trial seeks to maximize external validity to ensure that the results can be generalized. There are limitations for both types of trials. Surprisingly, methodologic quality criteria awards the same number of points whether it is a pragmatic trial or a placebo-controlled trial as shown in Table 2 (28,31,159,160).

In modern medicine, pragmatic or practical clinical trials measuring effectiveness are considered more appropriate than explanatory trials measuring efficacy (10,12-14,67-72,161-165). Explanatory trials are most commonly conducted in academic settings measuring the efficacy, whereas pragmatic or practical trials are best designed to provide the results of benefit of the treatment produced in routine clinical practice (10,12-14,163-165). In addition, practical clinical trials address the questions about the risks, benefits, and costs of an intervention as they occur in routine clinical practice better than an explanatory trial in an academic setting (164). The issue of lack of a placebo group is addressed in pragmatic trials with the treatment response accounting for the total difference between 2 treatments, including both treatment and associated placebo effect. Consequently, the treatment response in a pragmatic trial is a combination of the treatment effect and placebo effect, as this will best reflect the likely clinical response in actual clinical practice.

Multiple pragmatic trials have been conducted in interventional pain management settings (70-72,161,162).

Vesely and De Almeida (166) described that much of the controversy surrounding EBM is due to the fact that EBM has been reduced to mainly RCTs or gold standard trials. Multiple shortcomings of RCTs (165) include 1) the length of time to complete due to the difficulty of obtaining large sample sizes to achieve statistical significance (167) in order to equalize confounding factors between study groups by randomization (168) and due to a very low incidence of disease under study, requiring cooperation between multiple groups and settings; 2) possibly never addressing a clinician's question; 3) the proposal of a lengthy and costly RCT for a therapy that may never achieve popularity or may soon become obsolete (169); and 4) it may not lend itself to particular types of research questions, which may nevertheless be very important (170,171). Consequently, clinical researchers focus on topics where the methodologic criteria of reviewers and editors can be most easily met, rather than studying real-life clinical problems (172).

Table 2. *Modified and weighted Cochrane methodologic quality assessment criteria as described by Koes et al (28).*

CRITERION		Weighted Score
1. Study population		35
A	Homogeneity	2
B	Comparability of relevant baseline characteristics	5
C	Randomization procedure adequate	4
D	Drop-outs described for each study group separately	3
E	< 20% loss for follow-up	2
	< 10% loss for follow-up	2
F	> 50 subject in the smallest group	8
	> 100 subjects in the smallest group	9
2. Interventions		25
G	Interventions included in protocol and described	10
H	Pragmatic study	5
I	Co-interventions avoided	5
J	Placebo-controlled	5
3. Effect		30
K	Patients blinded	5
L	Outcome measures relevant	10
M	Blinded outcome assessments	10
N	Follow-up period adequate	5
4. Data-presentation and analysis		10
O	Intention-to-treat analysis	5
P	Frequencies of most important outcomes presented for each treatment group	5
TOTAL SCORE		100

Errors in the design of large scale RCTs can compromise, if not completely invalidate, the conclusions generated by such a study (173-177). Even a study with impeccable experimental design and implementation may be questioned as to the generalizability of the results, since most studies have very particular eligibility criteria upon which treatment groups are chosen, and the conclusions may only be applicable to others that strictly fit such criteria (178). Another disadvantage of placebo-controlled

RCTs is patients are chosen from tertiary or quaternary care settings and usually differ significantly from those for which the treatment is intended, not only in physiologic commonality (176,179,180), but also with concomitant medical and psychological profiles. In fact, Charlton (180) argues that the results of an RCT offer knowledge only at the population level in megatrials, and to apply this to the care of individual patients would be a classic example of the ecological fallacy. Further, RCTs rarely actually study long-term outcomes, and therefore many RCTs look instead at surrogate endpoints (166,181). This essentially translates that placebo-controlled RCTs fail to apply to the general patient population, fail to investigate mechanisms, and fail to accurately look at outcomes (existence and size of effect on a long-term basis).

Due to the disadvantages of placebo-controlled trials, physicians and other medical decision-makers should choose practical clinical trials to obtain high quality evidence-based on head-to-head comparisons of clinically relevant alternatives. Characteristic features of practical clinical trials are 1) interventions compared in the trial are clinically relevant alternatives, 2) participants are diverse and reflect the underlying population affected with the disease, 3) participants come from a heterogenous group of practice settings and geographic locations, and 4) endpoints of the trial reflect a broad range of meaningful clinical outcomes (182). However, in spite of the inherent usefulness of practical clinical trials, they have been relatively rare compared to placebo-controlled trials. Multiple reasons described for this disparity include the primary mission of major funding sources for clinical trials being RCTs rather than practical clinical trials (163,183,184). MacPherson (185) described in detail practical clinical trials, along with the differences between explanatory and pragmatic trials, as illustrated in Table 3. Multiple practical clinical trials have been conducted in various disciplines, including interventional pain management (65,67-72,161,162,186-189). However, a great need continues for more practical clinical trials in interventional pain management. This is exemplified by the fact that practical clinical trials are crucial for developing practice guidelines and quality indicators, and for formulating evidence-based coverage policies for public and private insurers. This is not only true but essential, as large numbers of traditional, placebo-controlled trials continually expand the range of high-cost new technologies that are available to physicians and patients

Table 3. *Characteristics of explanatory (placebo-control) and pragmatic (active-control) trials.*

EXPLANATORY TRIALS	PRAGMATIC TRIALS
1. Placebo-controlled	Not placebo-controlled
2. Experimental setting	Routine care setting
3. Evaluate efficacy	Compare effectiveness
4. More suitable for acute conditions	More suitable for chronic conditions
5. Standardized treatment	Routine treatment
6. Simple interventions	Complex interventions
7. Practitioner skilled for standard protocol	Practitioner skilled in routine care
8. Patients blinded to minimize bias	Patients unblinded to maximize synergy
9. Aim to equalize non-specific effects	Aim to optimize non-specific effects
10. Usually short-term follow-up	Often long-term follow-up
11. May manage with smaller sample sizes	May need larger sample sizes
12. Low relevance and impact on practice	High relevance and impact on practice
13. Homogenous group of patients	Heterogeneous group of patients
14. More commonly used	Less commonly used
15. Provide comparative information of interventions	Do not provide comparative information of interventions
16. Minimal ethical concerns	Major ethical concerns
17. IRB approval difficult	IRB approval relatively easier
18. High internal validity	High external validity
19. Generally large withdrawals	Generally fewer withdrawals
20. Disincentive for physicians and patients with lack of preference	Enhanced preferences and incentives for patients and physicians

Adapted and modified from MacPherson H. Pragmatic clinical trials. *Complement Ther Med* 2004; 12:136-140 (185).

rather than evaluating the existing technologies. Finally, most disorders treated with interventional techniques are chronic in nature, thus it is much easier and more practical for practical clinical trial design rather than placebo-controlled trials, removing the disparity between the course of chronic pain and the length of the trials, an obvious weakness in design in placebo-controlled randomized trials.

6. What Is A Control Group Design?

The choice of control group is always a critical decision in designing a clinical trial, as the choice affects the inferences that can be drawn from the trial, the ethical acceptability of the trial, the degree to which bias in conducting and analyzing the study can be minimized, the types of subjects that can be recruited and the pace of recruitment, the kind of endpoints that can studied, the public and scientific credibility of the results, the acceptability of the results by regulatory authorities, and many other features of the study, its conduct, and its interpretation (56). The control group experience essentially provides the knowledge of what would have happened to patients if they had not received the test treatment or if they had received a different treatment known to be effective.

Table 4 and Figure 1 illustrate the methodology to choose the control group for demonstration of efficacy and usefulness of specific control types in various situations. As shown in Table 4, the best type of trial is the placebo + active + dose-response type of trial.

7. What Are The Types Of Controls?

Control groups in clinical trials can be classified on the basis of either the type of treatment used or the method of determining who will be in the control group. Four types of control groups have been described. These include: 1) placebo, 2) active treatment, 3) no treatment, 4) different dose or regimen of the study treatment, and 5) external or historical control. The first 4 are concurrently controlled, usually with random assignment to treatment, distinguished by the type of control as described above; whereas, external or historical control groups, regardless of the comparator treatment, are considered as different and probably inferior due to the serious concerns about the ability of such trials to ensure comparability of tests and control groups and their ability to minimize important biases, making the external or historical control design usable only in unusual circumstances. However, more than one type of control group may be utilized.

7.1 Placebo-Control

In a placebo-controlled trial, subjects are randomly assigned to a test treatment or to an identical-appearing treatment that does not contain the test drug (56). Such trials are always double-blind and the treatments may be titrated to effect or tolerance or may be given at one or more fixed doses. As suggested by the name itself, the purpose is to control for placebo effect. However, placebo control design also controls

Table 4. *Usefulness of specific control types in various situations.*

Trial Objective	Type of Control						
	Placebo Control	Active Control	Dose Response (D/R)	Placebo + Active	Placebo + D/R	Active + D/R	Placebo + Active + D/R
Measure *Absolute* effect size	Y	N	N	Y	Y	N	Y
Show existence of effect	Y	Y	Y	Y	Y	Y	Y
Show dose-response relationship	N	N	Y	N	Y	Y	Y
Compare therapies	N	Y	N	Y	N	P	Y

Y=Yes, N=No, P=Possible, depending on whether there is historical evidence of sensitivity to drug effects.
Source: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Choice of Control Group and Related Issues in Clinical Trials E10. July 20, 2000 (56).

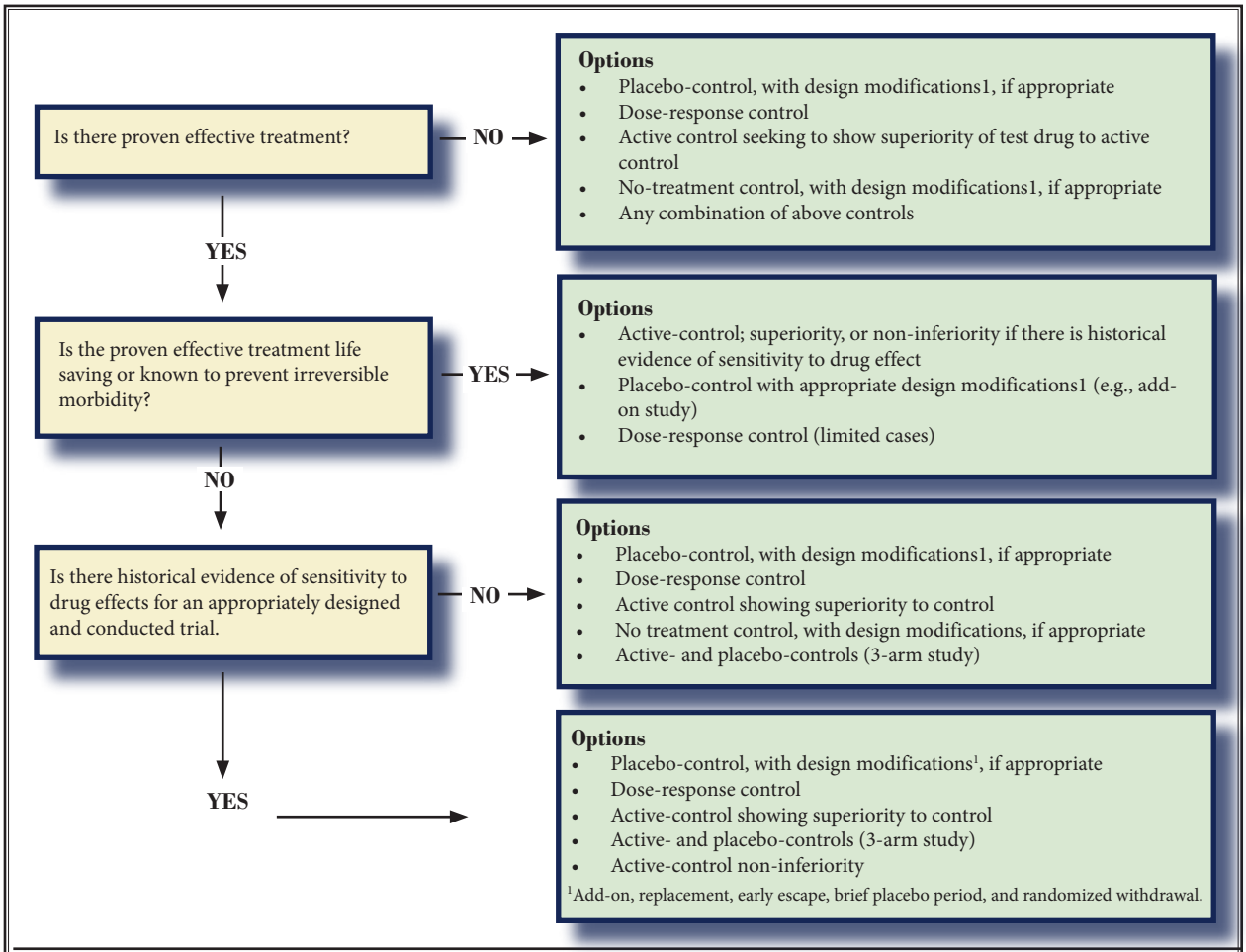

Fig. 1. *Choosing the concurrent control for demonstrating efficacy.*
Source: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Choice of Control Group and Related Issues in Clinical Trials E10. July 20, 2000 (56).

for all potential influences on the actual or apparent course of the disease other than those arising from the pharmacologic action of the test drug. In general, placebo-controlled trials seek to show a difference between treatments when they are studying effectiveness, but may also seek to show a lack of differences of specified size, in evaluating a safety measurement.

A placebo is a "dummy" treatment that appears as identical as possible to the test treatment with respect to physical characteristics such as color, weight, taste, and smell, or needle placement and injection of a solution in interventional pain management, but the treatment does not contain the same drug or procedure. There is a wide variety of designs that can be used successfully in placebo control – namely parallel or crossover designs, single fixed dose or titration in the active drug group, or several fixed doses. However, all placebos are not completely inactive. For example, some studies of interventional techniques described placebo controls in which sodium chloride solution was injected over a nerve root, into a joint, or into an epidural space or local anesthetic or other agents were injected exerting a significant effect (63-80). This may or may not impair the ability of the design to measure the specific effect of the test agent or procedure.

7.1.1 Minimizing the Bias

The placebo-controlled trial, using randomization and blinding, generally minimizes subject and investi-

gator bias. Such trials, however, are not impervious to blind-breaking through recognition of the pharmacologic effects of one treatment.

7.1.2 Ethical Issues

When a new treatment is tested for a condition for which no effect is known, there is usually no ethical problem with a study comparing the new treatment to placebo. However, when an effective treatment is available for various painful conditions with interventional techniques, use of a placebo control may raise problems of ethics, acceptability, and feasibility. A long-lasting placebo-controlled trial of 1 to 2 years or longer in an interventional pain management setting is not the same as a short-term placebo-controlled trial of a new antihypertensive agent or antidepressant agent in managing mild essential hypertension or depression with no end-organ disease, which may be considered generally acceptable.

7.1.3 Advantages

The major advantage of a randomized double-blind controlled trial is that when used to show effectiveness of a treatment, it is free of assumptions and reliance on external information. Placebo-controlled trials may also provide the maximum ability to distinguish adverse effects caused by a drug or procedure from those resulting from underlying disease or intercurrent illness. Further, a placebo-controlled trial contains internal evidence of assay sensitivity, lending to interpretation without reference to external findings when a difference is demonstrated.

Disadvantages of placebo-controlled trials include ethical concerns, patient and physician practical concerns, lack of generalizability, and lack of comparative information. When an effective therapy is known to prevent pain in a particular population, that population usually cannot be ethically studied in placebo-controlled trials. Thus, for interventional techniques, active control designs are the most ideal.

Physicians and/or patients may be reluctant to accept the possibility that the patient will be assigned to the placebo treatment, even if there is general agreement that withholding or delaying treatment will not result in harm. Further, it is unlikely that in interventional pain management, one can withhold treatment for pain relief and improvement of function on a long-term basis. In addition, patients also sense that they are not improving and withdraw from treatment because they attribute a lack of effect to having been treated with placebo, even though they are in an active group, complicating the analysis of the study, despite the use of intent-to-treat analysis.

Generalizability is an issue with randomized trials. A placebo-controlled trial essentially represents an artificial environment that gives results different from true "real world" effectiveness. Study populations in randomized trials may be unrepresentative of the patient populations because of ethical or practical concerns. Another practical problem is that it may be extremely difficult to get such studies approved by local institutional review boards (IRBs). However, it was concluded that participation in RCTs is not associated with greater risks than receiving the same treatment outside RCTs, challenging the assertion that the results of RCTs are not applicable to usual practice (190).

Placebo-controlled trials also fail to provide comparative effectiveness, information that is of importance and interest to patients and physicians in most circumstances. Such information cannot reliably be obtained from cross-study comparisons, as the conditions of the studies may have been quite different.

Finally, preferences, incentives, and disincentives to participation by clinicians and patients in RCTs have been described (191,192). King et al (191) in a systematic review of the effects of participants' and professionals' preferences in RCTs concluded that preferences significantly compromise the internal and external validity of trials. Rendell et al (192) concluded that the impact of factors varied across the studies. Thus, researchers need to be aware that aspects of the design and conduct of trials can affect clinicians' willingness to invite patients to participate.

7.1.4 Other Design Considerations

Ethical or practical limitations of placebo-controlled trials may be addressed by using modified study designs that still retain the inferential advantages of these trials. Placebo-controlled trials can also be made more informative by including additional treatment groups (Table 4 and Fig. 1).

7.2 Active (positive) Control

This type of trial is also considered as a pragmatic or practical clinical trial (10,12-14,163-165). In this design of active or positive controlled trial, patients are randomly assigned to the test treatment or to an active control treatment. Such trials are usually double-blind, but this is not always possible in certain circumstances. Active control trials can have 2 distinct objectives with

respect to showing efficacy, which include the demonstration of efficacy of the test treatment by showing it as good as a known effective treatment, or to show efficacy by showing superiority of the test treatment to the active control. They may also be used with the primary objective of comparing the efficacy and/or safety of the 2 treatments. Whether the purpose of the trial is to show efficacy of the new treatment or to compare 2 treatments, the question of whether the trial would be capable of distinguishing effective from less effective or ineffective treatments is critical.

A trial using any of the control types may demonstrate efficacy of the test treatment by showing that it is superior to the control. Further, an active control trial may, in addition, demonstrate efficacy in some cases by showing the new treatment to be similar in efficacy to a known effective treatment, thus, establishing the efficacy of the test treatment. However, this is only feasible when the active control treatment is effective under the conditions of the trial, as 2 treatments would also look similar if neither were effective in the trial.

7.2.1 Minimizing the Bias

A randomized, blinded, active control trial generally minimizes subject and investigator bias, but investigators and patients know that all patients are getting an active drug or intervention, although they do not know which one. Consequently, this could lead to a tendency toward categorizing borderline cases as successes in partially subjective evaluations.

7.2.2 Ethical Issues

Active controls are generally considered to pose fewer ethical and practical problems than placebo-controlled trials. Consequently, it should be appreciated that subjects receiving a new treatment are not receiving standard therapy (just as a placebo control group is not) and may be receiving an ineffective or even harmful drug or intervention.

7.2.3 Advantages

When a new treatment shows an advantage over an active control, the study is readily interpreted as showing efficacy, just as any other superiority trial is, assuming that the active control is not actually harmful. However, an active control may also be used to assess the comparative efficacy.

The major advantages of the active control design include ethical and practical advantages and larger samples. The active control design, whether intended to show non-inferiority, equivalence, or superiority, reduces ethical concerns that arise from failure to use drugs or interventions with documented important health benefits. Further, it addresses physician and patient concerns about failure to use documented effective therapy. In essence, this will facilitate IRB approval and also recruitment, leading to a study of larger samples. Further, there may be fewer withdrawals due to lack of effectiveness.

The disadvantages of active control trials include difficulty in quantitating safety outcomes, lack of direct assessment of effect size, and the requirement for large sample sizes.

7.4 No-treatment Control

In a no treatment-controlled trial, patients are randomly assigned to test treatment or to no study treatment. This design principally differs from the placebo control design in that subjects and investigators are not blind to the treatment assignment. Consequently, this design is useful only when it is difficult or impossible to double-blind. Further, the study designers must have reasonable confidence that study endpoints are objective and that the results of the trial are unlikely to be influenced by various factors intended to minimize the potential biases resulting from differences in management, treatment, or assessment of patients, or interpretation of results that could arise from subject or investigator knowledge of the assigned treatment. Some of these effects may be obtunded by a blinded evaluator. Many of the interventions both in interventional pain management and surgery fall into this category (54,70-72, 126,161,162,193-203), which generally use usual treatment and compare it with an active intervention.

7.5 Dose-response Control

In the dose-response control design, subjects are randomized to one of the several fixed-dose groups. The intended comparison of this design is between the groups on their final dose. Consequently, subjects may be placed on either fixed dose initially or be raised to that dose gradually. Dose-response trials are usually double-blind. They may include a placebo (zero dose) and/or an active control. In this type of a trial, treatment groups are titrated to several fixed-concentration windows and this trial is conceptually similar to a fixed-dose, dose-response trial. In a dose-response controlled trial, patients are randomized to 2 or more

doses of the study drug (e.g., 3 mg, 6 mg, or 12 mg of epidural betamethasone).

7.5.1 Minimizing the Bias

In a dose response control, if the study is blinded, bias is minimized similar to randomized and blinded designs. Consequently, when a drug or a procedure has effects that could break the blinding for some patients or investigators, it may be easier to preserve blinding in a dose-response study than in a placebo control trial.

7.5.2 Ethical Issues

The ethical and practical concerns related to a dose-response study are similar to those affecting placebo control trials. Consequently, when there is therapy known to be effective in preventing morbidity, it is no more ethically acceptable to randomize deliberately to subaffective control therapy than it is to randomize to placebo. However, in interventional pain management or surgery settings, this design may be acceptable.

7.5.3 Advantages

A blinded dose-response study is useful for the determination of efficacy and safety in situations where a placebo-controlled trial would be useful and has similar credibility. The other advantages include efficiency and possible ethical advantage. In this design, even though a comparison of a large, fully effective dose to placebo may be maximally efficient for showing efficacy, this design may produce unacceptable toxicity and gives no dose-response information. Thus, when the dose response is monotonic, the dose-response trial is reasonably efficient in showing efficacy and also yields dose-response information. Further, it may be more prudent to study a range of doses than to choose a single dose that may prove to be suboptimal or to have unacceptable adverse effects if the optimally effective dose is not known. Further, possible ethical advantages include the dose-related efficacy and dose-related important toxicity compared to a placebo-controlled trial.

The disadvantages of dose-response studies include the necessity to recognize that a positive dose-response trend without significant pair-wise differences may leave uncertainty as to which doses, other than the largest, are actually effective. In addition, no differences between doses in a dose-response study may be observed if there is no placebo group, which may lead to an uninformative outcome.

7.6 External Control

An externally controlled trial is one in which the control group consists of patients who are not part of the same randomized study as the group receiving the investigational agent or technique. Consequently, there is no randomized control group. The control group is thus derived from a different population than the treated population, such as a historical control. Even though the control group is a well-documented population of patients observed at an earlier time (historical control), it could be a group at another institution observed contemporaneously, or even a group at the same institution but outside the study. Baseline-controlled studies describe the patient as his or her own control, and they do not in fact have an internal control.

7.6.1 Minimizing the Bias

Inability to control bias is the major and well-recognized limitation of externally controlled trials and is sufficient in many cases to make the design unsuitable. It is difficult, and in some cases impossible, to establish comparability of the treatment and control group and thus to fulfill the major purpose of a control group.

7.6.2 Advantages

The main advantage of an externally controlled trial is that all patients can receive a promising drug or therapy, making the study more attractive to patients and physicians. Other advantages include that externally controlled trials are most likely to be persuasive when the study endpoint is objective, when the covariates influencing outcomes of the diseases are well characterized, and when the control closely resembles the study group in all known relevant baseline, treatment, other than the study procedure drug, and observational variables.

Disadvantages of externally controlled study groups are lack of blinding and observer and analyst bias, resulting in erroneous conclusions, despite all precautions.

8. How to Choose Control Groups?

In most cases, evidence of efficacy is most convincingly demonstrated by showing superiority to a control treatment. If a superiority trial is not feasible or is inappropriate for ethical or practical reasons, and if a defined treatment effect of the active control is regularly seen, as it is for multiple interventional techniques, a non-inferiority or equivalence trial can be

used and can be persuasive. Consequently, choosing the control group is important. Various choices of control demonstrate efficacy, whereas some designs also allow comparisons of test and control agents. However, the choice of control can be affected by the availability of therapies and by medical practice in specific regions.

9. Types Of Trials

Randomized trials may be double-blind, single-blind, or open.

9.1 Blinding

Clinical trials are often double-blind meaning that both subjects and investigators, as well as sponsor or investigator staff involved in the treatment or clinical evaluation of subjects, are unaware of each subject's assigned treatment. The main purpose of blinding is to ensure that subjective assessments and decisions are not affected by knowledge of the treatment assignment.

9.2 Double-blind Trial

In a double-blind trial, neither researcher nor patient is aware of allocation and intervention. Therefore, double-blind randomized trials tend to give the most accurate results.

Unlike allocation concealment, blinding may not always be appropriate or possible. Blinding is particularly important when outcome measures involve some subjectivity, such as assessment of pain, improvement in functional status, or cause of death. Lack of blinding in any trial can lead to other problems, such as attrition. In certain trials, especially interventional pain management or surgical trials, double-blinding is difficult or impossible; however, blinded assessment of outcome can often be achieved even in open trials (204,205).

Blinding is not limited to the caregiver and the patient. Blinding should extend to evaluators, monitors, and data analysts (206-208). Further, the blinding mechanism, similarity of characteristics of treatments, and explanation must be provided if care providers or evaluators were not blinded. The revised CONSORT statement for reporting randomized trials (108) states that authors frequently do not report whether or not blinding was used, and when blinding is specified, details are often missing (130,208-211). Blinding in its methodologic sense appears to be understood worldwide and is acceptable for reporting clinical trials (204,212).

The description of the mechanism used for blinding may provide such an assurance of successful blinding, which can sometimes be evaluated directly by asking participants, caregivers, and outcome assessors which treatment they think they received (213). However, if participants do successfully identify their assigned intervention more often than expected by chance, it may not mean that the blinding was unsuccessful (108). Based on clinical outcome clinicians are likely to assume, though not always correctly, that a patient who had a favorable outcome was more likely to have received the active intervention rather than control. If the active intervention is indeed beneficial, their "guesses" would be likely to be better than those produced by chance (214).

Evaluators of the literature and studies may misinterpret the effectiveness of blinding. Carragee et al (37) criticized blinding in the study conducted by Lord et al (66). The Task Force reported that blinding was in doubt, as 42% of the active group developed long-term anesthetic or dysesthetic areas of skin and none of the control developed changes. They stated that these changes revealed the treatment assigned in nearly half of the active treatment group. Lord et al (66) were unable to avoid such an issue and in fact, this is a problem with any of the sham procedures in interventional pain management. Dreyfuss and Baker (98) stated that Lord et al (66) maintained blinding of subjects admirably well and the evidence of difficulty of performing such a study is demonstrated by an extremely limited number of published sham studies involving an invasive treatment. Even then, Carragee et al (215) maintained their criticism.

9.3 Single-blind Trial

In a single-blind trial, the researcher knows the details of the treatment, but the patient does not. Because the patient does not know which treatment is being administered, there might be no placebo effect. In practice, since the researcher knows, it is possible for him to treat the patient differently or to subconsciously hint to the patient important treatment-related details, thus influencing the outcome of the study.

9.4 Open Trial

In an open trial, the researcher and the patient know the full details of the treatment.

10. Randomization In Clinical Trials

Randomization procedure and allocation concealment are 2 processes involved in randomizing patients to different interventions. Randomization procedure refers to generating a random and unpredictable sequence of allocations which may be a simple random assignment of patients to any of the groups at equal probabilities or it may be a complex design (Fig. 2) (96).

10.1 Randomization Procedures

Multiple issues to consider in generating the randomization sequences (216) include balance, selection bias, and accidental bias. Balance refers to the fact that it is desirable for the randomization procedure to generate similarly-sized groups, since most statistical tests are most powerful when the groups being compared have equal sizes. Selection bias refers to the fact that investigators may be able to infer the next group assignment by guessing which of the groups has been assigned the least up to that point, which essentially breaks allocation concealment and can lead to bias in the selection of patients for enrollment in the study. Accidental bias refers to bias generated by covariates that are related to the outcomes which are ignored in the statistical analysis. This potential magnitude of the bias, if any, will depend on the randomization procedure.

10.1.1 Random Allocation Sequence

Participants should be assigned to comparison groups in the trial on the basis of a chance random allocation (process characterized by unpredictability). The method used to assign treatments or other interventions to trial participants is a crucial aspect of the clinical trial design. Random assignment is the preferred method as it has been successfully used in trials for more than 50 years (17-19). Randomization has 3 advantages (217): first, it eliminates bias in assignment of treatments; second, random allocation facilitates blinding (218); and third, random assignment permits the use of probability theory to express the likelihood that any difference in outcome between intervention groups merely reflects chance (219). Thus, preventing selection and confounding biases is the most important advantage of randomization (220).

There are many methods of sequence generation available that are considered adequate. It has been found that only 32% of reports published in specialty journals (221) and 48% of reports in medical journals (222) specified an adequate method of randomization.

Randomization is of 2 types: simple randomization and restricted randomization.

10.1.2 Simple Randomization

Randomization based on a single sequence of random assignments is known as simple randomization (96,223). The most common and basic method of simple randomization is flipping a coin, as originally described by Fisher in 1926 (17). However, in the modern world, a random number table found in a statistics book or computer-generated random numbers can also be used for simple randomization of participants.

The disadvantages of simple randomization include an unequal distribution in each group. By chance alone, the smaller the sample size, the larger the likelihood of a major imbalance in the number of patients or the baseline characteristics in each group (224,225). This disadvantage will happen when flipping a coin or odd and even numbers are utilized. In small trials or in larger trials with planned in-term analysis, simple randomization can result in imbalanced group numbers. Even if the groups have equal numbers, there can be important differences in baseline characteristics that would distort the apparent treatment effect (41).

Simple randomization can also result in chronological bias in which one treatment is predominantly assigned earlier and the other later in trial (226). However, chronologic bias is only important with new procedures in which an investigator gains experience as time passes.

10.1.3 Restricted Randomization

Restricted randomization describes any procedure to control the randomization to achieve balance between groups in size or characteristics. This is achieved by blocking, stratification, or covariate adaptation.

10.1.3.1 Blocking

Blocking is used to ensure close balance of the numbers in each group at any time during the trial. After a block of every 10 participants was assigned, for example, 5 would be allocated to each arm of the trial and after a block of every 20 participants

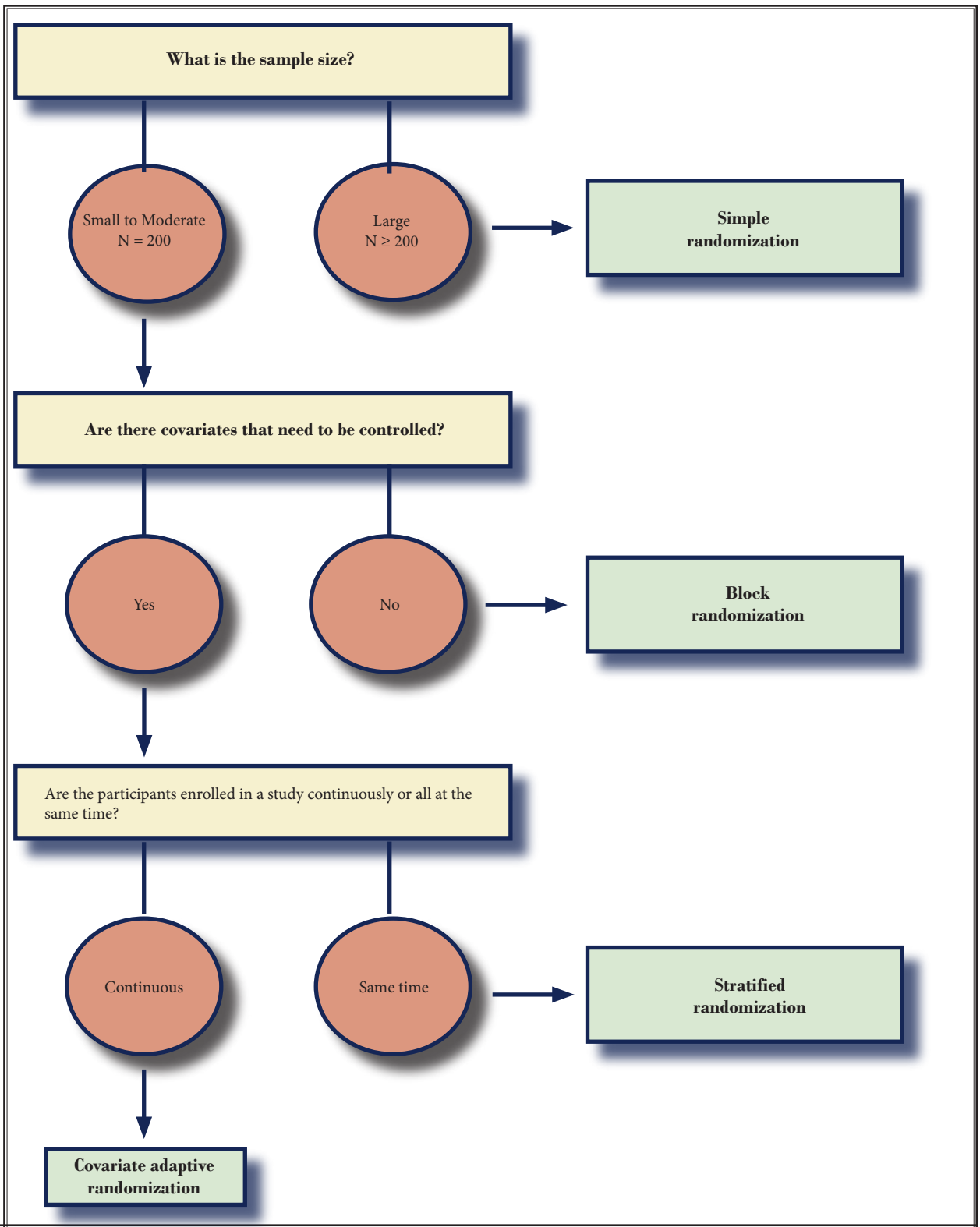

Fig. 2. *Flow chart for selecting appropriate randomization technique (the gray boxes represent appropriate techniques).*
Adapted and modified from Kang M et al. Issues in outcomes research: An overview of randomization techniques for clinical trials. J Athl Train 2008; 43:215-221 (96).

was assigned, 10 would be allocated to each arm of the trial (223). However, improved balance comes at the cost of reducing the unpredictability of the sequence. Even though the order of interventions varies randomly within each block, a person running a trial could deduce some of the next treatment allocations if they discovered the block size (224). Blinding the interventions, using larger block sizes, and randomly varying the block size can ameliorate this problem.

Even though the balance in sampling may be achieved with this method, groups may be generated that are rarely comparable in terms of certain covariates (227-229). It has been observed that one group may have more participants with secondary diseases such as comorbid medical conditions that could confound the data and may negatively influence the results of the clinical trial (96). The importance of controlling for these covariates due to the serious consequences to the interpretation of the results has been stressed (230). Hence, in small trials, sample size and covariates must be balanced, since imbalance could introduce bias in the statistical analysis and reduce the power of the study (225,229,231).

10.1.3.2 Stratification

An imbalance may weaken the trial's credibility (227) if the trial is small and study groups are not well matched per baseline characteristics, such as age and stage of the disease. Such imbalances can be avoided without sacrificing the advantages of randomization. Consequently, stratification ensures that the numbers of participants receiving each intervention are closely balanced with each stratum.

Stratified randomization is achieved by performing a separate randomization procedure within each of the 2 or more subsets of participants based on certain characteristics such as age, smoking, or disease severity. Stratification by center is common in multicenter trials and stratification requires blocking within the strata. However, without blocking, it is, ineffective.

Stratified randomization, which is a relatively simple and useful technique, specifically for smaller clinical trials, may become complicated to implement if multiple covariates must be controlled (232-238). Since stratified random allocation is carried out by blocks used to control the covariates of sex (male and female) or body mass index (3 levels – normal weight, underweight, overweight); between study arms, with these 2 covariates, possible block combinations total 6. However, if covariates are increased, they would be multiplied. Thus, too many block combinations may lead to imbalances in overall treatment allocations because a large number of blocks can generate small participant numbers within the block.

10.1.3.3 Covariate Adaptive Randomization

Covariate adaptive randomization has been recommended by many researchers as a valid alternative randomization method for clinical trials (41,96,239). When treatment assignment is based on patient characteristics, the adaptive randomization procedure, also known as minimization, assigns the next treatment to minimize any imbalance in prognostic factors among previously enrolled patients. For the computer algorithm to run, minimization should be limited to larger trials (240).

Multiple types of covariate adaptive randomizations have been described (95,230,233, 239,241-250). Even though covariate adaptive randomization produces less imbalance than other conventional randomization methods and can be used successfully to balance important covariates among control and treatment groups (229), these approaches are made controversial by losing predictability and being susceptible to bias.

10.1.4 Randomized Designs Incorporating Preferences

Randomized trials may incorporate patient and physician preferences. These designs avoid failure to enroll patients into surgical or interventional RCTs based on lack of preferences or serve as a theoretical threat to validity (41,251-256). Similarly, there could be a problem of variation between the physician skill and preference, resulting in expertise bias (226).

10.1.4.1 Patient Preference Trials

Multiple designs are available for patient preference trials (253,254,257). Patients randomized to their preferred treatment can perform better because of increased compliance or placebo effect, and patients randomized to their non-preferred treatment can perform worse (255). This issue may be resolved by measuring baseline patient preferences and mathematically adjusting for the interaction between preference and treatment, but, this approach may increase sample size (256). Other solutions include modification of

trial designs to incorporate patient preferences using multiple designs (191,253,254,257-260).

Even though patient preference trials are an alternative to RCTs, the downsides include the potential for additional differences between treatment groups other than preference and increased sample size requirements or costs to complete a trial (255,257).

10.1.4.2 Physician Preferred Trials

Physician preferred or expertise-based trials differ from a conventional RCT because physicians perform only the procedure at which they believe they are most skilled. Proponents of this technique argue that expertise-based trials minimize bias resulting from differences in technical competency and physician preference, decrease crossover from one intervention to the other, and can be more ethical than conventional RCTs (226). However, these types of RCTs present challenges in coordinating trials in which there are few experts for one or both procedures, changing physicians after the initial patient contact, or generalizing the results to physicians with less expertise (226).

10.2 Allocation Concealment

Allocation concealment is a technique used to prevent selection bias by concealing the allocation sequence from those assigning participants to intervention groups, until the moment of assignment. Allocation concealment prevents researchers from influencing which participants are assigned to a given intervention group, either unconsciously, or otherwise. In practice, in taking care of individual patients, clinical investigators may often find it difficult to maintain impartiality. Breaking allocation concealment in RCTs is much more problematic because in principle the randomization should have minimized such biases. Great care for allocation concealment must go into the clinical trial protocol and be reported in detail in the publication. Studies (261,262) have found that not only do most publications not report their concealment procedure; most of the publications that do not report also have unclear concealment procedures in the protocols. A general allocation schedule with using allocation concealment minimizes bias. The decision to accept or reject a participant should be made and an informed consent should be obtained from the participant, in ignorance of the next assignment in the sequence (263-265).

10.2.1 Methods of Allocation Concealment

Some standard methods of ensuring allocation concealment for interventional techniques include:

- Central randomization
- Pharmacy controlled
- Sequentially numbered, opaque, sealed envelopes (SNOSE)
- Sequentially numbered containers

Table 5 illustrates generation and implementation of a random sequence of treatments adapted from the CONSORT statement for reporting randomized trials (108). Kunz et al (110) in a Cochrane collaboration evaluation of randomization to protect against selection bias in health care trials concluded that on average, non-randomized trials and randomized trials with inadequate concealment of allocation tend to result in larger estimates of effect than randomized trials with adequately concealed allocation. Further, they were unable to predict the magnitude, or even the direction of possible selection biases in consequent distortions of treatment effects. However, not using concealed random allocation may also result in smaller estimates or even a reversal of the direction of

Table 5. *Generation and implementation of a random sequence of treatments.*

Generation	Implementation
Preparation of the random sequence	Enrolling participants Assessing eligibility Discussing the trial Obtaining informed consent Enrolling patient in trial
Preparation of an allocation system (such as coded bottles or envelopes), preferably designed to be concealed from the person assigning participants to groups	Ascertaining treatment assignment (such as by opening the next envelope) Administering intervention

Source: Altman DG et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001; 134:663-694 (108).

the effect (from harmful to beneficial or vice versa).

11. Outcomes

Many instruments and procedures have been developed to assess the impact of chronic pain on the quality of life (266-284). The available instruments include measures evaluating the disease-specific disability (269), general pain measures (270,279), or broader measures associated with health and illness (271,283). The disease-specific measures, Oswestry Disability Index (ODI), Roland-Morris Disability Questionnaire (RDQ), and Neck Pain Disability Index (NDI) have emerged as the most commonly recommended condition-specific outcome measures for spinal disorders (271-276).

11.1 Pain Intensity Scales and Assessment of Pain

Even though uniquely subjective and highly variable, reduction in chronic pain and intensity is frequently employed as a primary outcome in RCTs. Assessment of pain intensity and reduction with interventions is considered to have both face validity and intuitive appeal (284) even though pain measurement scales and their properties, what constitutes clinically meaningful change, how responders are defined, and the manner in which pain is assessed have potentially problematic aspects to consider. Thus, pain measurements are challenged for numerous reasons – subjectivity, underlying etiology, external practice influencing pain perception, variable measurement tools, and understanding what constitutes a clinically meaningful pain reduction. Because of pain's subjective and multidimensional nature, interpreting the results of RCTs with chronic pain outcomes presents challenges. Since pain is perceived differently (285), heterogeneous responses to interventions must be considered. While there are numerous aspects to the choice and application of rating scales, 2 widely used scales in RCTs include the visual analog scale (VAS) and numeric rating scale (NRS) (286).

11.1.1 Visual Analog vs. Numeric Rating Scales

The experience of chronic pain is multidimensional with emotional, physical, and functional aspects other than sensory phenomenon (287-289). Gracely and Kwilosz (287) devised a 20-segment scale having descriptive labels with a 3-unit change reflecting a 50% pain intensity reduction. Similarly, Price et al (290) proposed that the metric of experimentally induced pain intensity followed some power function which was non-linear.

One systematic review of the literature (284) found a high degree of methodological heterogenicity that precludes comparisons across studies for either NRS or VAS (291-307). There were 4 studies evaluating low back pain with NRS (291,294-296) and 3 using VAS (291,302,303). It was opined that the body of evidence does not permit conclusions about the magnitude of change in either NRS or VAS pain scales that is clinically significant among chronic pain patients.

VAS has been used for more than 80 years (308), however, approximately 10% of individuals may have difficulty completing a VAS to assess pain intensity (309,310). VAS scales have been described on a scale of 0-100 mm or 0-10 cm. VAS scores tend to be better suited for parametric analyses (311).

NRS assesses pain intensity or other attributes, using a 0 to 10, 0 to 20, or even 0 to 100 point scale (286,291,294-296). Zero is referenced as no pain and the extreme by descriptors similar to those employed in a VAS. NRS is used more widely than VAS. The ratio properties were utilized in many clinical studies (66,70-79,312-338).

11.1.2 Clinical Versus Statistical Significance

It is important to acknowledge that statistical significance and clinical significance are not necessarily equivalent. A treatment may produce a statistically significant change, and yet be clinically meaningless (339).

The reliability of change index (RCI) is used to calculate the difference between pre- and post-treatment scores and then to divide this difference by a standard error of measure that includes not only the standard deviation of the measure, but also its reliability coefficient (339). The RCI values that result can be referenced to the normal distribution, and values that exceed 1.96 are unlikely ($p < 0.05$) unless an actual change in scores occurs between pre- and post-treatment assessment. When the absolute value of RCI exceeds 1.96, the improvement is considered to have reliably occurred at the alpha level of 0.05. That is, changes that exceed this magnitude can be considered to be reflecting more than the normal measurement of fluctuations that occur with repeated testing with a measure that has less-than-perfect reliability. Hence, it is deemed to be clinically significant.

Farrar et al (340) evaluated differences in pain scores by multiple methodologies including: absolute pain intensity difference (PID) (0-10 scale), percent-

age pain intensity difference (PID%) (0-100% scale), pain relief (PR) – 0 (none), 1 (slight), 2 (moderate), 3 (lots), 4 (complete), sum of the pain intensity difference (SPID) (over 60 min), percentage of maximum total pain relief (% max TOTPAR) (over 60 min), and global medication performance 0 (poor), 1 (fair), 2 (good), 3 (very good), 4 (excellent)). The best cut-off point for both the % Max TOTPAR and the PID% was 33%. The best cut-off points for the absolute scales were absolute pain intensity difference of 2, pain relief of 2 (moderate), and SPID of 2 (340).

The responsiveness of the NRS in a broad population of patients with various musculoskeletal conditions has been investigated and the (minimal clinically important difference [MCID]) has been identified to be 2 points (292). Additionally, in a patient population with low back pain, the scales also showed an MCID of 2 points (294).

Two of the statistical methods are the effect size (ES) statistic and the RCI.

The ES statistic is a method whereby mean differences between pretreatment and post-treatment scores can be standardized to quantify an intervention's effect in units of standard deviation (SD). It is therefore independent of measuring units and can be used to compare outcomes (341). ES statistics are widely used to assess the magnitude of treatment- related changes over time and can be applied both to group data and to data recorded from a single patient (342).

The RCI is similar to the ES statistic in that it calculates mean differences between pretreatment and posttreatment scores but divides the difference by a standard error of measure that includes not only the SD of the measure but also its reliability coefficient. RCI values can be referenced to the normal distribution, and values that exceed 1.96 are unlikely ($P \leq .05$) unless an actual and reliable change has occurred (339).

To assess a patient's own impressions of change, the Patient's Global Impression of Change (PGIC), a tool which was a global scale from "much better" through "no change" to "much worse" is commonly used (Table 6) (292,341,343). Since patients themselves make a subjective judgment about the meaning of the change to them following treatment, this scale is often taken as the external criterion or "gold standard" of clinically important change (292).

11.1.3 General Pain Measures

The Multidimensional Pain Inventory (MPI) and the McGill Pain Questionnaire (MPQ) are considered the most commonly used pain-specific assessment instruments (270,279,344-348). Both instruments have been validated and widely used.

11.2 Quality of Life

Chronic pain may affect physical and emotional functioning directly as well as indirectly, affecting the quality of life adversely. Physical functioning is not commonly reported in randomized trials, whereas pain intensity is invariably reported (349). However, the importance of quality of life improvement can never be underestimated (350).

Quality of life and functional status improvement is evaluated by many available tests. Most commonly utilized measures include Short-Form-36 (SF-36) (283), ODI (272), RDQ (274), and the NDI (275).

Table 6. *Patient's Global Impression of Change (PGIC) scale*

Since beginning treatment at this clinic, how would you describe the change (if any) in ACTIVITY LIMITATIONS, SYMPTOMS, EMOTIONS and OVERALL QUALITY OF LIFE, related to your painful conditions (check ONE box).		
1	☐	No change (or condition has got worse)
2	☐	Almost the same, hardly an change at all
3	☐	A little better, but no noticeable change
4	☐	Somewhat better, but the change has not made any real difference
5	☐	Moderately better, and a slight but noticeable change
6	☐	Better, and a definite improvement that has made a real and worthwhile difference
7	☐	A great deal better, and a considerable improvement that has made all the difference

Source: Modified from Grotle M et al (291). Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. *Spine* 2004; 29:E492-E501.

While all these instruments are considered as objective evaluations, they all depend on subjective information.

11.2.1 Short Form-36

The SF-36 is a multipurpose, short-form health survey with 36 questions. It yields an 8-scale profile of functional health and well-being scores, as well as psychometrically-based physical and mental health summary measures and a preference-based health utility index (283). It is a generic measure, as opposed to one that targets a specific age, disease, or treatment group.

Studies of validity from many types of research have yielded content, concurrent, criterion, construct, and predictive evidence of validity. The SF-36 has been shown to be sensitive to change (351,352) and able to differentiate between treatment responders and non-responders (353,354). It has been used as a validation tool in the development of new disease-specific instruments (355,356), including a pain-specific tool (356,357).

11.2.2 Oswestry Disability Index (ODI)

The ODI is the most commonly used condition specific outcome measure for assessment of low back pain with its development in 1976 and publication of the questionnaire in 1980 (269) with wide dissemination since 1981 (272). In 2000, Fairbank and Pynsent (272) reviewed the role of ODI 20 years after the introduction. During these years, numerous versions have been published, along with advances in understanding of the instrument. ODI 2.0 is the commonly utilized version in the United States in interventional pain management settings. Table 7 illustrates questions and the scoring of ODI 2.0.

11.2.2.1 Scoring

The standard scoring method, as shown in Table 8 can be used in most circumstances.

11.2.2.2 Validity and Reliability

ODI has been evaluated in normal populations and multiple other conditions including spondylolisthesis, primary back pain, psychiatric patients, idiopathic scoliosis, neurogenic claudication, chronic back pain, sciatica, neck pain, and other conditions, with severe and minor symptoms. The studies have shown face and content validity, test-retest stability, and internal consistency with validation by comparison with other tests. The ODI has been shown to have a moderate correlation with pain measures such as MPQ (279,358) and the VAS (359). It has been used to validate the Pain Disability Index (PDI) (359-361), the Low Back Outcome Score (LBOS) (362), functional capacity evaluations (363), SF-36 (364) as a predictor of return to work (365,366), and it has been evaluated as a predictor of isokinetic performance (367), isometric endurance (368), pain with sitting and standing (but not lifting) (369), prediction for centralization by the McKenzie system of evaluation (370), and multiple physical tests (371), except range of movement (372). The ODI has a linear correlation with disability, and thus, a person with a score of 40 is twice as disabled as one with a score of 20. Consequently, it is assumed that disability can be viewed as a continuum from "non-disabled" "to severely disabled." Thus, the change has been expressed as a percent of the original score, arguing that it is better to shift a patient from 20% to 10% than to go from 60% to 50% (373). Another approach in reporting is to aggregate the index into several categories, originally, 5 levels of the score were suggested (0% - 20%, 21% - 40%, 41% - 60%, 61% - 80%, 81% - 100%). Some investigators have used this system to categorize their patients (374-376). However, others have divided their patient population into 2 groups above and below a criterion, such as 40% (70-72,377).

11.2.2.3 Clinically Significant Change

The U.S. Food and Drug Administration (FDA) have chosen a minimum 15-point change in patients who undergo spinal fusion before surgery and at follow-up. Others have described a change of 4 points as the minimum difference in mean scores between the groups that carried clinical significance.

The change in the total score and change in components of the ODI have been investigated. Sources of error include inconsistencies in the answering of a questionnaire, the natural fluctuations of symptoms, as well as clinical improvements.

The ODI has been directly compared with RDQ in several studies. The 2 scales correlate. However, the ODI tends to score higher than RDQ; thus, it is likely that the ODI is better at detecting change in the more seriously disabled patients, whereas RDQ may well have an advantage in patients with minor disability (272).

11.2.3 Roland-Morris Disability Questionnaire (RDQ)

Table 7. Questions and the scoring of Oswestry Disability Index (ODI) 2.0.

Section 1: Pain Intensity	**Score**
☐ I have no pain at the moment	0
☐ The pain is very mild at the moment	1
☐ The pain is moderate at the moment	2
☐ The pain is fairly severe at the moment	3
☐ The pain is very severe at the moment	4
☐ The pain is the worst imaginable at the moment	5

Section 2: Personal Care (eg. washing, dressing)	**Score**
☐ I can look after myself normally without causing extra pain	0
☐ I can look after myself normally but it causes extra pain	1
☐ It is painful to look after myself and I am slow and careful	2
☐ I need some help but can manage most of my personal care	3
☐ I need help every day in most aspects of self-care	4
☐ I do not get dressed, wash with difficulty and stay in bed	5

Section 3: Lifting	**Score**
☐ I can lift heavy weights without extra pain	0
☐ I can lift heavy weights but it gives me extra pain	1
☐ Pain prevents me lifting heavy weights off the floor but I can manage if they are conveniently placed eg. on a table	2
☐ Pain prevents me lifting heavy weights but I can manage light to medium weights if they are conveniently positioned	3
☐ I can only lift very light weights	4
☐ I cannot lift or carry anything	5

Section 4: Walking	**Score**
☐ Pain does not prevent me walking any distance	0
☐ Pain prevents me from walking more than 2 kilometres	1
☐ Pain prevents me from walking more than 1 kilometre	2
☐ Pain prevents me from walking more than 500 metres	3
☐ I can only walk using a stick or crutches	4
☐ I am in bed most of the time	5

Section 5: Sitting	**Score**
☐ I can sit in any chair as long as I like	0
☐ I can only sit in my favourite chair as long as I like	1
☐ Pain prevents me sitting more than one hour	2
☐ Pain prevents me from sitting more than 30 minutes	3
☐ Pain prevents me from sitting more than 10 minutes	4
☐ Pain prevents me from sitting at all	5

Section 6: Standing	**Score**
☐ I can stand as long as I want without extra pain	0
☐ I can stand as long as I want but it gives me extra pain	1
☐ Pain prevents me from standing for more than 1 hour	2
☐ Pain prevents me from standing for more than 30 minutes	3
☐ Pain prevents me from standing for more than 10 minutes	4
☐ Pain prevents me from standing at all	5

Section 7: Sleeping	**Score**
☐ My sleep is never disturbed by pain	0
☐ My sleep is occasionally disturbed by pain	1
☐ Because of pain I have less than 6 hours sleep	2
☐ Because of pain I have less than 4 hours sleep	3
☐ Because of pain I have less than 2 hours sleep	4
☐ Pain prevents me from sleeping at all	5

Section 8: Sex Life (if applicable)	**Score**
☐ My sex life is normal and causes no extra pain	0
☐ My sex life is normal but causes some extra pain	1
☐ My sex life is nearly normal but is very painful	2
☐ My sex life is severely restricted by pain	3
☐ My sex life is nearly absent because of pain	4
☐ Pain prevents any sex life at all	5

Section 9: Social Life	**Score**
☐ My social life is normal and gives me no extra pain	0
☐ My social life is normal but increases the degree of pain	1
☐ Pain has no significant effect on my social life apart from limiting my more energetic interests e.g. sport	2
☐ Pain has restricted my social life and I do not go out as often	3
☐ Pain has restricted my social life to my home	4
☐ I have no social life because of pain	5

Section 10: Travelling	**Score**
☐ I can travel anywhere without pain	0
☐ I can travel anywhere but it gives me extra pain	1
☐ Pain is bad but I manage journeys over two hours	2
☐ Pain restricts me to journeys of less than one hour	3
☐ Pain restricts me to short necessary journeys under 30 minutes	4
☐ Pain prevents me from travelling except to receive treatment	5

Source: *www.tac.vic.gov.au/upload/Oswestry.pdf*

Table 8. *Scoring system for Oswestry Disability Index (ODI).*

For each section of six statements the total score is 5; if the first statement is marked the score = 0; if the last statement is marked it = 5. Intervening statements are scored according to rank. If more than one box is marked in each section, take the highest score. If all 10 sections are completed the score is calculated as follows:
Example: if 16 (total scored) out of 50 (total possible score) x 100 = 32%. **If one section is missed (or not applicable) the score is calculated:** **Example: 16 (total scored)/45 (total possible score) x 100 = 35.6%.**
So the final score may be summarized as: **(total score/[5 x number of questions answered]) x 100%.**
It is suggested rounding the percentage to a whole number for convenience.

Source: Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine* 2000; 25:2940-2952 (272).

Similar to ODI (272), RDQ (274,378) has been used in a wide variety of situations over many years and is available in a number of languages as a condition-specific health status measure for low back pain. RDQ (274) is a health status measure designed to be completed by patients to assess physical disability due to low back pain. It was designed for use in research as an outcome measure for clinical trials, but has also been found useful for monitoring patients in clinical practice (378). The RDQ was derived from the Sickness Impact Profile (SIP) (379), which is a 136 item health status measure covering all aspects of physical and mental function. Of these, 25 items relating specifically to physical functions that were likely to be affected by low back pain were selected. The RDQ focuses on a limited range of physical functions, which include walking, bending over, sitting, lying down, dressing, sleeping, self-care, and daily activities as illustrated in Table 9. Since these functions are chosen specifically for low back pain, the scoring system does not permit or require a non-applicable response. Further, the statements in the RDQ focus almost exclusively on physical function, with only one question on mood. Some aspects of physical function are not explicitly included, for example, lifting and twisting or turning.

11.2.3.1 Scoring

Patients completing the RDQ are asked to place a check mark beside a statement if it applies to them that day. Consequently, this approach emphasizes short-term changes in response to treatment. The RDQ score is calculated by adding up the number of items checked. Items are not weighted. The scores therefore range from 0 (no disability) to 24 (maximum disability). Although designed for administration on paper, the RDQ has also been satisfactorily administered on computer and by telephone (378). The RDQ is short, simple to complete, and readily understood by the patient.

11.2.3.2 Reliability and Validity

The limited range of RDQ is considered both a strength and weakness in its content validity. The questionnaire covers only a limited range of the problems that a patient with back pain may face and, in particular, does not address psychological or social problems. Consequently, it is essential to combine RDQ with specific measures of other issues including psychological or social problems. It has been stated that the restricted nature of the domains covered by the RDQ is a strength, as it makes the scores easy to understand and interpret (378).

In assessing the construct validity, RDQ scores correlate well with other measures of physical function, including the physical subscales of SF-36, the SIP (379-382), the Quebec Back Scale (383), the ODI (384,385), and pain ratings (303). However, it shows only modest correlation with direct measures of physical function (380,386,387), which is in common with other self-reported disability measures.

RDQ has been shown to have good psychometric properties, evidenced by internal consistency and responsiveness (378). Reproducibility in chronic low back pain, when evaluated 39 days apart, has been shown to be 0.72 (382).

11.2.3.3 Clinically Significant Change

It has been suggested that the smallest change likely to be clinically significant lies between 2.5 and

Table 9. *The Roland-Morris Disability questionnaire (RDQ).*

Only mark the sentence if you are sure that it describes you today.

1. ☐ I stay at home most of the time because of the pain in my back.
2. ☐ I change position frequently to try and make my back comfortable.
3. ☐ I walk more slowly than usual because of the pain in my back.
4. ☐ Because of the pain in my back, I am not doing any of the jobs that I usually do around the house.
5. ☐ Because of the pain in my back, I use a handrail to get upstairs.
6. ☐ Because of the pain in my back, I lie down to rest more often.
7. ☐ Because of the pain in my back, I have to hold on to something to get out of a reclining chair.
8. ☐ Because of the pain in my back, I ask other people to do things for me.
9. ☐ I get dressed more slowly than usual because of the pain in my back.
10. ☐ I only stand up for short periods of time because of the pain in my back.
11. ☐ Because of the pain in my back, I try not to bend or kneel down.
12. ☐ I find it difficult to get out of a chair because of the pain in my back.
13. ☐ My back hurts most of the time.
14. ☐ I find it difficult to turn over in bed because of the pain in my back.
15. ☐ My appetite is not very good because of the pain in my back.
16. ☐ I have trouble putting on my socks (or stockings) because of the pain in my back.
17. ☐ I only walk short distances because of the pain in my back.
18. ☐ I sleep less because of the pain in my back.
19. ☐ Because of the pain in my back, I get dressed with help from someone else.
20. ☐ I sit down for most of the day because of the pain in my back.
21. ☐ I avoid heavy jobs around the house because of the pain in my back.
22. ☐ Because of the pain in my back, I am more irritable and bad tempered with people.
23. ☐ Because of the pain in my back, I go upstairs more slowly than usual.
24. ☐ I stay in bed most of the time because of the pain in my back.

Source: www.rmdq.org

5 points. However, this may vary depending on the level of disability of the patients (378). Stratford et al (388) suggest that minimal clinically important change (MCIC) in scores is 1 to 2 points for patients with little disability, 7 to 8 points for patients reporting high levels of disability, and 5 points in unselected patients. Others (381) have suggested a change of 2 to 3 points as the MCID for a 23-item version of RDQ. It has been suggested that for sample size calculations for clinical trials, the changes in scores of 2 to 3 points on the RDQ be used since setting the MCID as high as 5 in designing a clinical trial would risk underpowering the trial because fewer patients are needed if a trial is designed on the basis of a large change in score.

11.2.3.4 Oswestry Disability Index vs. Roland-Morris Disability Questionnaire

RDQ and ODI are similar in many aspects (378). However, a greater proportion of patients score in the top half of the distribution of RDQ scores than in the top half of ODI scores (389), but at high levels of disability, the ODI may still show change when RDQ scores are maximal. At the other end of the scale, RDQ scores may still discriminate when ODI scores are at a minimum (390). Consequently, it has been recommended by Roland and Fairbank (378) to use ODI in

patients who are likely to have persistent severe disability and the RDQ in patients who are likely to have relatively little disability.

The RDQ and ODI scores are highly correlated with similar test-retest reliability and internal consistency (390,391). It has been found that the properties of the 2 instruments were very similar in discriminating power, including ability to detect change over time (385). However, some have reported that the ODI performs better (384) or the reverse (392), or that the result depends on the exact composition being made (391). In a recent evaluation (393) responsiveness of the ODI and the RDQ, for patients with mild to moderate low back pain and disability, the ODI was the most responsive measure for patients with mild to moderate low back pain disability.

In summary, the RDQ and ODI are simple methods of self-rated assessments of physical function in patients with low back pain. Since the majority of the patients in interventional pain management do suffer with moderate to severe disability, ODI may be more appropriate than RDQ.

11.2.4 Neck Pain Disability Index

The NDI developed by Vernon and Mior (275) was derived from the ODI and has been validated in several study populations (277,394-398). It consists of 10 items referring to various activities and pain with 6 possible answers for each item. The score of each item varies between 0 and 5, resulting in a total score of 0 to 50. Table 10 illustrates the NDI and scoring. One of the strengths of the NDI is that it has been validated against multiple measures of function, pain, and clinical signs/symptoms (276).

11.2.4.1 Validity and Reliability

Concurrent criterion validity was established by a correlation coefficient of 0.6 between the NDI and the VAS evaluating overall improvement (276). The correlation coefficient with the MPQ is 0.7 (279). It was also shown that a single item (pain intensity) and the total NDI score were the only significant predictors of pain scores. It was concluded that both NDI and NRS exhibit fair to moderate test-retest reliability in patients with mechanical neck pain with adequate responsiveness (398). However, the MCID required to be certain that the change in scores has surpassed a level that could be contributed to measurement error for the NDI was twice that which was previously being reported.

A significant correlation between the NDI and both physical and mental health components of SF-36 has been identified (399). Sensitivity to change was further substantiated by calculating the effect sizes for change scores of both the NDI and SF-36 (400). Moderate reliability and construct validity of NDI and patients specific functional scale in patients with cervical radiculopathy has also been illustrated (401).

11.2.4.2 Scoring

The NDI is scored using a percentage of the maximal pain and disability score. The items are organized by type of activity and followed by 6 different assertions expressing progressive levels of functional capacity, similar to ODI.

11.2.4.3 Clinically Significant Change

Studies have been published evaluating neck pain, effect sizes, and standard response means, minimal detectable change (MDC), and MCIC (277,292,294,401-403). Adequate responsiveness in patients with mechanical neck pain has been shown (398); however, the MCID required to be certain that the change in scores has surpassed a level that could be contributed to measurement error for the NDI was twice that which has previously been reported. Others (277,403) identified the MDC as 5 points or 10% change.

11.3 Minimal Clinically Important Difference (MCID)

MCID was defined as, "… the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" (388,404). Test responsiveness refers to the ability of a test to detect clinically important change over time. Thus, it is crucial to distinguish between the responsiveness as a test property and the MCID as a quantity useful in interpreting study results.

MCIC was preferred by others (295) for the change of health status and the term MCID to indicate differences between patients. Estimating the MCIC of relevant outcomes measures enables a comparison between interventions on the patient level. Linking the MCIC to economic evaluations may contribute to the relevance and interpretability of these studies. Several reviews and studies have presented a clear overview of the different methods to assess MCIC and provided some priorities for future research (405-410).

12. Data Presentation and Analysis

Understanding of research methods in the mod-

Table 10. *Questions and scoring of Neck Pain Disability Index (NDI).*

Section 1: Pain Intensity	**Score**
☐ I have no pain at the moment	0
☐ The pain is very mild at the moment	1
☐ The pain is moderate at the moment	2
☐ The pain is fairly severe at the moment	3
☐ The pain is very severe at the moment	4
☐ The pain is the worst imaginable at the moment	5

Section 2: Personal Care (eg. washing, dressing)	**Score**
☐ I can look after myself normally without causing extra pain	0
☐ I can look after myself normally but it causes extra pain	1
☐ It is painful to look after myself and I am slow and careful	2
☐ I need some help but can manage most of my personal care	3
☐ I need help every day in most aspects of self-care	4
☐ I do not get dressed, wash with difficulty and stay in bed	5

Section 3: Lifting	**Score**
☐ I can lift heavy weights without extra pain	0
☐ I can lift heavy weights but it gives me extra pain	1
☐ Pain prevents me lifting heavy weights off the floor but I can manage if they are conveniently placed eg. on a table	2
☐ Pain prevents me lifting heavy weights but I can manage light to medium weights if they are conveniently positioned	3
☐ I can only lift very light weights	4
☐ I cannot lift or carry anything	5

Section 4 - Reading	**Score**
☐ I can read as much as I want to with no pain in my neck.	0
☐ I can read as much as I want to with slight pain in my neck.	1
☐ I can read as much as I want to with moderate pain in my neck.	2
☐ I cannot read as much as I want because of moderate pain in my neck.	3
☐ I cannot read as much as I want because of severe pain in my neck.	4
☐ I cannot read at all.	5

Section 5: Headaches	**Score**
☐ I have no headaches at all.	0
☐ I have slight headaches which come infrequently.	1
☐ I have moderate headaches which come infrequently.	2
☐ I have moderate headaches which come frequently.	3
☐ I have severe headaches which come frequently.	4
☐ I have headaches almost all the time.	5

Section 6: Concentration	**Score**
☐ I can concentrate fully when I want to with no difficulty.	0
☐ I can concentrate fully when I want to with slight difficulty.	1
☐ I have a fair degree of difficulty in concentrating when I want to.	2
☐ I have a lot of difficulty in concentrating when I want to.	3
☐ I have a great deal of difficulty in concentrating when I want to.	4
☐ I cannot concentrate at all.	5

Section 7: Work	**Score**
☐ I can do as much work as I want to.	0
☐ I can only do my usual work, but no more.	1
☐ I can do most of my usual work, but no more.	2
☐ I cannot do my usual work.	3
☐ I can hardly do any work at all.	4
☐ I cannot do any work at all.	5

Section 8: Driving	**Score**
☐ I can drive my car without any neck pain.	0
☐ I can drive my car as long as I want with slight pain in my neck.	1
☐ I can drive my car as long as I want with moderate pain in my neck.	2
☐ I cannot drive my car as long as I want because of moderate pain in my neck.	3
☐ I can hardly drive at all because of severe pain in my neck.	4
☐ I cannot drive my car at all.	5

Section 9: Sleeping	**Score**
☐ I have no trouble sleeping.	0
☐ My sleep is slightly disturbed (less than 1 hour sleepless).	1
☐ My sleep is mildly disturbed (1-2 hours sleepless).	2
☐ My sleep is moderately disturbed (2-3 hours sleepless).	3
☐ My sleep is greatly disturbed (3-5 hours sleepless).	4
☐ My sleep is completely disturbed (5-7 hours)	5

Section 10: Recreation	**Score**
☐ I am able to engage in all of my recreational activities with no neck pain at all.	0
☐ I am able to engage in all of my recreational activities with some pain in my neck.	1
☐ I am able to engage in most, but not all of my recreational activities because of pain in my neck.	2
☐ I am able to engage in a few of my recreational activities because of pain in my neck.	3
☐ I can hardly do any recreational activities because of pain in my neck.	4
☐ I cannot do any recreational activities at all.	5

Note: The score of each item varies between 0 (no pain and no functional limitation) and 5 (worst pain and maximal limitation) resulting in a total score of 0 (no disability) to 50 (totally disabled)

Source: www.srisd.com/NDI.pdf

ern day environment in evaluation of EBM, along with statistical techniques used to assist in drawing conclusions is essential. However, the methods of statistical inference in current use are not "evidence-based" and thus have contributed to a widespread misconception (411). The misperception is that absent any consideration of biological plausibility and prior evidence, statistical methods can provide a number that by itself reflects a probability of reaching erroneous conclusions. It is believed that this has damaged the quality of scientific reasoning and discourse, primarily by making it difficult to understand how the strength of evidence in a particular study can be related to and combined with the strength of other evidence.

Statistical methods are important in comparing groups for determination of sample size, outcomes, and additional analysis. Multiple deficiencies have been described in the publication of medical statistics (132,411-414).

12.1 Sample Size

For scientific and ethical reasons the sample size for a trial needs to be planned carefully, with a balance between clinical and statistical considerations. Ideally, a study should be large enough to have a high probability (power) of detecting a statistically significant clinically important difference of a given size if such a difference exists. The size of effect deemed important is inversely related to the sample size necessary to detect it; that is, large samples are necessary to detect small differences (108). Elements of the sample size calculation include the estimated outcomes in each group (which implies the clinically important target difference between the intervention groups), type I and type II error levels and the standard deviation (SD) of the measurements for continuous outcomes (415).

It has been widely stated that reports of studies with small samples frequently include the erroneous conclusion that the intervention groups do not differ, when too few patients were studied to make such a claim (416). Reviews of published trials have consistently found that a high proportion of trials have very low power to detect clinically meaningful treatment effects (417,418). However, in reality, small but clinically valuable true differences are likely, which require large trials to detect (419). The median sample size was 54 patients in 196 trials in arthritis (211), 46 patients in 73 trials in dermatology (130), and 65 patients in 2000 trials in schizophrenia (13). Many reviews have found that few authors report how they determined the sample size (13,130,135,222); however, there is little merit in calculating the statistical power once the results of the trial are known as the power is then appropriately indicated by confidence intervals (CIs) (417).

In interventional pain management studies, a sample size of 50 in the smallest group has been considered to be appropriate (Table 2) (28). For interventional techniques, in the methodologic quality assessment, a sample size of 50 or less patients in the smallest group will lose – 17% of the total points available.

12.1.1 Determination of Sample Size

In interventional pain management, sample size determinations are essential to demonstrate the existence of a difference or relationship, as well as the estimated magnitude of the relationship in a clinical trial. The sample size calculations are based on significance tests, using the power of a test to help choose the sample size required to detect a difference if it exists. The power of a test is related to the postulated difference in the population, the standard error of the sample difference, and the significance level. These quantities are linked by an equation which enables us to determine any of them given the others.

12.1.2 Parameter Definition

An appropriate sample size generally depends on 5 study design parameters (415,420,421). These are 1) minimum expected difference or the effect size, 2) estimated measurement variability, 3) desired statistical power, 4) significance criterion, and 5) whether a one- or two-tailed statistical analysis is planned.

12.1.2.1 Minimum Expected Difference

This parameter is the smallest measured difference between comparison groups that the investigator would like the study to detect (420). As the minimum expected difference is made smaller, the sample size needed to detect statistical significance increases. The setting of this parameter is subjective and is based on clinical judgment and experience with the problem being investigated. The results of pilot studies or a literature review can guide the selection of a reasonable minimum difference.

12.1.2.2 Estimated Measurement Variability

This parameter is represented by the expected SD in

the measurements made within each comparison group (420). As statistical variability increases, the sample size needed to detect the minimum difference increases. If preliminary data are not available, this parameter may have to be estimated on the basis of subjective experience, or a range of values may be assumed. A separate estimate of measurement variability is not required when the measurement being compared is a proportion (in contrast to a mean), because the SD is mathematically derived from the proportion.

12.1.2.3 Statistical Power

Statistical power is the power that is desired from the study. Power is increased as sample size increases. While high power is always desirable, there is an obvious trade-off with the number of individuals that can feasibly be studied, given the usually fixed amount of time and resources available to conduct a study. The statistical power is customarily set to a number greater than or equal to 0.80 in RCTs. However, many clinical trial experts now are advocating a power of 0.90 (420).

12.1.2.4 Significance Criterion

This parameter is the maximum *P* value for which a difference is to be considered statistically significant. As the significance criterion is decreased (made more strict), the sample size needed to detect the minimum difference increases. The significance criterion is customarily set to 0.05.

12.1.2.5 One- or Two-Tailed Statistical Analysis

Generally it is not known whether or one- or two-tailed statistical analysis is performed. In a few cases, it may be known before the study that any difference between comparison groups is possibly in only one direction. In such cases, use of one-tailed statistical analysis, which would require a smaller sample size for detection of the minimum difference than would a two-tailed analysis, may be considered. Two-tail analysis is most commonly performed.

12.1.2.6 Unequal Numbers in Each Group

For a given total sample size, the maximum power is achieved by having equal numbers of subjects in 2 groups. However, in some clinical trials, the numbers of subjects taking one treatment may have to be limited, so to achieve the necessary power one has to allocate more patients to the other treatment (422). If one were to maintain the same sample size as calculated for a 1:1 ratio but then allocated in the ratio 2:1, the loss in power would be quite small (around 5%) (415). However, if the allocation ratio is allowed to exceed 2:1 with the same total sample size, the power falls very quickly (a loss of around 25% in power for a ratio of 5:1) and consequently, a considerably larger total sample size is required to maintain a fixed power with an imbalanced study than with a balanced one.

12.1.2.7 Minimizing the Sample Size

Multiple strategies have been described for minimizing the sample size (423). These include use of continuous measurements instead of categories, more precise measurements, paired measurements, unequal group sizes, and expanding the minimum expected difference.

In interventional pain management, continuous measurements are commonly utilized. Thus, statistical tests that incorporate the use of continuous values are mathematically more powerful than those used for proportions, given the same sample size. In addition, in interventional pain management, precision can be increased by simply repeating the measurement. However, equations are more complex for studies involving repeated measurements in the same individuals (424).

The sample size may also be minimized by using paired measurements such as paired t-tests which are mathematically more powerful for a given sample size than are unpaired tests. Because of the paired tests, each measurement is matched with its own control.

The additional power and reduction sample size are due to the SD being smaller for changes within individuals than for overall differences between groups of individuals. Thus, studies with long-term follow-up provide higher statistical power with a smaller sample size.

Utilizing unequal group sizes may also assist in minimizing the sample size. Even though, it is statistically most efficient if the 2 groups are equal in size, there is still benefit gained by studying more individuals, even if the additional individuals all belong to one of the groups. However, more complex equations are necessary for calculating sample sizes when comparing means and proportions of unequal group sizes (425-429).

Finally, the expansion of the minimum expected difference that has been specified, especially if the planned study is a preliminary one, can significantly minimize the sample size. The results of a preliminary study could be used to justify a more ambitious follow-up study of a larger number of individuals and a smaller minimum difference.

12.1.2.8 Estimating Sample Sizes and Power

The task of calculating sample size and power requires clinical knowledge, detailed knowledge of the measurement tools, and statistical knowledge. The steps for experimental studies are first, determine the null hypothesis and either a one- or two-tailed alternative hypothesis, second, select an appropriate statistical test, third, choose a reasonable effect size (and variability), fourth, set an α (alpha) and β (beta) , and fifth, use the appropriate table or formula to estimate sample size (423).

12.1.2.8.1 Hypotheses

Hypotheses are needed in studies that will use tests of statistical significance to compare findings among groups. They also help to focus on the primary objective of the study.

A null hypothesis is the formal basis for testing statistical significance, indicating no association between the predictor and the outcome variable.

In contrast, the proposition that there is an association is called the alternative hypothesis. This cannot be tested directly; it is accepted by exclusion if the test of statistical significance rejects the null hypothesis.

A one-tailed alternative hypothesis specifies the direction of the association between the predictor and outcome variables. For example, the prediction that more pain relief will be experienced by patients receiving local anesthetic alone than those receiving local anesthetic and steroid is a one-tailed hypothesis. A two-tailed hypothesis states only that an association exists. A one-tailed hypothesis has the statistical advantage of permitting a smaller sample size than a two-tailed hypothesis, but it is not often appropriate and should be used with caution.

Hypotheses are also needed in studies which seek to measure the strength of the linear association between 2 continuous variables.

12.1.2.8.2 Statistical Tests

Variables are either continuous, discrete, or categorical. Continuous variables can take on any value within a defined range of values, and measurement is possible within whole units and fractional parts of units, e.g., age, height, weight. Discrete variables deal only with whole numbers, they can take on only certain definite and separate values, e.g., number of employees in an organization, number of receptionists on duty at a time. Categorical variables are further classified as nominal (unordered) or ordinal (ordered), and according to whether or not they are dichotomous (only 2 categories, e.g., sex).

The t-test is commonly used to determine whether the mean value of a continuous outcome variable in one group differs significantly from that in another group. The t-test assumes the distribution (spread) of the variable in the 2 groups approximates a normal (bell-shaped) curve.

12.1.2.8.3 Effect Size and Variability

Generally, data from other studies or pilot tests can be used to make an informed guess. If these are not available, researchers need to choose the smallest effect that would be clinically meaningful.

If using continuous data, variability of the response will also need to be estimated. Variability is determined by estimating the SD. This will necessitate looking at historical data derived under similar conditions, or when there is no such data, undertaking a pilot study.

Statistical tests depend on being able to show a difference between the groups being compared. The greater the variability (or spread) in the outcome variable among the subjects, the more likely it is that the values in the groups will overlap, and the more difficult it will be to demonstrate an overall difference between them.

12.1.2.8.4 α, β, and power

The probability of a type I error (i.e., rejecting the null hypothesis when it is actually true) is called α (alpha), or the "level of statistical significance." The probability of a type II error (failing to reject the null hypothesis when it is actually false) is called β (beta). Ideally α and β would be set at zero, eliminating the possibility of false-positive and false-negative results. In practice they are made as small as possible.

The quantity [1 - β] is called power. If β is set at 0.10, then a 10% chance of missing an association of a given effect size is accepted. This represents a power of 0.90, i.e. 90% chance of finding an association of that size if it exists.

Many studies set arbitrary values: α at 0.05, and β at 0.20 (a power of 0.80). When data are analyzed, statistical tests determine the *P* value - the probability of obtaining the study results by chance if the null hypothesis is true. The null hypothesis is rejected in favor of its alternative if the *P* value is less than the predetermined level of statistical significance.

Table 11. *Factors that affect sample size calculations.*

Factor	Magnitude	Impact on Identification of Effect	Required Sample Size
P value	Small	Stringent criterion; difficult to achieve "significance"	Large
	Large	Relaxed criterion; "significance" easier to attain	Small
Power	Low	Identification more probable	Small
	High	Difficult to identify	Large
Effect	Small	Easy to identify	Large
	Large		Small

Source: Whitley E, Ball J. Statistics review 4: sample size calculations. Crit Care 2002; 6:335-341.

12.1.2.8.5 Estimating Sample Size

Table 11 illustrates factors that affect sample size calculations (430).

12.1.2.8.6 An Example

The research objective is to compare 2 types of caudal epidural injections either with local anesthetic alone or with local anesthetic and steroid in the treatment of low back pain with or without lower extremity pain. The outcome variable used is that of NRS reduction after 12 months. A previous study has reported the mean NRS reduction is 4.6 with a standard deviation 1.46. The objective is to detect a difference of 25% or more in NRS between the 2 treatment groups. How many patients are required in each group ("local anesthetic and steroid") at α (2-tailed) = 0.05 and power = 0.80 (β=0.20)?

1. H0 = Both groups "local anesthetic and steroid" have the same reduction in NRS
2. H1 = local anesthetic group is inferior to steroid treatment
3. Efficacy size = 1.15 (25% of 4.6)
4. Standardized effect size (SE) = effect size / standard deviation

   Standardized effect size =1.15/1.46 = 0.79 $\cong$ 0.8
5. α = 0.05 and β = 0.20 (power = 80%)

Looking across from a standardized effect size 0.8 in the leftmost column of Table 12, and down from α (two-tailed) = 0.05 and β = 0.20, 26 patients are required per group.

If α = 0.05 and β = 0.10 (power = 90%)

Looking across from a standardized effect size 0.8 in the leftmost column of Table 12, and down from α (two-tailed) = 0.05 and β = 0.10, 34 patients are required per group.

Table 12 illustrates sample size required per group when using a t-test to compare means of continuous variables.

12.1.2.8.7 Power

The power of a statistical test is the probability that the test will reject a false null hypothesis (that it will not make a type II error). As power increases, the chances of a type II error decrease. The probability of a type II error is referred to as the false negative rate (β). Therefore power is equal to 1 – β.

Power analysis can either be done before (a priori) or after (post hoc) data is collected. A priori power analysis is conducted prior to the research study, and is typically used to determine an appropriate sample size to achieve adequate power. Post-hoc power analysis is conducted after a study has been completed, and uses the obtained sample size and effect size to determine what the power was in the study, assuming the effect size in the sample is equal to the effect size in the population.

12.2 Statistical Methods

Data can be analyzed in many ways, some of which may not be strictly appropriate in a particular situation. Almost all methods of analysis yield an estimate of the treatment effect, which is a contrast between the outcomes and the comparison groups. In addition, 95% CIs for the estimated effect is essential as it indicates a range of uncertainty for the true treatment effect. Study findings can also be assessed in terms of their statistical significance. The *P* value represents the probability that the observed data could have arisen by chance when the interventions did not differ. Ac-

Table 12 *Sample size required per group when using t-test to compare means of continuous variables.*

One-sided α = Two-sided α = E/S* β =	0.005 0.01			0.025 0.05			0.05 0.10		
	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20
0.10	3,565	2978	2,338	2,600	2,103	1,571	2,166	1,714	1,238
0.15	1,586	1325	1040	1,157	935	699	963	762	551
0.20	893	746	586	651	527	394	542	429	310
0.25	572	478	376	417	338	253	347	275	199
0.30	398	333	262	290	235	176	242	191	139
0.40	225	188	148	164	133	100	136	108	78
0.50	145	121	96	105	86	64	88	70	51
0.60	101	85	67	74	60	45	61	49	36
0.70	75	63	50	55	44	34	45	36	26
0.80	58	49	39	42	**34**	**26**	35	28	21
0.90	46	39	21	34	27	21	28	22	16
1.00	38	32	26	27	23	17	23	18	14

*E/S is the standardized effect size, computed as E (expected effect size) divided by S (standard deviation of the outcome variable). To estimate the sample size, read across from the standardized effect size, and down from the specified values of α and β for the required sample size in each group.

Source: Browner WS et al. Estimating sample size and power. In Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB (eds). Designing Clinical Research: An Epidemiologic Approach, 2nd ed. Lippincott Williams & Wilkins, Philadelphia, 2001, pp 65-84 (418).

tual *P* values are preferred to imprecise threshold reports (429,431).

Standard methods of analysis assume that the data are "independent." However, for controlled trials, it usually means that there is one observation per participant. Treating multiple observations from one participant as independent data is a serious error and such data are produced when outcomes can be measured on different parts of the body, as in interventional pain management. Data analysis should be based on counting each participant as one (432,433) or should be done by using more complex statistical procedures (434). Further, subgroup analysis or additional analysis requires specific statistical methods (435-448).

12.2.1 Parametric vs. Non-Parametric Statistics

Typically used parametric tests are t-tests, as well as Analysis of Covariance (ANCOVA), whereas Mann-Whitney is the non-parametric alternative. When the data are sampled from a normal distribution, the t-test has very slightly higher power than Mann-Whitney. However, when data are sampled from any one of a variety of non-normal distributions, Mann-Whitney is superior, often by a large amount.

Parametric as well as non-parametric statistics are utilized in the analysis of randomized trials (449). Altman (450) states that "parametric methods require the observations within each group to have an approximately normal distribution . . . if the raw data do not satisfy these conditions . . . a non-parametric method should be used." Further, introductory statistics textbooks typically advise against the use of parametric methods, such as the t-test for the analysis of randomized trials unless data approximate to a normal distribution. In addition, it has been stated that, parametric methods are applicable examples if the sample size suitably large: "for reasonably large samples (say, 30 or more observations in each sample) . . . the t-test may be computed on almost any set of continuous data (451)."

The rationale for recommending non-parametric over parametric methods, unless certain conditions are met, is rarely made explicit (449). However, techniques for statistical inference from randomized trials

can only fail in one of 2 ways, either by inappropriately rejecting the null hypothesis of no difference between groups (false-positive or type I error) or inappropriately failing to reject the null hypothesis (false-negative or type II error). Consequently, Vickers (449) recommends that any recommendation to favor one technique over another must be based on the relative rates of type I or type II errors. The empirical statistical research has clearly demonstrated that the t-test does not inflate type I (false-positive error) except in 5% of the time (452). Thus, concern over the relative advantages of parametric and non-parametric methods is focused on type II errors or false-negative results (453-456).

Where an endpoint is measured at baseline and again at follow-up, the t-test is not the recommended parametric method. Instead, ANCOVA, where a baseline score is added as a covariate in a linear regression, has been shown to be more powerful than the t-test (424,457-459).

12.2.2 The P Value

The *P* value is defined as the probability, under the assumption of no effect or no difference (the null hypothesis) of obtaining a result equal to or more extreme than what was actually observed (Fig. 3) (411). As shown in Fig. 3, the bell-shaped curve represents the probability of every possible outcome under the null hypothesis, both α (the type I error rate) and the *P* value are "tail areas" under this curve. The tail area for α is set before the experiment and a result can fall anywhere within it. The *P* value tail area is known only after the result is observed, and, by definition, the result will always lie on the border of that area (411).

Fisher (460) proposed a *P* value as an informed index to be used as a measure of discrepancy between the data and null hypothesis. Fisher also suggested that it be used as a part of the fluid, non-quantifiable process of drawing conclusions from observations, a process that included combining the *P* value in some unspecified way with background information.

Most researchers and readers think that a *P* value of 0.05 means that the null hypothesis has a probability of only 5%. However, this may be an unfortunate misinterpretation of the *P* value (459-464). A *P* value of 0.05 represents that there is a 95% or greater chance

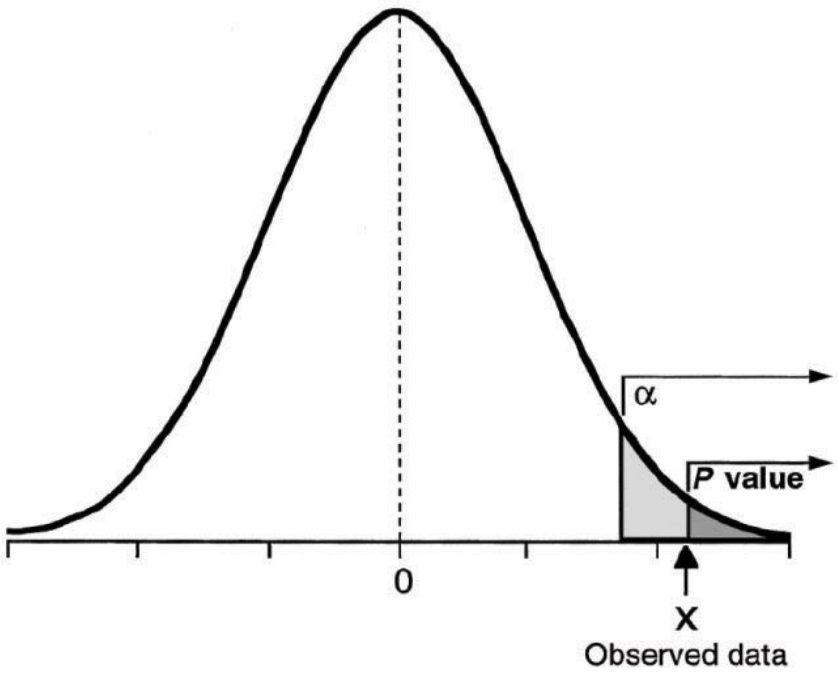

Fig. 3. *Illustration of a bell-shaped curve representing the probability of every possible outcome under the null hypothesis*

that the null hypothesis is correct. While this is an understandable, but categorically wrong interpretation because the *P* value is calculated on the assumption that the null hypothesis is true, it cannot therefore be a direct measure of the probability that the null hypothesis is false. Thus, this error may reinforce the mistaken notion that the data alone can tell the probability that a hypothesis is true (411).

The most powerful criticism of the *P* value was that it was a measure of evidence that did not take into account the size of the observed effect. When the *P* value was proposed, some scientists and statisticians attacked the logical basis and practical utility (465,466). A small effect in a study with large sample size can have the same *P* value as a large effect in a small study. This criticism is the foundation for the emphasis on CIs rather than *P* values (467-470).

The chance factor (P value) set arbitrarily at 5% or 0.5 and accepted as the standard is also called α. However, if α is set too high at 10% or 0.1, the risk of making an "α error" increases when the difference exists, which could be that the difference was due to a chance. Further, if α is set too low, the risk of missing a difference exists (471). The possibility of concluding that a difference does not exist when it does is called a "β error." By conversion, a β of 0.2 or 20% is thought to be the minimum needed. Consequently, researchers are more willing to risk making a β error (incorrectly concluding that a difference does not exist), then they are making an α error by incorrectly concluding that a difference does exist.

Generally it is believed that standard statistical methods are a great improvement over the chaos that preceded them and that they have proved enormously useful in practice (472-474).

12.2.3 Confidence Intervals

The CIs, along with *P* values, are crucial to determine the likelihood that a difference in a study is due to chance. Consequently, citing CIs is becoming more frequent and some journals do not accept manuscripts unless CIs are cited.

CIs are far from a panacea (411). In essence, CIs embody many of the same problems that afflict current methods, albeit in a subtler form (475). The most important drawback of CIs is that they offer no mechanism to unite external evidence with that provided by an experiment. Thus, CIs are a not a solution to the most serious problem created by frequentist methods, even though they are a step in the right direction (411).

If an article concludes no difference was found, the manuscript should describe the level of certainty (power with which they can make a conclusion). If a statistical difference is seen, then by definition there was a sufficient power. If the power is really high with a huge sample size compared with the number actually needed so that the power is 99% or so, statistical differences can be seen even when very small real clinical differences exist with narrow CIs. This means the difference is likely to be real and not due to chance, but the question remains if the difference is clinically significant.

12.2.4 Hypothesis Tests

Neyman and Pearson (476) thought Fisher's *P* value was an incomplete answer to the problem of developing an inferential method. In their hypothesis test, the authors pose 2 hypotheses: a null hypothesis indicating a statement that there is a null effect, and an alternative hypothesis, which is usually the opposite of the null hypothesis indicating that there is a non-zero effect (411). However, the researchers risk 2 types of errors – behaving as though 2 therapies differ when they are actually the same, committing either a type I error or a type II error. However, these errors can be calculated with mathematical formulas deductively and therefore objectively. In practice, this reports only whether or not the results were statistically significant and acting in accordance with that verdict, which may be considered profoundly non-scientific, even though this is often held up as a paradigm of the scientific methodology (411). Hypothesis testing is described as equivalent to a system of justice that is not concerned with which individual defendant is found guilty or innocent, but tries instead to control the overall number of incorrect verdicts.

12.2.5 Subgroup Analysis

Subgroup analysis and additional analysis are important in clinical trials (435-444). Subgroup analysis means any evaluation of treatment effects for a specific endpoint in some groups of patients are defined by baseline characteristics. Such analysis, which assesses the heterogeneity of treatment effects in subgroups of patients, may provide useful information for the care of patients and for future research (435). However, subgroup analysis also introduces analytic challenges and can lead to overstated and misleading results (436-442). Further, the subgroup analysis requires an appropriate statistical method for assessing the heterogeneity of treatment effects among the

levels of a baseline variable begins with a statistical test for interaction (232,443-444).

There have been multiple subgroup analyses performed in interventional pain management (65,69,444-448). van Wijk et al (445), in a study which appeared elegant and technically competent, described radiofrequency denervation of lumbar facet joints in the treatment of chronic low back pain. They concluded that the combined outcome measure and VAS showed no difference between radiofrequency and sham, though in both groups, significant VAS improvement occurred. The study was riddled with multiple flaws, but even then, in a subgroup analysis (445) concluded that there was substantial pain reduction. Karppinen et al (65) in a study of lumbar transforaminal epidural steroid injections showed negative results; however, in a subgroup analysis of the same data (69), they showed that steroid injections produced significant treatment effects and short-term improvement with cost benefits. However, Manchikanti et al (447,448) described difficulties in the subgroup analysis of psychological characteristics and age on the diagnosis of facet joint pain.

Subgroup analysis can be wrong in 2 ways. First, they can falsely indicate their treatment is beneficial in a particular subgroup when the trial shows no overall effect – the situation in which subgroup analysis is most commonly done (477-481). Simulations of RCTs power to determine the overall effect of treatment suggests that false subgroup effects will be noted by chance in 7% to 21% of analyses depending on other factors, which is seen more commonly in 41% to 66% of the simulated subgroups (479). Thus, the benefit is most likely to be absent in small subgroups, which probably explains the recurrent and usually mistaken finding that treatments are ineffective in subgroups, who tend to be under-represented in RCTs.

Some investigators avoid the issue of multiplicity of testing by tabulating the observed outcomes for the subgroups of interest without undertaking any formal statistical analysis (482). Further, the common practice of performing subgroup-specific tests of treatment effect is flawed in that it is testing the wrong hypothesis (483). The hypothesis that should be tested is whether the treatment effect in a subgroup is significantly different from that in the overall population.

The appropriate test to use when analyzing heterogeneity of responses among subgroups are interaction tests (439,479), for which multiple examples are available (483,484). Given the risks of false-positive findings when multiple subgroup analyses are performed, it is not surprising if a subgroup-specific test shows a significant ($P < 0.05$) or suggestive ($P = 0.05$ to $P = 0.10$) effect of treatment, even when the trial failed to do so overall (439,441).

13. Design Of Protocol And Reporting

RCTs today are usually conducted using an elaborate set of rules and procedures (108,109). The details for conducting the study are defined in the study protocol. In addition, for any controlled trial, prior to the beginning of the trial, the investigation must be reviewed by an IRB to evaluate the quality of the study design, the ethics of conducting the study, and the safeguards provided for patients, including a review of the informed consent statement. Further, informed consent also must be reviewed for compliance under the Health Insurance Portability and Accountability Act (HIPAA) regulations required to ensure the confidentiality of study data. In addition, all controlled trials must be registered with the U.S. National Institutes of Health Clinical Trial Registry of the United States at www.clinicaltrials.gov.

A study must be design based on the CONSORT statement for reporting randomized trials as illustrated in Table 1 (108). The CONSORT statement was developed to alleviate the problem of inadequate reporting of RCTs (108). An extension of the CONSORT recommendations to non-inferiority and equivalence trials was also developed (109). Most RCTs aim to determine whether an intervention is effective. By contrast, equivalence trials (485) aim to determine whether one intervention is therapeutically similar to another (109). A non-inferiority trial seeks to determine whether a treatment is no worse than the reference treatment (109).

The investigative team also should consider methodologic quality criteria utilized for assessment as described in Table 2 (28). It is essential to provide high quality, well-designed, and properly executed RCTs eligible for publication and utilized in systematic reviews and guideline preparation as high quality trials. It has been repeatedly shown that critical appraisal of the quality of clinical trials of the design, conduct, and analysis of RCTs is far from being transparent, incomplete, and confounded by poor methodology (108).

The CONSORT statements provide checklists and flow diagrams that authors can use for reporting RCTs (108,109). Checklists of items to include when report-

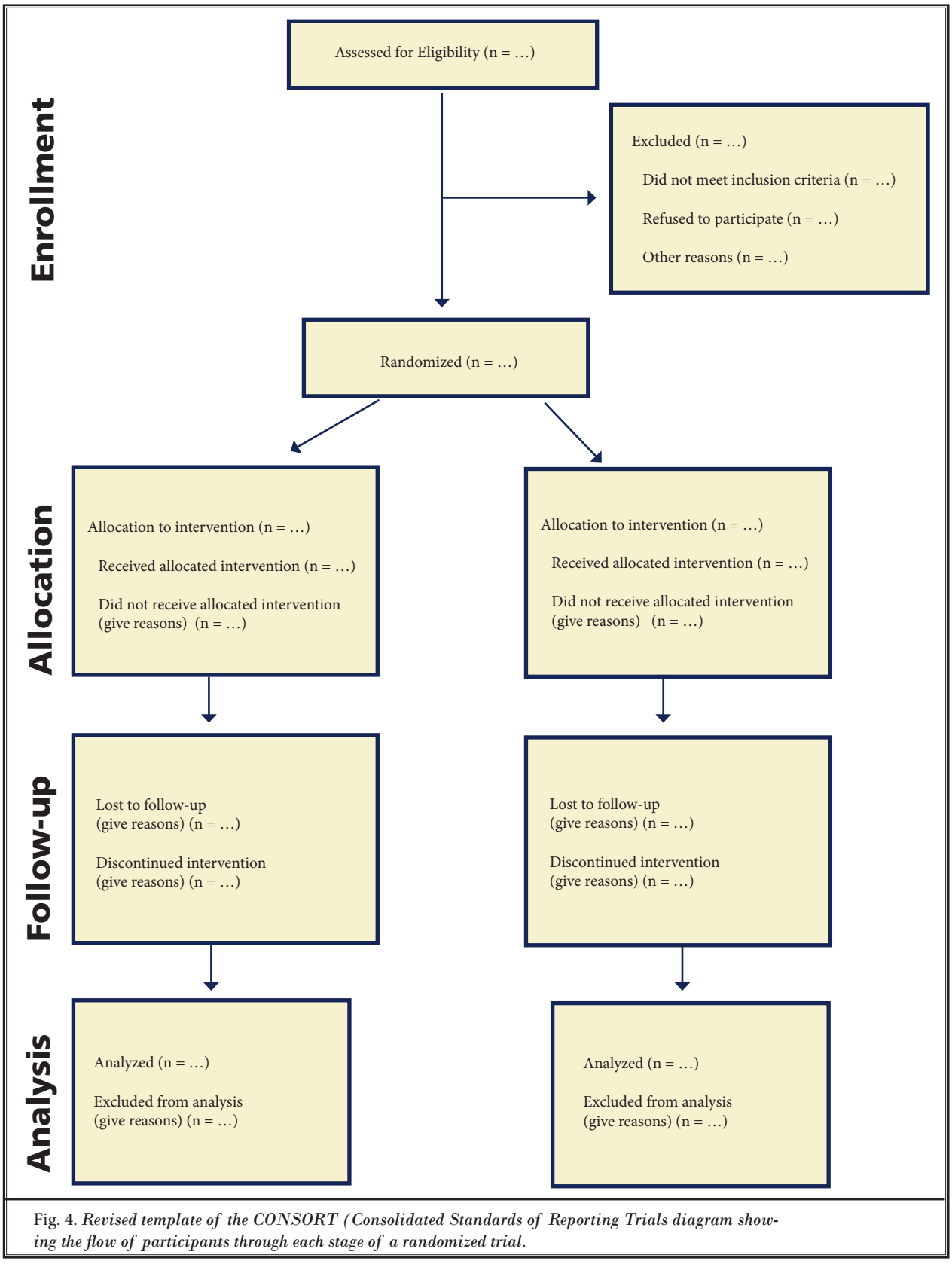

Fig. 4. *Revised template of the CONSORT (Consolidated Standards of Reporting Trials diagram showing the flow of participants through each stage of a randomized trial.*

ing randomized trials are shown in Table 1. Figure 4 shows the revised template of the CONSORT diagram showing the flow of participants through each stage of a randomized trial.

13.1 Title and Abstract

The title should illustrate clearly the allocation. It is recommended that superiority or equivalence trials and open trials should be titled in a manner that it is understood.

The structured abstract must provide a series of headings pertaining to the design, conduct, and analysis of a trial with standardized information appearing under each heading (108,109,486). It has been shown that structured abstracts are of higher quality than the more traditional descriptive abstracts (487), and they also allow readers to find information more easily (488).

Authors should address in detail conflicts of interest, and provide full information about extent of industry involvement.

13.2 Introduction

The introduction includes the scientific background and an explanation of rationale. Typically, it includes free-flowing text, without a structured format, in which the authors explain the scientific background of the context and scientific rationale for their trial. In general, a rationale is of 2 types, either explanatory or pragmatic. Pragmatic is to guide practice by comparing the clinical effects of 2 alternative treatments, whereas explanatory is with a placebo group. The introduction should provide an appropriate explanation for how the intervention might work and the research involving people should be based on a thorough knowledge of the scientific literature (489,490). Ideally, the need for a new trial should be justified and also reference to previous systematic reviews or similar trials should be made.

In reporting of non-inferiority and equivalence randomized trials, the rationale for using a non-inferiority or equivalence design also must be described.

13.3 Methods

Methods include a description of the eligibility criteria for participants; the settings and the locations where the data were collected; precise details of the interventions intended for each group and how and when they were actually administered; specific objectives and hypothesis; clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements such as multiple observations, training of assessors, etc.; determination of the sample size and an explanation of any interim analysis and stopping rules; the method used to generate random allocation sequence, details of any restriction of randomization, and the method used to implement the random allocation sequence namely allocation concealment; implementation of randomization, blinding or masking, and the success of such a process; and statistical methods used to compare groups for primary outcomes and methods for subgroup or additional analysis.

13.3.1 Participants

Eligibility criteria for participants and the settings and locations where the data were collected must be described. Further, in the reporting of non-inferiority and equivalence trials eligibility criteria for participants describing whether participants in the non-inferiority or equivalence trial are similar to those in any trials that establish efficacy of the reference treatment and the settings and locations.

Each and every RCT in effect addresses an issue relevant to some population with the condition of interest. Eligibility criteria are utilized to restrict the population and performance of the trial to one or a few centers (108). Typical selection criteria may relate to age, sex, clinical diagnosis, and comorbid conditions. Exclusion criteria are used to ensure safety. Eligibility criteria must be explicitly defined and described.

As a cautionary measure any known relevant inaccuracy in patients' diagnosis should be discussed to avoid that factor in affecting the power of the trial (491). The distinction between inclusion and exclusion criteria is not essential (108,492).

Descriptions of the characteristics of participants and the setting in which they were studied are essential for external validity or generalizability of the trial results. Of particular importance is the method of recruitment, such as by referral or self-selection (for example, through advertisements). The method of recruitment applied before randomization and eligibility criteria do not affect the internal validity of a trial, but method of recruitment may affect the external validity.

Settings and locations are important as health care institutions vary greatly in their organization, experience, and resources, and the baseline risk for the medical condition under investigation, which affects external

validity. In addition, climate and other physical factors, economics, geography, and the social and cultural milieu can all affect a study's external validity (108). Consequently, it is essential that the number and type of settings and care providers involved is reported to assess the external validity. The setting description should include the country, city, and immediate environment, for example office based practice, hospital outpatient practice, type of specialty center (spine, interventional pain management, pain medicine, etc.), hospital inpatient setting, tertiary referral center, primary care center, or ambulatory surgery center.

13.3.2 Interventions

Precise details of the interventions intended for each group, including the detailed description of whether the reference treatment in the non-inferiority or equivalence trial is identical or very similar to that in any trial(s) that establish efficacy, and how and when they were actually administered must be described. The characteristics of a placebo and the way in which it was described should also be reported. It is especially important to describe thoroughly the "usual care" given to a control group or an intervention that is in fact a combination of interventions (108). It is also important to describe the experience of the investigator or investigators, as it is necessary to describe the number, training, and experience of interventionalists or surgeons in addition to the intervention itself (108,493). In a case of multiple-component interventions, details of timing and duration of interventions must be reported. Further, any differences between the control intervention in the trial and the reference treatment in the previous trial(s) in which efficacy was established should be reported and explained (109). Differences may exist because background treatment and patient management change with time and concomitant therapies may differ (494). As an example, a change in a dose of the reference treatment may also result in reduced efficacy and an increased tolerability may be an issue overestimating the new treatment's advantages.

13.3.3 Objectives

Objectives are the questions that the trial was designed to answer, often relating to the efficacy of a particular therapeutic or preventive intervention. In contrast, the hypothesis or prescribed questions being tested to help meet the objectives are more specific than objective and are amenable to explicit statistical evaluation. However, in practice, objectives and hypothesis are not always easily differentiated. This is the most commonly met requirement in randomized trials (108). Further specific objectives and hypothesis must also include the hypothesis concerning non-inferiority or equivalence.

13.3.4 Outcomes

All RCTs assess response variables, or outcomes, for which the groups are compared. However, some outcomes are of more interest than others, thus the primary outcome measure is the prespecified outcome of greatest importance and is usually the one used in the sample size calculation. Any other outcome assessments are considered as secondary outcomes. Thus, primary and secondary outcome measures, detailing whether the outcomes in the non-inferiority or equivalence trial are identical (or very similar) to those in any trial(s) that establish efficacy of the reference treatment and, when applicable, any methods used to enhance the quality of measurements such as multiple observations, training of assessors, etc., must be clearly defined (108,109).

Both primary and secondary outcomes should be identified and completely defined. It is also helpful to provide the details of prespecified time points of primary interest and methods of outcome assessment. As described earlier, it is also essential to describe the outcome instruments, how they were chosen, and minimal clinical identifiable changes for each instrument.

13.3.5 Sample Size

As detailed in this document, for scientific and ethical reasons, the sample size for a trial needs to be planned carefully, with a balance between clinical and statistical considerations (Table 11). A description should include, in detail, the determination of the sample size, along with details whether it was calculated using a non-inferiority or equivalence criterion and specifying the margin of equivalence with the rationale for its choice.

Whenever applicable, an explanation of any interim analysis and stopping rules should be described.

13.3.6 Randomization

Under the heading of randomization, multiple items are essential which include sequence generation, allocation concealment, and implementation.

13.3.6.1 Sequence Generation

The description of the method used to generate the random allocation sequence, including details of

any restriction such as blocking, stratification, etc. is essential (Table 5).

13.3.6.2 Allocation Concealment

While sequence generation discusses generation of an unpredictable sequence of assignments, it is of considerable importance to understand how the sequence is applied when participants are enrolled into the trial. Multiple methods for allocation concealment have been described (Table 5). Allocation concealment should not be confused with blinding.

13.3.7 Implementation

In addition to knowing the methods used in concealment of the allocated intervention at the time of enrollment, it is also important to understand how the random sequence was implemented; specifically, who generated the allocation sequence, who enrolled the participants, and who assigned the participants to trial groups.

13.3.8 Blinding

Blinding refers to keeping study participants, healthcare providers, and assessors unaware of the assigned intervention. The trials may be open, single-blind, or double-blind. Further, it is essential to know how the success of blinding was evaluated.

13.3.9 Statistical Methods

While data can be analyzed in many ways, it is essential to specify which statistical procedure was used for each analysis, and further clarification for each analysis. In addition, methods for additional analysis, such as subgroup analysis and adjusted analysis, should be reported. Thus, statistical methods are used to compare groups for primary outcome(s), specifying whether a 1- or 2-sided or CI approach was used.

13.4 Results

Results of a properly conducted randomized trial should include participant flow, method of recruitment, baseline data, numbers analyzed, outcomes and estimation, ancillary analysis, and adverse events.

13.4.1 Participant Flow

Participant flow is an extremely important part of the equation. Thus, it should be described in detail as shown in Figure 4. Occasionally, the design and execution of some RCTs is straightforward, and the flow of participants through each phase of the study can be

Table 13. *Information required to document the flow of participants through each stage of a randomized, controlled trial.*

Stage	Number of People Included	Number of People Not Included or Excluded	Rationale
Enrollment	People evaluated for potential enrollment	People who did not meet the inclusion criteria People who met the inclusion criteria but declined to be enrolled	These counts indicate whether trial participants were likely to be representative of all patients seen; they are relevant to assessment of external validity only, and they are often not available
Randomization	Participants randomly assigned		Crucial count for defining trial size and assessing whether a trial has been analyzed by intention to treat
Treatment allocation	Participants who received treatment as allocated, by study group	Participants who did not receive treatment as allocated, by study group	Important counts for assessment of internal validity and interpretation of results; reasons for not receiving treatment as allocated should be given
Follow-up	Participants who completed treatment as allocated, by study group Participants who completed follow-up as planned, by study group	Participants who did not complete treatment as allocated, by study group Participants who did not complete follow-up as planned, by study group	Important counts for assessment of internal validity and interpretation of results; reasons for not completing treatment or follow-up should be given
Analysis	Participants included in main analysis, by study group	Participants excluded from main analysis, by study group	Crucial count for assessing whether a trial has been analyzed by intention to treat; reasons for excluding participants should be given

Source: Altman DG, et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001; 134:663-694 (108).

described adequately in a few sentences. However, in more complex studies, it may be difficult to discern whether and why some participants did not receive the treatment as allocated, were lost to follow-up, or were excluded from the analysis (495). Table 13 shows the information required to document the flow of participants through each stage of the RCT, through various stages of the study, starting from enrollment to data analysis, with the number of people included through various stages, and the number of people not included or excluded, along with the rationale for each action. The flow diagram may be expanded to describe treatment allocation or any other aspects and also other information may be added (108).

A participant flow diagram and description also should describe protocol deviations from the study as planned, together with the reasons.

13.4.2 Numbers Analyzed

The number of participants in each group is an essential element of the results. Even though the flow diagram indicates the number of participants for whom outcomes were available, these numbers may vary for different outcome measures. Thus, the clear descriptions of the number of participants (denominator in each group should be included in each analysis) and whether "intention-to-treat" and an alternative analysis were conducted should be reported. Further, the results should be stated in an absolute number when feasible (e.g., 30 of 60, not 50%). Failure to include all participants may bias trial results. Most trials do not yield perfect data. Further, "protocol violations" may occur, such as when patients do not receive the full intervention or the correct intervention or few ineligible patients are randomly allocated in error.

13.4.3 Recruitment

It is essential to describe when a study took place and over what period participants were recruited. It is also important to understand the rate at which participants were recruited. Medical and surgical therapies, including concurrent therapies, evolve continuously and may affect the routine care given to patients during a trial. Further, the length of follow-up, which is not always a fixed a period after randomization, must be shown with a median duration of follow-up (496,497). It is also important to report if the trial was stopped owing to the results of interim analysis of data, as early stopping will lead to a discrepancy between the planned and actual sample sizes. In addition, trials that stop early are likely to overestimate the treatment effect (498).

13.4.4 Baseline Data

Even though the eligibility criteria indicates who was eligible for the trial, it is also important to describe the characteristics of participants who were actually recruited. This information provides the importance and clinical relevance of the trial. Generally RCTs aim to compare groups of participants that differ only with respect to intervention, thus maintaining similar baseline characteristics among the groups. While proper random assignment prevents selection bias, it does not guarantee that the groups are equivalent at baseline and any difference in baseline characteristics are, however, the result of chance rather than bias (222), in contrast to widely held misbeliefs (37). Baseline data are extremely valuable when the outcome measure can also be measured at the start of the trial. Baseline information should be presented efficiently for both groups. The variability of the data should be reported with average values and continuous variables can be summarized for each group by the mean and SD. If continuous data have an asymmetrical distribution, a preferable approach may be to quote the median and percentile range (25th and 75th percentiles) (431). However, standard errors and CIs are not appropriate for describing variability as they are inferential rather than descriptive characteristics.

Despite many warnings about their inappropriateness (221,222,499), significance tests of baseline differences are common and reported in 50% of the trials (108,439). The trial protocol and the result description should state whether or not adjustment is made for nominated baseline variables by using ANCOVA (90,500).

13.4.5 Intention-to-Treat Analysis

The intention-to-treat strategy is commonly recommended to handle such issues as protocol violations and withdrawals to analyze all participants according to their original group assignment, regardless of what subsequently occurred. However, this is not always straightforward to implement. It is common for some patients not to complete a study – they may drop out or be withdrawn from active treatment – and thus are not assessed at the end. Even though those participants cannot be included in the analysis, it is customary still to refer to the analysis of all available participants as an intention-to-treat analysis.

However, the term is also used inappropriately when some participants for whom data are available are excluded; specifically, if the patients received none of the intended treatment because of non-adherence to the protocol, etc. Conversely, analysis can be restricted to only participants who fulfill the protocol in terms of eligibility, interventions, and outcome assessment. This analysis may be considered as per protocol analysis and may be compared with intention-to-treat analysis. However, non-compliance with assigned therapy may mean that the intention-to-treat analysis underestimates the real benefit of the treatment (501,502). Consequently, additional analysis may be considered. It has been reported that studies reporting an intention-to-treat analysis were associated with some other aspects of good study design and reporting, such as describing a sample size calculation (503).

The inclusion of all subjects in intention-to-treat analyses regardless of their follow-up status requires investigators to deal with the resulting missing data by one of several approaches.

13.4.5.1 Last Observation Carried Forward

The most common approach is the replacement of each subject's missing data with his or her last non-missing observation, a method called last observation carried forward (LOCF) (504). This method works best if the observations are expected to remain at some level or if there are only a few missing values. If the observations in a test are expected to increase or decrease over time this method does not work very well.

13.4.5.2 Best or Worst Case Imputation

Two other methods when dealing with missing data are best case and worst case imputation. Here, the best or worst data is imputed. This leads to either an under or over evaluation of the data and can be used "to assess a lower bound of efficacy as a demonstration of robustness" (505). There are different ways of using worst case imputation.

13.4.5.3 Mean Value Methods

A natural method of imputation is to use the mean value of the recorded observations. This method leads to lower variance and a concern here is that the dropouts might be more likely to be patients with more extreme values (i.e., a very ill patient might not show up). Another aspect of using the mean value is that it is not always clear on which data you should calculate the mean value. One method is mean value for the whole period and the other one is mean of previous and next visit.

13.4.5.4 Regression Methods

Linear regression methods can be used for imputation. Calculations need to control for factors studied which are not being investigated for association.

13.4.6 Outcomes

For both primary and secondary outcomes, results should be reported as a summary of the outcome in each group, together with the contrast between the groups and contrast between baseline and predetermined follow-up periods known as the effect size. Appropriate CIs should be presented for the contrast between groups. However, the CONSORT statement (108) cautions that a common error is the presentation of separate CIs for the outcome in each group rather than for the treatment effect (506). In general, trial results are often more clearly displayed in a table rather than in the text.

CIs to indicate the precision of the estimate should be provided for all outcome measures (108,430,507). The results should be provided, using either 95% CIs or in conjunction with *P* values, rather than solely as *P* values (506-513). Further, results should be reported for all planned primary and secondary endpoints, not just for analysis that was statistically significant. Selective reporting has been considered to be widespread and a serious problem, even though there is little empirical evidence of within-study selective reporting (211,514,515). It may also be beneficial to calculate the number needed to treat for benefit or harm (516,517).

13.4.7 Ancillary Analysis

Authors should resist the temptation to perform many subgroup analyses (436-442). Multiple analyses of the same data create considerable risk for false-positive findings. Consequently, analyses that were prespecified in the trial protocol are much more reliable than those suggested by the data.

However, if subgroup analysis is to be undertaken, the authors should report appropriate methodology rather than selective reporting which could lead to bias (518). Similarly, analysis in which adjustments were made for baseline variables should be clearly reported in both formats, adjusted and unadjusted.

13.4.8 Adverse Events

Interventions may have unintended and often undesirable effects in addition to expected and in-

tended effects. Thus, all important adverse events or side effects in each intervention group should be reported. Adverse events are crucial in the application of the data in practice and may have a major impact on whether a particular intervention will be deemed acceptable and useful or not. However, some reported adverse events may not be as much a consequence of the intervention as the progression of a disease, etc. Thus, it is considered that controlled trials offer the best approach for providing safety data, as well as efficacy data, even though rare adverse effects are not detected in randomized trials (108). Many reports of RCTs are considered to provide inadequate information on adverse events (108). Thus, at a minimum, estimates of the frequency of the main severe adverse events and reactions for treatment discontinuation should be provided separately for each intervention group. It is also essential to provide the operational definition for their measures of severity of adverse events in the protocol, as well as in the report (519).

13.5 Discussion

Discussion should describe interpretation of the results, generalizability of the results, and overall evidence.

13.5.1 Interpretation

The CONSORT statement (108) describes that the discussion sections of scientific reports are filled with rhetorics supporting the authors' findings and provide little measured argument of the pros and cons of the study and its results (520). In fact, some journals have encouraged a structure for the authors' discussion of their results (521,522). For example, the Annals of Internal Medicine (521) recommends that authors structure the discussion section by presenting: 1) a brief synopsis of the key findings; 2) consideration of possible mechanisms and explanation; 3) comparison with relevant findings from other published studies; 4) limitations of the present study and methods used to minimize and compensate for those limitations; and 5) a brief section that summarizes the clinical and research implications of the work, as appropriate. It is of particular importance to discuss the weaknesses and limitations of the study (108,523,524). Along with the limitations, discussion of any imprecision of the results is essential to be included in the weakness. Imprecision may arise in connection with several aspects of a study, including measurement of a primary outcome or diagnosis (108).

Finally, the authors should describe the difference between statistical significance and clinical importance.

13.5.2 Generalizability

External validity, also known as generalizability or applicability, is the extent to which the results of a study can be generalized to other circumstances (108,525). However, internal validity is also essential as it is a prerequisite for external validity. The results of a flawed study are invalid and the question of its external validity becomes irrelevant. Generalizability essentially conveys whether the results are applicable to an individual patient or groups that differ from those enrolled in the trial with regard to age, sex, severity of disease, and comorbid conditions or they can be applied at the primary, secondary, and tertiary levels of care. Since external validity is a matter of judgment and depends on the characteristics of the participants included in the trial setting, the treatment regimens tested, and the outcomes assessed, it is crucial that adequate information be provided about eligibility criteria and the setting and location, the interventions and how they were administered, the definition of outcomes, and the period of recruitment and follow-up (108,526,527). Even though several considerations are important when results of a trial are applied to an individual patient or general population, in general, interventions found to be beneficial in a narrow range of patients have broader applications in actual practice. Multiple measures that incorporate baseline risk and therapeutic effects, such as the number needed to treat to obtain one additional favorable outcome and the number needed to treat to produce one adverse event, are helpful in assessing the benefit-to-risk ratio (108,527). This will facilitate the clinician's ability to integrate the information in patient management.

13.5.3 Overall Evidence

Overall evidence shows the general interpretation of the results in the context of current evidence. This can be achieved by including a formal systematic review in the results or discussion of the report, provided such reports exist. This provides a basis for interpretation of the evidence as the totality, along with information about whether the results are similar to those of other trials in the same topic area and the degree of similarity.

14. Discussion

Assessment of healthcare interventions can be misleading unless investigators ensure unbiased com-

parisons. Even though N of 1 RCT is considered to be at the top of the hierarchy of strength of evidence, randomized trials remain the only method that eliminates selection and confounding biases. Basically, one may state that the standard RCTs are in fact set up to show that treatments do not work, rather than to demonstrate that treatments do work. While RCTs are considered to provide the most internally valued evidence for medical decision-making, they provide only partial answers. In interventional pain management settings, results from clinical trials, both randomized and observational, with substantial impact on patient care, have been ruled ineffective based on flawed methodology and evidence synthesis. However, recent results also have provided empirical evidence that some RCTs have biased results and in some cases there was no difference between observational and randomized trials. The poorly executed trials, whether randomized or non-randomized, tend to exaggerate treatment effects and to have important biases. Thus, it is essential to produce high-quality research, which consistently eliminates bias and shows significant effect size.

The design, implementation, and reporting of an RCT require methodologic as well as clinical expertise including meticulous effort, a high index of suspicion for unanticipated difficulties, potentially unnoticed problems, and methodological deficiencies, and skills to report the findings appropriately with close attention to minimizing bias. Sound reporting encompasses adequate reporting, and the conduct of ethical trials rests on the footing of sound science, which will not subject readers to speculation (108). Interventional pain specialists must understand the differences between multiple types of trials – placebo-controlled, pragmatic, other types of controls, and types of blinding, and must have a clear understanding of the randomization procedures including random allocation sequence and allocation concealment. Further, clearly defined primary and secondary outcome measures and appropriate presentation of the results are also essential. The most commonly utilized pain assessment instruments are NRS or VAS, and functional assessment instruments are the ODI and NDI. Other measures may include psychological assessment.

References

1. Guyatt G, Drummond R. Part 1. The Basics: Using the Medical Literature. 1A. Introduction: The philosophy of evidence-based medicine. In *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. American Medical Association, Chicago, 2002, pp 3-12.
2. Napodano RJ. *Values in Medical Practice*. Human Sciences Press, New York, 1986.
3. Haynes RB, Sackett RB, Gray JM, Cook DC, Guyatt GH. Transferring evidence from research into practice: 1. The role of clinical care research evidence in clinical decisions. *ACP J Club* 1996; 125: A14-A16.
4. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The n-of-1 randomized controlled trial: Clinical usefulness. Our three-year experience. *Ann Intern Med* 1990; 112:293-299.
5. Larson EB, Ellsworth AJ. Randomized clinical trials in single patients during a 2-year period. *JAMA* 1993; 270:2708-2712.
6. Guyatt G, Drummond R. Part 2. The Basics: Using and Teaching the Principles of Evidence-Based Medicine. 2B1. Therapy and validity. In *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. American Medical Association, Chicago, 2002, pp 247-308.
7. Mahon J, Laupacis A, Donner A, Wood T. Randomised study of n of 1 trials versus standard practice. *BMJ* 1996; 312:1069-1074.
8. Nisbett R, Ross L. *Human Inference*. Prentice-Hall, Englewood Cliffs, 1980.
9. Guyatt G, Drummond R. Part 1. The Basics: Using the Medical Literature. 1C2. Diagnostic tests. In *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. American Medical Association, Chicago, 2002, pp 121-140.
10. Hotopf M. The pragmatic randomized controlled trial. *Adv Psychiatr Treat* 2002; 8:326-333.
11. Williams DD, Garner J. The case against 'the evidence': A different perspective on evidence-based medicine. *Br J Psychiatry* 2002; 180:8-12.
12. Hotopf M, Lewis G, Normand C. Putting trials on trial: The costs and consequences of small trials in depression: A systematic review of methodology. *J Epidemiol Community Health* 1997; 51:354-358.
13. Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998; 317:1181-1184.
14. Hotopf M, Churchill R, Lewis G. Pragmatic randomized controlled trials in psychiatry. *Br J Psychiatry* 1999; 175:217-223.
15. Miles A, Charlton B, Bentley P, Polychronis A, Grey J, Price N. New perspectives in the evidence-based healthcare debate. *J Eval Clin Pract* 2000; 6:77-84.
16. Healy D. Randomized Controlled Trials: Evidence Biased Psychiatry. Alliance for Human Research Protection, 2002. www.ahrp.org/COI/healyo802.php
17. Fisher RA. The arrangement of field experiments. *J Ministry Ag* 1926; 33:503-513.
18. Amberson JB, McMahon BT, Pinner MA. Clinical trial of sanocrysin in pulmonary tuberculosis. *Am Rev Tuberc* 1931; 24:401-435.
19. Marshall G, Blacklock JWS, Cameron C, Capon NB, Cruickshank R, Gaddum JH, Heaf FRG, Hill AB, Houghton LE, Hoyle JC, Raistrick H, Scadding JG, Tytler WH, Wilson GW, Hart P. Streptomycin treatment of pulmonary tuberculosis: A Medical Research Council investigation. *BMJ* 1948; 2:769-782.

20. American College of Occupational and Environmental Medicine (ACOEM) Low back Disorders. In *Occupational Medicine Practice Guidelines: Evaluation and Management of Common Health Problems and Functional Recovery of Workers, Second Edition*. American College of Occupational and Environmental Medicine Press, Elk Grove Village, 2007.
21. American College of Occupational and Environmental Medicine (ACOEM) Chronic Pain. In *Occupational Medicine Practice Guidelines: Evaluation and Management of Common Health Problems and Functional Recovery of Workers, Second Edition*. OAmerican College of Occupational and Environmental Medicine Press, Elk Grove Village, awaiting publication.
22. Manchikanti L, Singh V, Derby R, Schultz DM, Benyamin RM, Prager JP, Hirsch JA. Reassessment of evidence synthesis of occupational medicine practice guidelines for interventional pain management. *Pain Physician* 2008; 11:393-482.
23. Manchikanti L, Singh V, Derby R, Helm S, Trescot AM, Staats PS, Prager JP, Hirsch JA. Review of occupational medicine practice guidelines for interventional pain management and potential implications. *Pain Physician* 2008; 11:271-289.
24. Manchikanti L, Singh V, Helm S, Trescot AM, Hirsch JA. A critical appraisal of 2007 American College of Occupational and Environmental Medicine (ACOEM) practice guidelines for interventional pain management: An independent review utilizing AGREE, AMA, IOM, and other criteria. *Pain Physician* 2008; 11:291-310.
25. Chou R. Using evidence in pain practice: Part I: Assessing quality of systematic reviews and clinical practice guidelines. *Pain Med* 2008; 9:518-530.
26. Chou R. Using evidence in pain practice: Part II: Interpreting and applying systematic reviews and clinical practice guidelines. *Pain Med* 2008; 9:531-541.
27. Boswell MV, Trescot AM, Datta S, Schultz DM, Hansen HC, Abdi S, Sehgal N, Shah RV, Singh V, Benyamin RM, Patel VB, Buenaventura RM, Colson JD, Cordner HJ, Epter RS, Jasper JF, Dunbar EE, Atluri SL, Bowman RC, Deer TR, Swicegood JR, Staats PS, Smith HS, Burton AW, Kloth DS, Giordano J, Manchikanti L. Interventional techniques: Evidence-based practice guidelines in the management of chronic spinal pain. *Pain Physician* 2007; 10:7-111.
28. Koes BW, Scholten RJ, Mens JM, Bouter LM. Efficacy of epidural steroid injections for low-back pain and sciatica: A systematic review of randomized clinical trials. *Pain* 1995; 63:279-288.
29. Sanders SH, Harden RN, Benson SE, Vicente PJ. Clinical practice guidelines for chronic non-malignant pain syndrome patients II: An evidence-based approach. *J Back Musc Rehabil* 1999; 13:47-58.
30. van Tulder MWV, Koes BW, Bouter LM. Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions. *Spine* 1997; 22:2128-2156.
31. Nelemans PJ, Debie RA, DeVet HC, Sturmans F. Injection therapy for subacute and chronic benign low back pain. *Spine* 2001; 26:501-515.
32. Geurts JW, van Wijk RM, Stolker RJ, Groen GJ. Efficacy of radiofrequency procedures for the treatment of spinal pain: A systematic review of randomized clinical trials. *Reg Anesth Pain Med* 2001; 26:394-400.
33. Resnick DK, Choudhri TF, Dailey AT, Groff MW, Khoo L, Matz PG, Mummaneni P, Watters WC 3rd, Wang J, Walters BC, Hadley MN; American Association of Neurological Surgeons/Congress of Neurological Surgeons. Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 13: injection therapies, low-back pain, and lumbar fusion. *J Neurosurg Spine* 2005; 2:707-715.
34. Niemisto L, Kalso E, Malmivaara A, Seitsalo S, Hurri H. Cochrane Collaboration Back Review Group. Radiofrequency denervation for neck and back pain: A systematic review within the framework of the Cochrane collaboration back review group. *Spine* 2003, 28:1877-1888.
35. Levin JH. Prospective, double-blind, randomized placebo-controlled trials in interventional spine: What the highest quality literature tells us. *Spine J* 2008; Sep 11 [Epub ahead of print].
36. Bigos SJ, Boyer OR, Braen GR, Brown K, Deyo R. *Acute Low Back Problems in Adults*. Clinical Practice Guideline Number 4. AHCPR Publication No. 95-0642. Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services, Rockville, December 1994.
37. Carragee EJ, Hurwitz EL, Cheng I, Carroll LJ, Nordin M, Guzman J, Peloso P, Holm LW, Côté P, Hogg-Johnson S, van der Velde G, Cassidy JD, Haldeman S, Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. Treatment of neck pain: Injections and surgical interventions: Results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine* 2008; 33:S153-S169.
38. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H, Zanoli G. Chapter 4: European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006; 15:S192-S300.
39. Armon C, Argoff CE, Samuels J, Backonja MM; Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Assessment: Use of epidural steroid injections to treat radicular lumbosacral pain: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology* 2007; 68:723-729.
40. Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: Systematic review of randomised controlled trials. *Int J Prosthodont* 2006; 19:126-128.
41. Kao LS, Tyson JE, Blakely ML, Lally KP. Clinical research methodology I: Introduction to randomized trials. *J Am Coll Surg* 2008; 206:361-369.
42. World Health Organization. *International Clinical Trials Registry Platform (ICTRP)*. 2007. www.who.int/entity/ictrp/en/
43. McLeod RS. Issues in surgical randomized controlled trials. *World J Surg* 1999; 23:1210-1214.
44. Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized controlled trials in surgery. *Surgery* 1994; 115:707-712.
45. Hardin WD Jr, Stylianos S, Lally KP. Evidence-based practice in pediatric surgery. *J Pediatr Surg* 1999; 34:908-912.
46. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: Problems and possible solutions. *BMJ* 2002; 324:1448-1451.
47. Antman K, Ayash L, Elias A, Wheeler C, Hunt M, Eder JP, Teicher BA, Critchlow J,

Bibbo J, Schnipper LE, Frei III E. A phase II study of high-dose cyclophosphamide, thiotepa, and carboplatin with autologous marrow support in women with measurable advanced breast cancer responding to standard-dose therapy. *J Clin Oncol* 1992; 10:102-110.

48. Farquhar C, Marjoribanks J, Basser R, Hetrick S, Lethaby A. High dose chemotherapy and autologous bone marrow or stem cell transplantation versus conventional chemotherapy for women with metastatic breast cancer. *Cochrane Database Syst Rev* 2005; CD003142.
49. Peters WP, Shpall EJ, Jones RB, Olsen GA, Bast RC, Gockerman JP, Moore JO. High-dose combination alkylating agents with bone marrow support as initial treatment for metastatic breast cancer. *J Clin Oncol* 1988; 6:1368-1376.
50. The EC/IC Bypass Study Group. Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke. Results of an international randomized trial. *N Engl J Med* 1985; 313:1191-1200.
51. Weinstein PR, Rodriguez Y, Baena R, Chater NL. Results of extracranial-intracranial arterial bypass for intracranial internal carotid artery stenosis: Review of 105 cases. *Neurosurgery* 1984; 15:787-794.
52. Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? *Surgery* 1995; 118:459-467.
53. Pawlik TM, Abdalla EK, Barnett CC, Ahmad SA, Cleary KR, Vauthey JN, Lee JE, Evans DB, Pisters PW. Feasibility of a randomized trial of extended lymphadenectomy for pancreatic cancer. *Arch Surg* 2005; 140:584-589.
54. Balasubramanian SP, Wiener M, Alshameeri Z, Tiruvoipati R, Elbourne D, Reed MW. Standards of reporting of randomized controlled trials in general surgery: Can we do better? *Ann Surg* 2006; 244:663-667.
55. Jacquier I, Boutron I, Moher D, Roy C, Ravaud P. The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: A systematic review. *Ann Surg* 2006; 244:677-683.
56. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Choice of Control Group and Related Issues in Clinical Trials E10. July 20, 2000.
57. Braunholtz DA, Edwards SJ, Lilford RJ. Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect." *J Clin Epidemiol* 2001; 54:217-224.
58. Weijer C, Freedman B, Fuks A, Robbins J, Shapiro S, Skrutkowska M. What difference does it make to be treated in a clinical trial? A pilot study. *Clin Invest Med* 1996; 19:179-183.
59. Manchikanti L, Pampati V, Damron KS. The role of placebo and nocebo effects of perioperative administration of sedatives and opioids in interventional pain management. *Pain Physician* 2005; 8:349-355.
60. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med* 2001; 344:1594-1602.
61. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment. *J Intern Med* 2004; 256:91-100.
62. Koshi EB, Short CA. Placebo theory and its implications for research and clinical practice: A review of the recent literature. *Pain Pract* 2007; 7:4-20.
63. Carette S, Leclaire R, Marcoux S, Morin F, Blaise G, St. Pierre A, Truchon R, Parent F, Levesque J, Bergeron V, Montminy P, Blanchette C. Epidural corticosteroid injections for sciatica due to herniated nucleus pulposus. *N Engl J Med* 1997; 336:1634-1640.
64. Carette S, Marcoux S, Truchon R, Grondin C, Gagnon J, Allard Y, Latulippe M. A controlled trial of corticosteroid injections into facet joints for chronic low back pain. *N Engl J Med* 1991; 325:1002-1007.
65. Karppinen J, Malmivaara A, Kurunlahti M, Kyllonen E, Pienimaki T, Nieminen P, Ohinmaa A, Tervonen O, Vanharanta H. Periradicular infiltration for sciatica. A randomized controlled trial. *Spine* 2001; 26:1059-1067.
66. Lord SM, Barnsley L, Wallis BJ, McDonald GJ, Bogduk N. Percutaneous radiofrequency neurotomy for chronic cervical zygapophyseal-joint pain. *N Engl J Med* 1996; 5:1721-1726.
67. Riew KD, Park JB, Cho YS, Gilula L, Patel A, Lenke LG, Bridwell KH. Nerve root blocks in the treatment of lumbar radicular pain. A minimum five-year follow-up. *J Bone Joint Surg Am* 2006; 88:1722-1725.
68. Riew KD, Yin Y, Gilula L, Bridwell KH, Lenke LG, Lauryssen C, Goette K. The effect of nerve-root injections on the need for operative treatment of lumbar radicular pain. A prospective, randomized, controlled, double-blind study. *J Bone Joint Surg Am* 2000; 82-A:1589-1593.
69. Karppinen J, Ohinmaa A, Malmivaara A, Kurunlahti M, Kyllonen E, Pienimaki T, Nieminen P, Tervonen O, Vanharanta H. Cost effectiveness of periradicular infiltration for sciatica. *Spine* 2001; 26:2587-2595.
70. Manchikanti L, Singh V, Falco FJE, Cash KA, Pampati V. Lumbar facet joint nerve blocks in managing chronic facet joint pain: One-year follow-up of a randomized, double-blind controlled trial: Clinical Trial NCT00355914. *Pain Physician* 2008; 11:121-132.
71. Manchikanti L, Singh V, Falco FJ, Cash KM, Fellows B. Cervical medial branch blocks for chronic cervical facet joint pain: A randomized, double-blind, controlled trial with 1-year follow-up. *Spine* 2008; 33:1813-1820.
72. Manchikanti L, Singh V, Falco FJ, Cash KM, Pampati V. Effectiveness of thoracic medial branch blocks in managing chronic pain: A preliminary report of a randomized, double-blind controlled trial: Clinical Trial NCT00355706. *Pain Physician* 2008; 11:491-504.
73. Manchikanti L, Pampati V, Fellows B, Bakhit CE. The diagnostic validity and therapeutic value of lumbar facet joint nerve blocks with or without adjuvant agents. *Curr Rev Pain* 2000; 4:337-344.
74. Cuckler JM, Bernini PA, Wiesel SW, Booth RE Jr, Rothman RH, Pickens GT. The use of epidural steroid in the treatment of radicular pain. *J Bone Joint Surg* 1985; 67:63-66.
75. Manchikanti KN, Pampati V, Damron KS, McManus CD. A double-blind, controlled evaluation of the value of Sarapin in neural blockade. *Pain Physician* 2004; 7:59-62.
76. Leclaire R, Fortin L, Lambert R, Bergeron YM, Rossignol M. Radiofrequency facet joint nerve denervation in the treatment of low back pain: A placebo-controlled clinical trial to assess efficacy. *Spine* 2001; 26:1411-1416.
77. Manchikanti L, Pampati V, Rivera JJ, Beyer CD, Damron KS, Barnhill RC. Cau-

dal epidural injections with Sarapin or steroids in chronic low back pain. *Pain Physician* 2001; 4:322-335.

78. Manchikanti L, Damron KS, Cash KA, Manchukonda R, Pampati V. Therapeutic medial branch blocks in managing chronic neck pain: A preliminary report of a randomized, double-blind, controlled trial: Clinical Trial NCT00033272. *Pain Physician* 2006; 9:333-346.
79. Manchikanti L, Singh V, Rivera JJ, Pampati V, Beyer CD, Damron KS, Barnhill RC. Effectiveness of caudal epidural injections in discogram positive and negative chronic low back pain. *Pain Physician* 2002; 5:18-29.
80. Manchikanti L. Interventional pain management: Past, present, and future. The Prithvi Raj lecture: Presented at the 4th World Congress-World Institute of Pain, Budapest, 2007. *Pain Pract* 2007; 7:357-371.
81. Byar DP, Simon RM, Friedewald WT, Schlesselman JJ, DeMets DL, Ellenberg JH, Gail MH, Ware JH. Randomized clinical trials: Perspectives on some recent ideas. *N Engl J Med* 1976; 295:74-80.
82. Feinstein AR. Current problems and future challenges in randomized clinical trials. *Circulation* 1984; 70:767-774.
83. Abel U, Koch A. The role of randomization in clinical studies: Myths and beliefs. *J Clin Epidemiol* 1999; 52:487-497.
84. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982; 72:233-240.
85. Kane RL. Approaching the outcomes question. In Kane RL (ed). *Understanding Health Care Outcomes Research*. Aspen Publications, Gaithersburg, 1997, pp 1-15.
86. Shikata S, Nakayama T, Noguchi Y, Taji Y, Yamagishi H. Comparison of effects in randomized controlled trials with observational studies in digestive surgery. *Ann Surg* 2006; 244:668-676.
87. Groenwold RH, Van Deursen AM, Hoes AW, Hak E. Poor quality of reporting confounding bias in observational intervention studies: A systematic review. *Ann Epidemiol* 2008 Aug 8; [Epub ahead of print].
88. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *Int J Epidemiol* 2007; 36:666-676.
89. Hartz A, Bentler S, Charlton M, Lanska D, Butani Y, Soomro GM, Benson K. Assessing observational studies of medical treatments. *Emerg Themes Epidemiol* 2005; 2:8.
90. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000; 342:1878-1886.
91. Hartz A, Benson K, Glaser J, Bentler S, Bhandari M. Assessing observational studies of spinal fusion and chemonucleolysis. *Spine* 2003; 28:2268-2275.
92. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; 342:1887-1892.
93. Shrier I, Boivin JF, Steele RJ, Platt RW, Furlan A, Kakuma R, Brophy J, Rossignol M. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 2007; 166:1203-1209.
94. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG. Evaluating non-randomised intervention studies. *Health Technol Assess* 2007; 11:1-178.
95. Fleiss JL, Levin B, Paik MC. How to randomize. In *Statistical Methods for Rates and Proportions. 3rd ed.* John Wiley & Sons, Hoboken, 2003, pp 86-94.
96. Kang M, Ragan BG, Park JH. Issues in outcomes research: An overview of randomization techniques for clinical trials. *J Athl Train* 2008; 43:215-221.
97. Manchikanti L, Pampati V. Research designs in interventional pain management: Is randomization superior, desirable or essential? *Pain Physician* 2002; 5:275-284.
98. Dreyfuss P, Baker R. In response to treatment of neck pain. *Eur Spine J* 2008; 17:1270-1272.
99. Barnsley L. Percutaneous radiofrequency neurotomy for chronic neck pain: Outcomes in a series of consecutive patients. *Pain Med* 2005; 6:282-286.
100. Govind J, King W, Bailey B, Bogduk N. Radiofrequency neurotomy for the treatment of third occipital headache. *J Neurol Neurosurg Psychiatr* 2003; 74:88-93.
101. McDonald GJ, Lord SM, Bogduk N. Long term follow-up of patients treated with cervical radiofrequency neurotomy for chronic neck pain. *Neurosurgery* 1999; 45:61-67.
102. Sapir DA, Gorup JM. Radiofrequency medial branch neurotomy in litigant and nonlitigant patients with cervical whiplash. *Spine* 2001; 26:E268-E273.
103. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L. *Systems to Rate the Strength of Scientific Evidence*, Evidence Report, Technology Assessment No. 47. AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality, 2002. www.thecre.com/pdf/ahrq-system-strength.pdf
104. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282:1061-1066.
105. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995; 274:645-651.
106. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3:25.
107. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Clin Chem* 2003; 49:1-6.
108. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001; 134:663-694.
109. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ, CONSORT Group. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. *JAMA* 2006; 295:1152-1160.
110. Kunz R, Vist GE, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2007; MR000012.
111. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimen-

sions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-412.

112. Moher D. CONSORT: An evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated Standards of Reporting Trials. *JAMA* 1998; 279:1489-1491.
113. Kjaergard LL, Nikolova D, Gluud C. Randomized clinical trial in hepatology: Predictors of quality. *Hepatology* 1999; 30:1134-1138.
114. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ* 2001; 323:42-46.
115. Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, Hollingsworth JC, Ashton CM, Wray NP. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002; 347:81-88.
116. Gillespie WJ. Arthroscopic surgery was not effective for relieving pain or improving function in osteoarthritis of the knee. *ACP J Club* 2003; 138:49.
117. Jackson RW. Arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002; 347:1717-1719.
118. Morse LJ. Arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002; 347:1717-1719.
119. Chambers KG, Schulzer M. Arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002; 347:1717-1719.
120. Ellis TJ, Crawford D. Arthroscopic surgery for arthritis of the knee. *Curr Womens Health Rep* 2003; 3:63-64.
121. Ewing W, Ewing JW. Arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002; 347:1717.
122. Kelly MA. Role of arthroscopic debridement in the arthritic knee. *J Arthroplasty* 2006; 21:9-10.
123. Poehling GG. Degenerative arthritis arthroscopy and research. *Arthroscopy* 2002; 18:683-687.
124. Siparsky P, Ryzewicz M, Peterson B, Bartz R. Arthroscopic treatment of osteoarthritis of the knee: Are there any evidence-based indications? *Clin Orthop Relat Res* 2007; 455:107-112.
125. Hawker G, Guan J, Judge A, Dieppe P. Knee arthroscopy in England and Ontario: Patterns of use, changes over time, and relationship to total knee joint replacement. *J Bone Joint Surg A;* in press.
126. Kirkley A, Birmingham TB, Litchfield RB, Giffin R, Willits KR, Wong CJ, Feagan BG, Donner A, Griffin SH, D'Ascanio LM, Pope JE, Fowler PJ. A randomized trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2008: 359:1097-1107.
127. Felson DT, Buckwalter J. Debridement and lavage for osteoarthritis of the knee. *N Engl J Med* 2002; 347:132-133.
128. Veldhuyzen van Zanten SJ, Cleary C, Talley NJ, Peterson TC, Nyren O, Bradley LA, Verlinden M, Tytgat GN. Drug treatment of functional dyspepsia: A systematic analysis of trial methodology with recommendations for design of future trials. *Am J Gastroenterol* 1996; 91:660-673.
129. Talley NJ, Owen BK, Boyce P, Paterson K. Psychological treatments for irritable bowel syndrome: A critique of controlled treatment trials. *Am J Gastroenterol* 1996; 91:277-283.
130. Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol* 2000; 136:381-385.
131. Schor S, Karten I. Statistical evaluation of medical journal manuscripts. *JAMA* 1966; 195:1123-1128.
132. Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: Critical assessment of articles in BMJ from January to March 1976. *Br Med J* 1977; 1:85-87.
133. Hall JC, Hill D, Watts JM. Misuse of statistical methods in the Australasian surgical literature. *Aust N Z J Surg* 1982; 52:541-543.
134. Altman DG. Statistics in medical journals. *Stat Med* 1982; 1:59-71.
135. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987; 317:426-432.
136. Altman DG. The scandal of poor medical research. *BMJ* 1994; 308:283-284.
137. Moher D, Jones A, Lepage L, CONSORT Group (Consolidated Standards for Reporting of Trials). Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. *JAMA* 2001; 285:1992-1995.
138. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, Gaboury I. Does the CONSORT checklist improve the quality of reports of randomized controlled trials? A systematic review. *Med J Aust* 2006; 185:263-267.
139. Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT Statement by high impact factor medical journals: A survey of journal editors and journal 'Instructions to Authors'. *Trials* 2008; 9:20.
140. Prady SL, Richmond SJ, Morton VM, MacPherson H. A systematic evaluation of the impact of STRICTA and CONSORT recommendations on quality of reporting for acupuncture trials. *PLoS ONE* 2008; 3:e1577.
141. de Vet HCW, de Bie RA, van der Heijden GJMG, Verhagen AP, Sijpkes P, Kipschild PG. Systematic reviews on the basis of methodological criteria. *Physiotherapy* 1997; 83:284-289.
142. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981; 2:31-49.
143. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 1986; 4:942-951.
144. van der Heijden GJ, van der Windt DA, Kleijnen J, Koes BW, Bouter LM. Steroid injections for shoulder disorders: A systematic review of randomized clinical trials. *Brit J Gen Pract* 1996; 46:309-316.
145. Sindhu F, Carpenter L, Seers K. Development of a tool to rate the quality assessment of randomized controlled trials using a Delphi technique. *J Adv Nurs* 1997; 25:1262-1268.
146. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998; 52:377-384.
147. Reisch JS, Tyson JE, Mize SG. Aid to the evaluation of therapeutic studies. *Pediatrics* 1989; 84:815-827.
148. Seidenfeld J, Samson DJ, Aronson N, Albertson PC, Bayoumi AM, Bennett C, Brown A, Garber A, Gere M, Hasselblad V, Wilt T, Ziegler K. Relative Effectiveness and Cost-Effectiveness of Methods of Androgen Suppression in the Treatment of Advanced Prostate Cancer. Evidence Report/Technology Assessment No. 4. Rockville, MD. Agency for Health Care Policy and Research. AHCPR Publication No.99-E0012; 1999.
149. Lau J, Ioannidis J, Balk E, Milch C, Chew P, Terrin N, Lang TA, Salem D, Wong JB. Evaluating Ischemia in Emergency De-

partments: Evidence Report/Technology Assessment: No. 26. Rockville, MD. Agency for Healthcare Research and Quality. AHRQ Publication No. 01-E006 (Contract 290-97-0019 to the New England Medical Center); 2000.

150. Chestnut RM, Carney N, Maynard H, Patterson P, Mann NC, Helfand M. Rehabilitation for Traumatic Brain Injury. Evidence Report/Technology Assessment No. 2. Rockville, MD. Agency for Health Care Policy and Research. AHCPR Publication No. 99-E006; 1999.
151. Jadad AR, Boyle M, Cunningham C, Kim M, Schachar R. Treatment of Attention-Deficit/Hyperactivity Disorder. Evidence Report/Technology Assessment No. 11. Rockville, MD. Agency for Health care Research and Quality. AHRQ Publication No. 00-E005; 1999.
152. Heidenreich PA, McDonald KM, Hastie T, Fadel B, Hagan V, Lee BK, Hlatky MA. An Evaluation of Beta-Blockers, Calcium Antagonists, Nitrates, and Alternative Therapies for Stable Angina. Rockville, MD. Agency for Healthcare Research and Quality. AHRQ Publication No. 00-E003; 1999.
153. Mulrow CD, Williams JW, Trivedi M, Chiquette E, Aguilar C, Cornell JE. Treatment of Depression: Newer Pharmacotherapies. Evidence Report/Technology Assessment No. 7. Rockville, MD. Agency for Health Care Policy and Research. AHRQ Publication No. 00-E003; 1999.
154. Vickrey BG, Shekelle P, Morton S, Clark K, Pathak M, Kamberg C. Prevention and Management of Urinary Tract Infections in Paralyzed Persons. Evidence Report/Technology Assessment No. 6. Rockville, MD. Agency for Health Care Policy and Research. AHCPR Publication No. 99-E008; 1999.
155. West SL, Garbutt JC, Carey TS, Lux LJ, Jackman AM, Tolleson-Rinehart S, Lohr KN, Crews FT. Pharmacotherapy for Alcohol Dependence. Evidence Report/Technology Assessment No. 5; Rockville, MD. Agency for Health Care Policy and Research. AHCPR Publication No. 99-E004; 1999.
156. McNamara RL, Miller MR, Segal JB, Goodman SN, Kim NL, Robinson KA, Powe NR. Management of New Onset Atrial Fibrillation. Evidence Report/Technology Assessment No. 12. Rockville, MD. Agency for Health Care Policy and Research; AHRQ Publication No. 01-E026; 2001.
157. Ross S, Eston R, Chopra S, French J. Management of Newly Diagnosed Patients With Epilepsy: A Systematic Review of the Literature. Evidence Report/Technology Assessment No. 39; Rockville, MD. Agency for Healthcare Research and Quality. AHRQ Publication No. 01-E-029; 2001.
158. Goudas L, Carr DB, Bloch R, Balk E, Ioannidis JPA, Terrin N, Gialeli-Goudas M, Chew P, Lau J. Management of Cancer Pain. Evidence Report/Technology Assessment. No. 35 (Contract 290-97-0019 to the New England Medical Center). Rockville, MD. Agency for Health Care Policy and Research. AHCPR Publication No. 99-E004; 2000.
159. Nelemans P, deBie R, deVet H, Sturmans F. WITHDRAWN: Injection therapy for subacute and chronic benign low-back pain. *Cochrane Database Syst Rev* 2007; 3:CD001824.
160. Koes BW, Scholten RJ, Mens JMA, Bouter LM. Epidural steroid injections for low back pain and sciatica: An updated systematic review of randomized clinical trials. *Pain Digest* 1999; 9:241-247.
161. Manchikanti L, Rivera JJ, Pampati V, Damron KS, McManus CD, Brandon DE, Wilson SR. One day lumbar epidural adhesiolysis and hypertonic saline neurolysis in treatment of chronic low back pain: A randomized, double-blind trial. *Pain Physician* 2004; 7:177-186.
162. Manchikanti L, Boswell MV, Rivera JJ, Pampati V, Damron KS, McManus CD, Brandon DE, Wilson SR. A randomized, controlled trial of spinal endoscopic adhesiolysis in chronic refractory low back and lower extremity pain. *BMC Anesthesiol* 2005; 5:10.
163. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials. Increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003; 290:1624-1632.
164. Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998; 316:285.
165. Barbui C, Hotopf M. Forty years of antidepressant drug trials. *Acta Psychiatr Scand* 2001; 104:92-95.
166. Vesely AE, De Almeida J. Evidence-Based Medicine: Part 1: More than just the randomized controlled trial. *Univ Toronto Med J* 2002; 2:129-132.
167. Cohen MA. MAST: What went wrong? An essay on the changes in medical practice. *Pharos Alph Omega Alpha Honor Med Soc* 2000; 63:22-26.
168. Holmberg L, Baum M, Adami HO. On the scientific inference from clinical trials. *J Eval Clin Pract* 1999; 5:157-162.
169. Maisonneuve H, Ojasoo T. From the life cycles of clinical evidence to the learning curve of clinical experience. *J Eval Clin Pract* 1999; 5:417-421.
170. Knottnerus A, Dinant GJ. Medicine based evidence, a prerequisite for evidence based medicine. *BMJ* 1997; 315:1109-1110.
171. Smith BH. Evidence based medicine. Quality cannot always be quantified. *BMJ* 1995; 311:258.
172. Rosser WW. Application of evidence from randomised controlled trials to general practice. *Lancet* 1999; 353:661-664.
173. Norman GR. Examining the assumptions of evidence-based medicine. *J Eval Clin Pract* 1999; 5:139-147.
174. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352:609-613.
175. Kirk-Smith MD, Stretch DD. Evidence-based medicine and randomized double-blind clinical trials: A study of flawed implementation. *J Eval Clin Pract* 2001; 7:119-123.
176. Mant D. Can randomised trials inform clinical decisions about individual patients? *Lancet* 1999; 353: 743-743.
177. Cho MK, Bero LA. The quality of drug studies published in symposium proceedings. *Ann Int Med* 1996; 124:485-489.
178. Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine." *Am J Med* 1997; 103:529-535.
179. Djulbegovic B, Morris L, Lyman GH. Evidentiary challenges to evidence-based medicine. *J Eval Clin Pract* 2000; 6:99-109.
180. Charlton BG. Evidence based medicine. Megatrials are subordinate to medical science. *BMJ* 1995; 311:257.
181. Owens DK. Response: Does the emperor have new clothes? *Pharos Alpha Omega Alpha Honor Med Soc* 2000: 63:29-30.
182. Siderowf AD. Evidence from clinical trials: Can we do better? *NeuroRx* 2004; 1:363-371.
183. Investigational New Drug Application. Code of Federal Regulations Title 21 Part 5, pp 300-499. Revised as of 4-

1-1999. Government Printing Office, Washington, DC, 1999.

184. Sung NS, Crowley WF Jr, Genel M, Salber P, Sandy L, Sherwood LM, Johnson SB, Catanese V, Tilson H, Getz K, Larson EL, Scheinberg D, Reece EA, Slavkin H, Dobs A, Grebb J, Martinez RA, Korn A, Rimoin D. Central challenges facing the national clinical research enterprise. *JAMA* 2003; 289:1278-1287.
185. MacPherson H. Pragmatic clinical trials. *Complement Ther Med* 2004; 12:136-140.
186. Mattson RH, Cramer JA, Collins JF. A comparison of valproate with carbamazepine for the treatment of complex partial seizures and secondarily generalized tonic-clonic seizures in adults. The Department of Veterans Affairs Epilepsy Cooperative Study No. 264 Group. *N Engl J Med* 1992; 327:765-771.
187. Mohr JP, Thompson JL, Lazar RM, Levin B, Sacco RL, Furier KL, Kistler JP, Albers GW, Pettigrew LC, Adams HP Jr, Jackson CM, Pullicino P, Warfarin-Aspirin Recurrent Stroke Study Group. A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke. *N Engl J Med* 2001; 345:1444-1451.
188. McBride R, Anderson DC, Asinger RW, Newburg SM, Farmer CC, Wang K, Bundlie SR, Koller RL, Jagiella WM, Kreher S, Jorgensen CR, Sharkey SW, Flaker GC, Webel R, Nolte B, Stevenson P, Byer J, Wright W, Chesebro JH. Preliminary report of the stroke prevention in atrial fibrillation study. *N Engl J Med* 1990; 322:863-868.
189. Chesebro JH, Wiebers DO, Holland AE, Linker S, Bardsley WT, Kopecky S, Litin SC, Meissner I, Zerbe DM, Flaker GC, Webel R, Nolte B, Stevenson P, Byer J, Jenkins JS, Wright W, Anderson DC, Asinger RW, Newburg SM. Warfarin versus aspirin for prevention of thromboembolism in atrial fibrillation. Stroke Prevention in Atrial Fibrillation II Study. *Lancet* 1994; 343:687-691.
190. Vist GE, Hagen KB, Devereaux PJ, Bryant D, Kristoffersen DT, Oxman AD. Outcomes of patients who participate in randomised controlled trials compared to similar patients receiving similar interventions who do not participate. *Cochrane Database Syst Rev* 2007; MR000009.
191. King M, Nazareth I, Lamper F, Bower P, Chandler M, Morou M, Sibbald B, Lai R. Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials. *Health Technol Assess* 2005; 9:1-186.
192. Rendell JM, Merritt RD, Geddes JR. Incentives and disincentives to participation by clinicians in randomised controlled trials. *Cochrane Database Syst Rev* 2007; MR000021.
193. van Kleef M, Barendse GA, Kessels A, Voets HM, Weber WE, de Lange S. Randomized trial of radiofrequency lumbar facet denervation for chronic low back pain. *Spine* 1999; 24:1937-1942.
194. Nath S, Nath CA, Pettersson K. Percutaneous lumbar zygapophysial (facet) joint neurotomy using radiofrequency current, in the management of chronic low back pain. A randomized double-blind trial. *Spine* 2008; 33:1291-1297.
195. Kumar K, Taylor RS, Jacques L, Eldabe S, Meglio M, Molet J, Thomson S, O'Callaghan J, Eisenberg E, Milbouw G, Buchser E, Fortini G, Richardson J, North RB. Spinal cord stimulation versus conventional medical management for neuropathic pain: A multicentre randomised controlled trial in patients with failed back surgery syndrome. *Pain* 2007; 132:179-188.
196. Kemler MA, Barendse GAM, van Kleef M, deVet HC, Rijks CP, Furnée CA, van den Wildenberg FA. Spinal cord stimulation in patients with chronic reflex sympathetic dystrophy. *N Engl J Med* 2000; 343:618-624.
197. 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; 56:98-107.
198. Malmivaara A, Slätis P, Heliövaara M, Sainio P, Kinnunen H, Kankare J, Dalin-Hirvonen N, Seitsalo S, Herno A, Kortekangas P, Niinimäki T, Rönty H, Tallroth K, Turunen V, Knekt P, Härkänen T, Hurri H; Finnish Lumbar Spinal Research Group. Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial. *Spine* 2007; 32:1-8.
199. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, Abdu WA, Hilibrand AS, Boden SD, Deyo RA. Surgical vs nonoperative treatment for lumbar disk herniation: The Spine Patient Outcomes Research Trial (SPORT): A randomized trial. *JAMA* 2006; 296:2441-2450.
200. Stadhouder A, Buskens E, de Klerk LW, Verhaar JA, Dhert WA, Verbout AJ, Vaccaro AR, Oner FC. Traumatic thoracic and lumbar spinal fractures: Operative or nonoperative treatment: Comparison of two treatment strategies by means of surgeon equipoise. *Spine* 2008; 33:1006-1017.
201. Fritzell P, Wessberg P, Nordwall A: Swedish Lumbar Spine Study Group. Chronic low back pain and fusion: A comparison of three surgical techniques: A prospective multicenter randomized study from the Swedish lumbar spine study group. *Spine* 2002; 27:1131-1141.
202. Mirza SK, Deyo RA. Systematic review of randomized trials comparing lumbar fusion surgery to nonoperative care for treatment of chronic back pain. *Spine* 2007; 32:816-823.
203. Tosteson AN, Skinner JS, Tosteson TD, Lurie JD, Andersson GB, Berven S, Grove MR, Hanscom B, Blood EA, Weinstein JN. The cost effectiveness of surgical versus nonoperative treatment for lumbar disc herniation over two years: Evidence from the Spine Patient Outcomes Research Trial (SPORT). *Spine* 2008; 33:2108-2115.
204. Day SJ, Altman DG. Statistics notes: Blinding in clinical trials and other studies. *BMJ* 2000; 321:504.
205. Poolman RW, Struijs PA, Krips R, Sierevelt IN, Marti RK, Farrokhyar F, Bhandari M. Reporting of outcomes in orthopaedic randomized trials: Does blinding of outcome assessors matter? *J Bone Joint Surg Am* 2007; 89:550-558.
206. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Guyatt GH. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001; 285:2000-2003.
207. Andrew E, Anis A, Chalmers T, Cho M, Clarke M, Felson D, Gotzsche P, Greene R, Jadad A, Jonas W, Klassen T, Knipschild P, Laupacis A, Meinert CL, Moher D, Nichol G, Oxman A, Penman MF, Pocock S. A proposal for structured reporting of randomized controlled trials. *JAMA* 1994; 272:1926-1931.
208. Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: Survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996; 312:742-744.
209. DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982;

306:1332-1337.

210. Cheng K, Smyth RL, Motley J, O'Hea U, Ashby D. Randomized controlled trials in cystic fibrosis (1966-1997) categorized by time, design, and intervention. *Pediatr Pulmonol* 2000; 29:1-7.
211. Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989; 10:31-56.
212. Lang T. Masking or blinding? An unscientific survey of mostly medical journal editors on the great debate. *MedGenMed* 2000; 2:E25.
213. Prasad AS, Fitzgerald JT, Bao B, Beck FW, Chandrasekar PH. Duration of symptoms and plasma cytokine levels in patients with the common cold treated with zinc acetate. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med* 2000; 133:245-252.
214. Quitkin FM, Rabkin JG, Gerald J, Davis JM, Klein DF. Validity of clinical trials of antidepressants. *Am J Psychiatry* 2000; 157:327-337.
215. Carragee EJ, Hurwitz EL, Cheng I, Carroll LJ, Nordin M, Guzman J, Peloso P, Holm LW, Côté P, Hogg-Johnson S, van der Velde G, Cassidy JD, Haldeman S, Secretariat of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. The authors' reply to the letter to the editor by Paul Dreyfuss et al. *Eur Spine J* 2008; 17:1273-1275.
216. Lachin JM, Matts JP, Wei LJ, Randomization in clinical trials: Conclusions and recommendations. *Control Clin Trials* 1988; 9:365-374.
217. Schulz KF. Randomized controlled trials. *Clin Obstet Gynecol* 1998; 41:245-256.
218. Armitage P. The role of randomization in clinical trials. *Stat Med* 1982; 1:345-352.
219. Greenland S. Randomization, statistics, and causal inference. *Epidemiology* 1990; 1:421-429.
220. Kleijnen J, Gøtzsche P, Kunz RA, Oxman AD, Chalmers I. So what's so special about randomisation. In Maynard A, Chalmers I (eds). *Non-Random Reflections on Health Services Research: On the 25th Anniversary of Archie Cochrane's Effectiveness and Efficiency*. BMJ Publishers, London, 1997, pp 93-106.
221. Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994; 272:125-128.
222. Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990; 335:149-153.
223. Altman DG, Bland JM. How to randomise. *BMJ* 1999; 319:703-704.
224. Schulz KF. Subverting randomization in controlled trials. *JAMA* 1995; 274:1456-1458.
225. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: Chance, not choice. *Lancet* 2002; 359:515-519.
226. Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, Sackett DL, Cinà CS, Walter SD, Haynes B, Schünemann HJ, Norman GR, Guyatt GH. Need for expertise based randomised controlled trials. *BMJ* 2005; 330:88.
227. Enas GG, Enas NH, Spradlin CT, Wilson MG, Wiltse CG. Baseline comparability in clinical trials. *Drug Inf J* 1990; 24:541-548.
228. Altman DG. Randomisation. *BMJ* 1991; 302:1481-1482.
229. Hedden SL, Woolson RF, Malcolm RJ. Randomization in substance abuse clinical trials. *Subst Abuse Treat Prev Policy* 2006; 1:6.
230. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 1975; 31:103-115.
231. McEntegart DJ. The pursuit of balance using stratified and dynamic randomization techniques: An overview. *Drug Inf J* 2003; 37:293-308.
232. Pocock SJ. *Clinical Trials: A Practical Approach*. John Wiley & Sons Ltd, Chichester, 1983.
233. Treasure T, MacRae KD. Minimisation: The platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does. *BMJ* 1998; 317:362-363.
234. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Public Health* 2004; 94:416-422.
235. Cornfield J. Randomization by group: A formal analysis. *Am J Epidemiol* 1978; 108:100–102.
236. COMMIT Research Group. Community Intervention Trial for Smoking Cessation (COMMIT): 1. Cohort results from a four-year community intervention. *Am J Public Health* 1995; 85:183-192.
237. Halloran ME, Longini IM Jr, Struchiner CJ. Design and interpretation of vaccine field studies. *Epidemiol Rev* 1999; 21:73-88.
238. Weir CJ, Lees KR. Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. *Stat Med* 2003; 22:705-726.
239. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials. A review. *Control Clin Trials* 2002; 23:662-674.
240. Bracken MB. On stratification, minimization and protection against types 1 and 2 error. *J Clin Epidemiol* 2001; 54:104-105.
241. Frane JW. A method of biased coin randomization, its implementation, and its validation. *Drug Inf J* 1998; 32:423-432.
242. Taves DR. Minimization: A new method of assigning patients to treatment and control groups. *Clin Pharmacol Ther* 1974; 15:443-453.
243. Begg CB, Iglewicz B. A treatment allocation procedure for sequential clinical trials. *Biometrics* 1980; 36:81-90.
244. Rosenberger WF. Randomized play-the-winner clinical trials: Review and recommendations. *Control Clin Trials* 1999; 20:328-342.
245. Kang M, Ragan BG, Park JH. Issues in outcomes research: An overview of randomization techniques for clinical trials. *J Athl Train* 2008; 43:215-221.
246. Bandyopadhyay U, Biswas A, Bhattacharya R. A covariate adjusted two-stage allocation design for binary responses in randomized clinical trials. *Stat Med* 2007; 26:4386-4399.
247. Berger VW. Quantifying the magnitude of baseline covariate imbalances resulting from selection bias in randomized clinical trials. *Biom J* 2005; 47:119-127.
248. Kuehn BM. Industry, FDA warm to "adaptive" trials. *JAMA* 2006; 296:1955-1957.
249. Schmidt C. Adaptive design may hasten clinical trials. *J Natl Cancer Inst* 2007; 99:108-109.
250. Morita S, Sakamoto J. Application of an adaptive design to a randomized phase II selection trial in gastric cancer: A report of the study design. *Pharm Stat* 2006; 5:109-118.

251. Goldberg HI, McGough H. The ethics of ongoing randomization trials: Investigation among intimates. *Med Care* 1991; 29:JS41-JS48.
252. Abraham NS, Young JM, Solomon MJ. A systematic review of reasons for nonentry of eligible patients into surgical randomized controlled trials. *Surgery* 2006; 139:469-483.
253. King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, Sibbald B, Lai R. Impact of participant and physician intervention preferences on randomized trials: A systematic review. *JAMA* 2005; 293:1089-1099.
254. Torgerson DJ, Sibbald B. Understanding controlled trials. What is a patient preference trial? *BMJ* 1998; 316:360.
255. Torgerson DJ, Klaber-Moffett J, Russell IT. Patient preferences in randomised trials: Threat or opportunity? *J Health Serv Res Policy* 1996; 1:194-197.
256. Halpern SD. Evaluating preference effects in partially unblinded, randomized clinical trials. *J Clin Epidemiol* 2003; 56:109-115.
257. Jadad A. *Randomized Controlled Trials: A User's Guide*. BMJ Books, London, 1998.
258. Adamson J, Cockayne S, Puffer S, Torgerson DJ. Review of randomised trials using the post-randomised consent (Zelen's) design. *Contemp Clin Trials* 2006; 27:305-319.
259. Torgerson DJ, Roland M. What is Zelen's design? *BMJ* 1998; 316:606.
260. Wennberg JE, Barry MJ, Fowler FJ, Mulley A. Outcomes research, PORTS, and health care reform. *Ann NY Acad Sci* 1993; 703:52-62.
261. Pildal J, Chan AW, Hrobjartsson A, Forfang E, Altman DG, Gotzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: Cohort study. *BMJ* 2005; 330;1049.
262. Forder PM, Gebski VJ, Keech AC. Allocation concealment and blinding: When ignorance is bliss. *Med J Aust* 2005; 182:87-89.
263. Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med* 1987; 6:315-328.
264. Schulz KF, Grimes DA. Allocation concealment in randomized trials: Defending against deciphering. *Lancet* 2002; 359:614-618.
265. Pocock SJ. Statistical aspects of clinical trial design. *Statistician* 1982; 31:1-18.
266. Wittink H, Turk DC, Carr DB, Sukiennik A, Rogers W. Comparison of the redundancy, reliability, and responsiveness to change among SF-36, Oswestry Disability Index, and Multidimensional Pain Inventory. *Clin J Pain* 2004; 20:133-142.
267. Turk DC, Melzack R. The measurement of pain and the assessment of people experiencing pain. In Turk DC, Melzack R (eds). *Handbook of Pain Assessment*. 2nd ed. Guilford, New York, 2001, pp 13-14.
268. Rogers W, Wittink H, Wagner A, Cynn D, Carr DB. Assessing individual outcomes during outpatient, multidisciplinary chronic pain treatment by means of an augmented SF-36. *Pain Med* 2000; 1:44-54.
269. Fairbank J, Couper J, Davies J, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980; 66:271-273.
270. Kerns RD, Turk DC, Rudy TE. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain* 1985; 23:345-356.
271. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30:473-483.
272. Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine* 2000; 25:2940-2952.
273. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G. Outcomes measures for low back pain research: A proposal for standardized use. *Spine* 1998; 23:2003-2013.
274. Roland M, Morris R. A study of the natural history of low back pain. Part 1: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983; 8:141-144.
275. Vernon H, Mior S. The Neck Disability Index: A study of reliability and validity. *J Manipulative Physiol Ther* 1991; 14:409-415.
276. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF. Standard scales for measurement of functional outcome for cervical pain or dysfunction. *Spine* 2002; 27:515-522.
277. Stratford PW, Riddle DL, Binkley JM, Spadoni G, Westaway MD, Padfield B. Using the Neck Disability Index to make decisions concerning individual patients. *Physiother Can* 1999; 51:107-112.
278. Trouli MN, Vernon HT, Kakavelakis KN, Antonopoulou MD, Paganas AN, Lionis CD. Translation of the Neck Disability Index and validation of the Greek version in a sample of neck pain patients. *BMC Musculoskeletal Disord* 2008; 9:106.
279. Melzack R. The McGill Pain Questionnaire: Major properties and scoring methods. *Pain* 1975; 1:277-299.
280. Ware JE. The status of health assessment 1994. *Ann Rev Public Health* 1995; 16:327-354.
281. World Health Organization. *Constitution of the WHO, Basic Documents*. 1948.
282. Kuenstner S, Langelotz C, Budach V, Possinger K, Krause B, Sezer O. The comparability of quality of life scores: A multitrait multimethod analysis of the EORTC QOL-C30, SF-36 and FLIC questionnaires. *Eur J Cancer* 2002; 38:339-348.
283. Ware JE. SF-36 Health Survey Update. *Spine* 2000; 25:3130-3139.
284. Measuring and reporting pain outcomes in randomized controlled trials. Blue Cross Blue Shield Association. Technology Evaluation Center. Assessment Program. Volume 21, No. 11, October 2006.
285. Nielsen CS, Price DD, Vassend O, Stubhaug A, Harris JR. Characterizing individual differences in heat-pain sensitivity. *Pain* 2005; 119:65-74.
286. Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In Turk DC, Melzack R (eds). *Handbook of Pain Assessment,* 2nd Edition. Guilford Press, New York, 2001, pp 15-34.
287. Gracely RH, Kwilosz DM. The descriptor differential scale: Applying psychophysical principles to clinical pain assessment. *Pain* 1988; 35:279-288.
288. Leavitt F, Garron DC, Whisler WW, Sheinkop MB. Affective and sensory dimensions of back pain. *Pain* 1978; 4:273-281.
289. Schofferman J. Restoration of function: The missing link in pain medicine? *Pain Med* 2006; 7:S159-S165.
290. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures

for chronic and experimental pain. *Pain* 1983; 17:45-56.

291. Grotle M, Brox JI, Vollestad NK. Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. *Spine* 2004; 29:E492-E501.
292. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001; 94:149-158.
293. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur J Pain* 2004; 8:283-291.
294. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. *Spine* 2005; 30:1331-1334.
295. van der Roer N, Ostelo RW, Bekkering GE, van Tulder MW, de Vet HC. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine* 2006; 31:578-582.
296. Suarez-Almazor ME, Kendall C, Johnson JA, Skeith K, Vincent D. Use of health status measures in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. *Rheumatology (Oxford)* 2000; 39:783-790.
297. van der Windt DA, van der Heijden GJ, de Winter AF, Koes BW, Deville W, Bouter LM. The responsiveness of the Shoulder Disability Questionnaire. *Ann Rheum Dis* 1998; 57:82-87.
298. Dunkl PR, Taylor AG, McConnell GG, Alfano AP, Conaway MR. Responsiveness of fibromyalgia clinical trial outcome measures. *J Rheumatol* 2000; 27:2683-2691.
299. Hanley MA, Jensen MP, Ehde DM, Robinson LR, Cardenas DD, Turner JA, Smith DG. Clinically significant change in pain intensity ratings in persons with spinal cord injury or amputation. *Clin J Pain* 2006; 22:25-31.
300. Spadoni GF, Stratford PW, Solomon PE, Wishart LR. The evaluation of change in pain intensity: A comparison of the P4 and single-item numeric pain rating scales. *J Orthop Sports Phys Ther* 2004; 34:187-193.
301. Smidt N, van der Windt DA, Assendelft WJ, Mourits AJ, Deville WL, de Winter AF, Bouter LM. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. *Arch Phys Med Rehabil* 2002; 83:1145-1150.
302. Hagg O, Fritzell P, Nordwall A, Swedish Lumbar Spine Study Group. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 2003; 12:12-20.
303. Beurskens AJ, de Vet HC, Koke AJ. Responsiveness of functional status in low back pain: A comparison of different instruments. *Pain* 1996; 65:71-76.
304. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, Bombardier C, Felson D, Hochberg M, van der Heijde D, Dougados M. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: The minimal clinically important improvement. *Ann Rheum Dis* 2005; 64:29-33.
305. Crossley KM, Bennell KL, Cowan SM, Green S. Analysis of outcome measures for persons with patellofemoral pain: Which are reliable and valid? *Arch Phys Med Rehabil* 2004; 85:815-822.
306. Auleley GR, Benbouazza K, Spoorenberg A, Collantes E, Hajjaj-Hassouni N, van der Heijde D, Dougados M. Evaluation of the smallest detectable difference in outcome or process variables in ankylosing spondylitis. *Arthritis Rheum,* 2002; 47:582-587.
307. Kropmans TJ, Dijkstra PU, Stegenga B, Stewart R, de Bont LG. Repeated assessment of temporomandibular joint pain: Reasoned decision-making with use of unidimensional and multidimensional pain scales. *Clin J Pain* 2002; 18:107-115.
308. Hayes M, Patterson D. Experimental development of the graphic rating method. *Psych Bull* 1921; 18:98-99.
309. Kremer E, Atkinson JH, Ignelzi RJ. Measurement of pain: Patient preference does not confound pain measurement. *Pain* 1981; 10:241-248.
310. Bonica JJ. Evolution and current status of pain programs. *J Pain Symptom Manage* 1990; 5:368-374.
311. Good M, Stiller C, Zausniewski JA, Anderson GC, Stanton-Hicks M, Grass JA. Sensation and Distress of Pain Scales: reliability, validity, and sensitivity. *J Nurs Meas* 2001; 9:219-238.
312. Manchikanti L, Manchikanti KN, Manchukonda R, Pampati V, Cash KA. Evaluation of therapeutic thoracic medial branch block effectiveness in chronic thoracic pain: A prospective outcome study with minimum 1-year follow up. *Pain Physician* 2006; 9:97-105.
313. Manchikanti L, Manchikanti KN, Damron KS, Pampati V. Effectiveness of cervical medial branch blocks in chronic neck pain: A prospective outcome study. *Pain Physician* 2004; 7:195-201.
314. Hansen HC, McKenzie-Brown AM, Cohen SP, Swicegood JR, Colson JD, Manchikanti L. Sacroiliac joint interventions: A systematic review. *Pain Physician* 2007; 10:165-184.
315. Buenaventura RM, Shah RV, Patel V, Benyamin R, Singh V. Systematic review of discography as a diagnostic test for spinal pain: An update. *Pain Physician* 2007; 10: 147-164.
316. Trescot AM, Chopra P, Abdi S, Datta S, Schultz DM. Systematic review of effectiveness and complications of adhesiolysis in the management of chronic spinal pain: An update. *Pain Physician* 2007; 10:129-146.
317. Datta S, Everett CR, Trescot AM, Schultz DM, Adlaka R, Abdi S, Atluri SL, Smith HS, Shah RV. An updated systematic review of diagnostic utility of selective nerve root blocks. *Pain Physician* 2007; 10:113-128.
318. Seghal N, Dunbar EE, Shah RV, Colson JD. Systematic review of diagnostic utility of facet (zygapophysial) joint injections in chronic spinal pain: An update. *Pain Physician* 2007; 10:213-228.
319. Abdi S, Datta S, Trescot AM, Schultz DM, Adlaka R, Atluri SL, Smith HS, Manchikanti L. Epidural steroids in the management of chronic spinal pain: A systematic review. *Pain Physician* 2007; 10:185-212.
320. Boswell MV, Colson JD, Sehgal N, Dunbar EE, Epter R. A systematic review of therapeutic facet joint interventions in chronic spinal pain. *Pain Physician* 2007; 10:229-253.
321. Barnsley L, Lord SM, Wallis BJ, Bogduk N. The prevalence of chronic cervical zygapophyseal joint pain after whiplash. *Spine* 1995; 20:20-26.
322. Barnsley L, Lord S, Wallis B, Bogduk N. False-positive rates of cervical zygapophysial joint blocks. *Clin J Pain* 1993; 9:124-130.
323. Lord SM, Barnsley L, Wallis BJ, Bogduk N. Chronic cervical zygapophysial joint pain with whiplash: A placebo-con-

trolled prevalence study. *Spine* 1996; 21:1737-1745.

324. Manchikanti L, Singh V, Pampati V, Damron KS, Beyer CD, Barnhill RC. Is there correlation of facet joint pain in lumbar and cervical spine? An evaluation of prevalence in combined chronic low back and neck pain. *Pain Physician* 2002; 5:365-371.
325. Manchikanti L, Singh V, Rivera J, Pampati V. Prevalence of cervical facet joint pain in chronic neck pain. *Pain Physician* 2002; 5:243-249.
326. Manchikanti L, Boswell MV, Singh V, Pampati V, Damron KS, Beyer CD. Prevalence of facet joint pain in chronic spinal pain of cervical, thoracic, and lumbar regions. *BMC Musculoskelet Disord* 2004; 5:15.
327. Manchikanti L, Singh V, Pampati V, Beyer C, Damron K. Evaluation of the prevalence of facet joint pain in chronic thoracic pain. *Pain Physician* 2002; 5:354-359.
328. Manchukonda R, Manchikanti KN, Cash KA, Pampati V, Manchikanti L. Facet joint pain in chronic spinal pain: An evaluation of prevalence and false-positive rate of diagnostic blocks. *J Spinal Disord Tech* 2007; 20:539-545.
329. Manchikanti L, Manchikanti KN, Pampati V, Brandon DE, Giordano J. The prevalence of facet-joint-related chronic neck pain in postsurgical and nonpostsurgical patients: A comparative evaluation. *Pain Pract* 2008; 8:5-10.
330. Schwarzer AC, Aprill CN, Derby R, Fortin J, Kine G, Bogduk N. Clinical features of patients with pain stemming from the lumbar zygapophysial joints. Is the lumbar facet syndrome a clinical entity? *Spine* 1994; 19:1132-1137.
331. Schwarzer AC, Wang SC, Bogduk N, McNaught PJ, Laurent R. Prevalence and clinical features of lumbar zygapophysial joint pain: A study in an Australian population with chronic low back pain. *Ann Rheum Dis* 1995; 54:100-106.
332. Manchikanti L, Singh V. Review of chronic low back pain of facet joint origin. *Pain Physician* 2002; 5:83-101.
333. Manchikanti L, Pampati V, Fellows B, Bakhit C. Prevalence of lumbar facet joint pain in chronic low back pain. *Pain Physician* 1999; 2:59-64.
334. Manchikanti L, Pampati V, Fellows B, Baha AG. The inability of the clinical picture to characterize pain from facet joints. *Pain Physician* 2000; 3:158-166.
335. Manchikanti L, Singh V, Pampati V, Damron K, Barnhill R, Beyer C, Cash K. Evaluation of the relative contributions of various structures in chronic low back pain. *Pain Physician* 2001; 4:308-316.
336. Manchikanti L, Hirsch JA, Pampati V. Chronic low back pain of facet (zygapophysial) joint origin: Is there a difference based on involvement of single or multiple spinal regions? *Pain Physician* 2003; 6:399-405.
337. Manchikanti L, Manchukonda R, Pampati V, Damron KS, McManus CD. Prevalence of facet joint pain in chronic low back pain in postsurgical patients by controlled comparative local anesthetic blocks. *Arch Phys Med Rehabil* 2007; 88:449-455.
338. Pauza KJ, Howell S, Dreyfuss P. A randomized, placebo-controlled trial of intradiscal electrothermal therapy for the treatment of discogenic low back pain. *Spine J* 2004; 4:27-35.
339. Turk DC. Statistical significance and clinical significance are not synonyms! *Clin J Pain* 2000; 16:185-187.
340. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. *Pain* 2000; 88:287-294.
341. Middel B, Stewart R, Bouma J, van Sonderen E, van den Heuvel W. How to validate clinically important change in health-related functional status. Is the magnitude of the effect size consistently related to magnitude of change as indicated by a global question rating? *J Eval Clin Pract* 2001;7:399-410.
342. Wyrwich KW, Wolinsky FD. Identifying meaningful intraindividual change standards for health-related quality of life measures. *J Eval Clin Pract* 2000;6:39-49.
343. Wyrwich K, Nienaber N, Tierney W, Wolinsky F. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999;37:469-478.
344. Mikail SF, DuBreuil S, D'Eon JL. A comparative analysis of measures used in the assessment of chronic pain patients. *Psychol Assess* 1993; 5:117-120.
345. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiat* 1961; 4:561-571.
346. Epker J, Gatchel RJ. Prediction of treatment-seeking behavior in acute TMD patients: Practical application in clinical settings. *J Orofac Pain* 2000; 14:303-309.
347. Olsson I, Bunketorp O, Carlsson SG, Styf J. Prediction of outcome in whiplash-associated disorders using West Haven-Yale Multidimensional Pain Inventory. *Clin J Pain* 2002;18:238-244.
348. Burton HJ, Kline SA, Hargadon R, Cooper BS, Schick R, Ong M. Assessing patients with chronic pain using the basic personality inventory as a complement to the multidimensional pain inventory. *Pain Res Manag* 1999; 4:121-129.
349. Turk DC. Clinical effectiveness and cost-effectiveness of treatments for patients with chronic pain. *Clin J Pain* 2002; 18:355-365.
350. 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.
351. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: Reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol* 1997; 50:79-93.
352. Essink-Bot ML, Krabbe PF, Bonsel GJ, Aaronson NK. An empirical comparison of four generic health status measures: The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. *Med Care* 1997; 35:522-537.
353. Bronfort G, Bouter LM. Responsiveness of general health status in chronic low back pain: A comparison of the COOP charts and the SF-36. *Pain* 1999; 83:201-209.
354. Kosinski M, Keller SD, Ware JE Jr., Hatoum HT, Kong SX. The SF-36 Health Survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: Relative validity of scales in relation to clinical measures of arthritis severity. *Med Care* 1999; 37:MS23-MS39.
355. Roos EM, Klassbo M, Lohmander LS.

WOMAC osteoarthritis index. Reliability, validity, and responsiveness in patients with arthroscopically assessed osteoarthritis. Western Ontario and MacMaster Universities. *Scand J Rheumatol* 1999; 28:210-215.

356. Smith BH, Penny KI, Purves AM, Munro C, Wilson B, Grimshaw J, Chambers WA, Smith WC. The Chronic Pain Grade questionnaire: Validation and reliability in postal research. *Pain* 1997; 71:141-147.
357. Elliott AM, Smith BH, Smith WC, Chambers WA. Changes in chronic pain severity over time: The Chronic Pain Grade as a valid measure. *Pain* 2000; 88:303-308.
358. Haas M, Nyiendo J. Diagnostic utility of the McGill Pain Questionnaire and the Oswestry Disability Questionnaire for classification of low back syndromes. *J Manipulative Physiol Ther* 1992; 15:90-98.
359. Gronblad M, Hupli M, Wennerstrand P, Jarvinen E, Lukinmaa A, Kouri JP, Karaharju EO. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. *Clin J Pain* 1993; 9:189-195.
360. Salen BA, Spangfort EV, Nygren AL, Nordemar R. The Disability Rating Index: An instrument for the assessment of disability in clinical settings. *J Clin Epidemiol* 1994; 47:1423-1435.
361. Strong J, Ashton R, Large RG. Function and the patient with chronic low back pain. *Clin J Pain* 1994; 10:191-196.
362. Greenough CG. Recovery from low back-pain: 1-5 year follow-up of 287 injury-related cases. *Acta Orthop Scand Suppl* 1993; 254:1-34.
363. Kaplan GM, Wurtele SK, Gillis D. Maximal effort during functional capacity evaluations: An examination of psychological factors. *Arch Phys Med Rehabil* 1996; 77:161-164.
364. Grevitt M, Khazim R, Webb J, Mulholland R, Shepperd J. The short form-36 health survey questionnaire in spine surgery. *J Bone Joint Surg Br* 1997; 79:48-52.
365. Loisel P, Poitras S, Lemaire J, Durand P, Southiere A, Abenhaim L. Is work status of low back pain patients best described by an automated device or by a questionnaire? *Spine* 1998; 23:1588-1594.
366. Nordin M, Skovron ML, Hiebert R, Weiser S, Brisson PM, Campello M, Harwood K, Crane M, Lewis S. Early predictors of delayed return to work in patients with low back pain. *J Musculoskelet Pain* 1997; 5:5-27.
367. Kaivanto KK, Estlander AM, Moneta GB, Vanharanta H. Isokinetic performance in low back-pain patients: The predictive power of the Self-Efficacy Scale. *J Occup Rehabil* 1995; 5:87-99.
368. Kuukkanen T, Malkia E. Muscular performance after a 3 month progressive physical exercise program and 9 month follow-up in subjects with low back pain. A controlled study. *Scand J Med Sci Sports* 1996; 6:112-121.
369. Fisher K, Johnson M. Validation of the Oswestry Low Back Pain Disability Questionnaire, its sensitivity as a measure of change following treatment and its relationship with other aspects of the chronic pain experience. *Physiother Theory Pract* 1997; 13:67-80.
370. Sufka A, Hauger B, Trenary M, Bishop B, Hagen A, Lozon R, Martens B. Centralization of low back pain and perceived functional outcome. *J Orthop Sports Phys Ther* 1998; 27:205-212.
371. Gronblad M, Jarvinen E, Hurri H, Hupli M, Karaharju EO. Relationship of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) with three dynamic physical tests in a group of patients with chronic low-back and leg pain. *Clin J Pain* 1994; 10:197-203.
372. Gronblad M, Hurri H, Kouri JP. Relationships between spinal mobility, physical performance tests, pain intensity and disability assessments in chronic low back pain patients. *Scand J Rehabil Med* 1997; 29:17-24.
373. Little D, MacDonald D. The use of the percentage change in Oswestry Disability Index Score as an outcome measure in lumbar spinal surgery. *Spine* 1994; 19:2139-2143.
374. Luoto S, Hupli M, Alaranta H, Hurri H. Isokinetic performance capacity trunk muscles. Part II. Coefficient of variation in isokinetic measurement in maximal effort and in submaximal effort. *Scand J Rehabil Med* 1996; 28:207-210.
375. Luoto S, Taimela S, Hurri H, Aalto H, Pyykko I, Alaranta H. Psychomotor speed and postural control in chronic low back pain patients: A controlled follow-up study. *Spine* 1996; 21:2621-2627.
376. Breitenseher MJ, Eyb RP, Matzner MP, Trattnig S, Kainberger FM, Imhof H. MRI of unfused lumbar segments after spondylodesis. *J Compt Assist Tomogr* 1996; 20:583-587.
377. Sanderson P, Todd B, Holt G, Getty CJ. Compensation, work status, disability in low back pain patients. *Spine* 1995; 20:554-556.
378. Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine* 2000; 25:3115-3124.
379. Bergner M, Bobbitt R, Carter W, Gilson BS. The Sickness Impact Profile: Development and final revision of a health status measure. *Med Care* 1981; 19:787-805.
380. Deyo R, Centor R. Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. *J Chronic Dis* 1986; 39:897-906.
381. Patrick D, Deyo R, Atlas S, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. *Spine* 1995; 20:1899-1908.
382. Jensen MP, Strom SE, Turner JA, Romano JM. Validity of the Sickness Impact Profile Roland Scale as a measure of dysfunction in chronic pain patients. *Pain* 1992; 50:157-162.
383. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, Williams JI. The Quebec Back Pain Disability Scale: Conceptualization and development. *J Clin Epidemiol* 1996; 49:151-161.
384. Leclaire R, Blier F, Fortin L, Proulx R. A cross-sectional study comparing the Oswestry and Roland-Morris functional disability scales in two populations of patients with low back pain of different levels of severity. *Spine* 1997; 22:68-71.
385. Stratford PW, Binkley J, Solomon P, Gill C, Finch E. Assessing change over time in patients with low-back-pain. *Phys Ther* 1994; 74:528-533.
386. Deyo RA. Measuring the functional status of patients with low-back pain. *Arch Phys Med Rehabil* 1988; 69:1044-1053.
387. Simmonds M, Olson S, Jones S, Hussein T, Lee CE, Novy D, Radwan H. Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. *Spine* 1998; 23:2412-2421.
388. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of

the Roland-Morris Back Pain Questionnaire: Part 1. *Phys Ther* 1998; 78:1186-1196.

389. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, William JI. The Quebec Back Pain Disability Scale: Measurement properties. *Spine* 1995; 20:341-352.
390. Co YY, Eaton S, Maxwell MW. The relationship between the St. Thomas and Oswestry disability scores and the severity of low back pain. *J Manipulative Physiol Ther* 1993; 16:14-18.
391. Kopec J, Esdaile J. Functional disability scales for back pain. *Spine* 1995; 20:1943-1949.
392. Beurskens A, de Vet H, Koke A, van der Heijden GJ, Knipschild PG. Measuring functional status of patients with low back pain: Assessment of the quality of four disease specific questionnaires. *Spine* 1995; 20:1017-1728.
393. Frost H, Lamb SE, Stewart-Brown S. Responsiveness of a patient specific outcome measure compared with the Oswestry Disability Index V2.1 and Roland and Morris Disability Questionnaire for patients with subacute and chronic low back pain. *Spine* 2008; 33:2450-2457.
394. Wloduka-Demaille S, Poiraudeau S, Catanzariti JF, Rannou F, Fermanian J, Reve M. French translation and validation of 3 functional disability scales for neck pain. *Arch Phys Med Rehabil* 2002; 83:376-382.
395. Ackelman B, Lindgren U. Validity and reliability of a modified version of the Neck Disability Index. *J Rehabil Med* 2002; 34:284-287.
396. Vos CJ, Verhagen AP, Koes BW. Reliability and responsiveness of the Dutch version of the Neck Disability Index in patients with acute neck pain in general practice. *Eur Spine J* 2006; 11:1729-1736.
397. Cook C, Richardson JK, Braga L, Menezes A, Soler X, Kume P, Zaninelli M, Sokolows F, Pietrobon R. Cross-cultural adaptation and validation of the Brazilian Portuguese version of the Neck Disability Index and Neck Pain and Disability Scale. *Spine* 2006; 31:1621-1627.
398. Cleland JA, Childs JD, Whitman JM. Psychometric properties of the Neck Disability Index and Numeric Pain Rating Scale in patients with mechanical neck pain. *Arch Phys Med Rehabil* 2008; 89:69-74.
399. Riddle D, Stratford P. Use of generic versus region-specific functional status measures on patients with cervical spine disorders. *Phys Ther* 1998; 78:951-963.
400. Jette DU, Jette AM. Physical therapy and health outcomes in patients with spinal impairments. *Phys Ther* 1996; 76:930-941.
401. Cleland JA, Fritz JM, Whitman JM, Palmer JA. The reliability and construct validity of the Neck Disability Index and Patient Specific Functional Scale in patients with cervical radiculopathy. *Spine* 2006; 31:598-602.
402. Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC. Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. *Spine* 2007; 32:3047-3051.
403. Westaway M, Stratford P, Binkley J. The Patient-Specific Functional Scale: Validation of its use in persons with neck dysfunction. *J Orthop Sports Phys Ther* 1998; 27:331-338.
404. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10:407-415.
405. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003; 56:395-407.
406. Bombardier C, Hayden J, Beaton DE. Minimal clinically important difference: Low back pain. Outcome measures. *J Rheumatol* 2001; 28:431-438.
407. Lee JS, Hobden E, Stiell IG, Wells GA. Clinically important change in the visual analog scale after adequate pain control. *Acad Emerg Med* 2003; 10:1128-1130.
408. Wells G, Beaton D, Shea B, Boers M, Simon L, Strand V, Brooks P, Tugwell P. Minimal clinically important differences: Review of methods. *J Rheumatol* 2001; 28:406-412.
409. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): A literature review and directions for future research. *Curr Opin Rheumatol* 2002; 14:109-114.
410. Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part 2. *Phys Ther* 1998; 78:1197-1207.
411. Goodman SN. Toward evidence-based medical statistics. 1: The *P* value fallacy. *Ann Intern Med* 1999; 130:995-1004.
412. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994; 69:979-985.
413. Tannock IF. False-positive results in clinical trials: Multiple significance tests and the problem of unreported comparisons. *J Natl Cancer Inst* 1996; 88:206-207.
414. Streiner D, Norman G. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press, New York, 1995.
415. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ* 1995; 311:1145-1148.
416. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; 311:485.
417. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994; 272:122-124.
418. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med* 1978; 299:690-694.
419. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; 3:409-422.
420. Eng J. Sample size estimation: How many individuals should be studied? *Radiology* 2003; 227:309-313.
421. Lerman J. Study design in clinical research: Sample size estimation and power analysis. *Can J Anaesth* 1996; 43:184-191.
422. Woodward M. Formulas for sample-size, power and minimum detectable relative risk in medical studies. *Statistician* 1992; 41:185-196.
423. Browner WS, Newman TB, Cummings SR, Hulley SB. Estimating sample size and power. In Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB (eds). *Designing Clinical Research: An Epidemiologic Approach*, 2nd ed. Lippincott, Williams & Wilkins, Philadelphia, 2001, pp 65-84.
424. Frison L, Pocock S. Repeated measurements in clinical trials: Analysis using mean summary statistics and its implications for design. *Stat Med* 1992;

11:1685-1704.

425. Rosner B. Estimation of sample size and power for comparing two means. In *Fundamentals of Biostatistics*. 5th ed. Duxbury Press, Pacific Grove, 2000, pp 307-308.
426. Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. Wiley-Interscience, New York, 1981.
427. Pagano M, Gauvreau K. *Principles of Biostatistics*. 2nd ed. Duxbury Press, Pacific Grove, 2000.
428. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Associates, Hillsdale, 1988.
429. Lang TA, Secic M. *How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers*. American College of Physicians (ACP) Press, Philadelphia, 1997.
430. Whitley E, Ball J. Statistics review 4: Sample size calculations. *Crit Care* 2002; 6:335-341.
431. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. In Altman DG, Machin D, Bryant TN, Gardner MJ (eds). *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. 2nd edition. BMJ Books, London, 2000, pp 171-190.
432. Altman DG, Bland JM. Statistics notes. Units of analysis. *BMJ* 1997; 314:1874.
433. Bolton S. Independence and statistical inference in clinical trial designs: A tutorial review. *J Clin Pharmacol* 1998; 38:408-412.
434. Greenland S. Principles of multilevel modelling. *Int J Epidemiol* 2000; 29:158-167.
435. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine – reporting of subgroup analyses in clinical trials. *N Engl J Med* 2007; 357:2189-2194.
436. Rothwell PM. Treating individuals 2. Subgroup analysis in randomized controlled trials: Importance, indications, and interpretation. *Lancet* 2005; 365:176-186.
437. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991; 266:93-98.
438. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; 355:1064-1069.
439. Pocock SJ, Assmann SF, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: Current practice and problems. *Stat Med* 2002; 21:2917-2930.
440. Hernández A, Boersma E, Murray G, Habbema J, Steyerberg E. Subgroup analyses in therapeutic cardiovascular clinical trials: Are most of them misleading? *Am Heart J* 2006; 151:257-264.
441. Parker AB, Naylor CD. Subgroups, treatment effects, and baseline risks: Some lessons from major cardiovascular trials. *Am Heart J* 2000; 139:952-961.
442. Lagakos SW. The challenge of subgroup analyses – reporting without distorting. *N Engl J Med* 2006; 354:1667-1669.
443. Halperin M, Ware JH, Byar DP. Testing for interaction in an IxJxK contingency table. *Biometrika* 1977; 64:271-275.
444. Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* 1985; 41:361-372.
445. van Wijk RM, Geurts JW, Wynne HJ, Hammink E, Buskens E, Lousberg R, Knape JT, Groen GJ. Radiofrequency denervation of lumbar facet joints in the treatment of chronic low back pain: A randomized, double-blind, sham lesion-controlled trial. *Clin J Pain* 2005; 21:335-344.
446. van Wijk RM, Geurts JW, Lousberg R, Wynne HJ, Hammink E, Knape JT, Groen GJ. Psychological predictors of substantial pain reduction after minimally invasive radiofrequency and injection treatments for chronic low back pain. *Pain Med* 2008; 9:212-221.
447. Manchikanti L, Cash KA, Pampati V, Fellows B. Influence of psychological variables on the diagnosis of facet joint involvement in chronic spinal pain. *Pain Physician* 2008; 11:145-160.
448. Manchikanti L, Manchikanti K, Cash KA, Singh V, Giordano J. Age-related prevalence of facet joint involvement in chronic neck and low back pain. *Pain Physician* 2008; 11:67-75.
449. Vickers AJ. Parametric versus nonparametric statistics in the analysis of randomized trials with non-normally distributed data. *BMC Med Res Methodol* 2005; 5:35.
450. Altman DG. *Practical Statistics for Medical Research*. Chapman and Hall, London, 1991.
451. Jekel JF, Katz DL, Elmore JG. *Epidemiology, Biostatistics and Preventive Medicine* WB Saunders Company, Philadelphia, 2001.
452. Heeren T, D'Agostino R. Robustness of the two independent samples t-test when applied to ordinal scaled data. *Stat Med* 1987; 6:79-90.
453. Sawilowsky SS. Comments on using alternative to normal theory statistics in social and behavioural science. *Canadian Psychology* 1993; 34:432-439.
454. Zimmerman DW, Zumbo BD. The effect of outliers on the relative power of parametric and nonparametric statistical tests. *Perceptual Mot Skills* 1990; 71:339-349.
455. Sawilowsky SS, Blair RC. A more realistic look at the robustness and Type II error properties of the t-test to departures from population normality. *Psychological Bulletin* 1992; 111:352-360.
456. Bridge PD, Sawilowsky SS. Increasing physicians' awareness of the impact of statistics on research outcomes: Comparative power of the t-test and Wilcoxon Rank-Sum test in small samples applied research. *J Clin Epidemiol* 1999; 52:229-235.
457. Senn S. *Statistical Issues in Drug Development*. John Wiley & Sons Ltd, Chichester, 1997.
458. Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: A simulation study. *BMC Med Res Methodol* 2001; 1:6.
459. Kalish LA, Begg CB. Treatment allocation methods in clinical trials: A review. *Stat Med* 1985; 4:129-144.
460. Fisher R. *Statistical Methods and Scientific Inference*. 3rd ed. Macmillan, New York, 1973.
461. Browner W, Newman T. Are all significant *P* values created equal? The analogy between diagnostic tests and clinical research. *JAMA* 1987; 257:2459-2463.
462. Diamond GA, Forrester JS. Clinical trials and statistical verdicts: Probable grounds for appeal. *Ann Intern Med* 1983; 98:385-394.
463. Lilford RJ, Braunholtz D. The statistical basis of public policy: A paradigm shift

is overdue. *BMJ* 1996; 313:603-607.

464. Freeman PR. The role of *p*-values in analysing trial results. *Stat Med* 1993; 12:1443-1552.
465. Berkson J. Tests of significance considered as evidence. Journal of the American Statistical Association 1942; 37:325-335. *Int J Epidemiol* 2003; 32:687-691.
466. Pearson E. Student as a statistician. *Biometrika* 1938; 38:210-250.
467. Altman DG. Confidence intervals in research evaluation. *Ann Intern Med* 1992; 116:A28-A29.
468. Berry G. Statistical significance and confidence intervals. *Med J Aust* 1986; 144:618-619.
469. Braitman LE. Confidence intervals extract clinically useful information from data. *Ann Intern Med* 1988; 108:296-298.
470. Simon R. Confidence intervals for reporting results of clinical trials. *Ann Intern Med* 1986; 105:429-435.
471. Godwin M. Hypothesis: The research page. Part 3: Power, sample size, and clinical significance. *Can Fam Physician* 2001; 47:1441-1443.
472. Altman DG, Bland J. Improving doctors' understanding of statistics. *J R Statis Soc* 1991; 154:223-267.
473. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med* 1993; 118:201-210.
474. Altman DG, Goodman SN. Transfer of technology from statistical journals to the biomedical literature. Past trends and future predictions. *JAMA* 1994; 272:129-132.
475. Feinstein AR. *P*-values and confidence intervals: Two sides of the same unsatisfactory coin. *J Clin Epidemiol* 1998; 51:355-360.
476. Neyman J, Pearson E. On the problem of the most efficient tests of statistical hypotheses. *Philos Trans R Soc, Series A* 1933; 231:289-337.
477. Stallones RA. The use and abuse of subgroup analysis in epidemiological research. *Prev Med* 1987; 16:183-194.
478. Pocock SJ, Hughes MD. Estimation issues in clinical trials and overviews. *Stat Med* 1990; 9:657-671.
479. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: Quantifying the risks of false-positives and false-negatives. *Health Technol Assess* 2001; 5:1-56.
480. Horwitz RI, Singer BH, Makuch RW, Viscoli CM. Clinical versus statistical considerations in the design and analysis of clinical research. *J Clin Epidemiol* 1998; 51:305-307.
481. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004; 57:229-236.
482. Cook DI, Gebski VJ, Keech AC. Subgroup analysis in clinical trials. *Med J Aust* 2004; 180:289-291.
483. Matthews JNS, Altman DG. Interaction 2: Compare effect sizes not *P* values. *BMJ* 1996; 313:808.
484. Altman DG, Bland JM. Interaction revisited: The difference between two estimates. *BMJ* 2003; 326:219.
485. Wellek S. *Testing Statistical Hypotheses of Equivalence*. Chapman Hall/CRC Press, Boca Raton, 2003.
486. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med* 1990; 113:69-76.
487. Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. Quality of nonstructured and structured abstracts of original research articles in the *British Medical Journal*, the *Canadian Medical Association Journal* and the *Journal of the American Medical Association*. *CMAJ* 1994; 150:1611-1615.
488. Hartley J, Sydes M, Blurton A. Obtaining information accurately and quickly: Are structured abstracts more efficient? *Journal of Information Science* 1996; 22:349-356.
489. World Medical Association declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. *JAMA* 1997; 277:925-926.
490. Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 1996; 313:1390-1393.
491. Rodgers A, MacMahon S. Systematic underestimation of treatment effects as a result of diagnostic test inaccuracy: Implications for the interpretation and design of thromboprophylaxis trials. *Thromb Haemost* 1995; 73:167-171.
492. Fuks A, Weijer C, Freedman B, Shapiro S, Skrutkowska M, Riaz A. A study in contrasts: Eligibility criteria in a twenty-year sample of NSABP and POG clinical trials. National Surgical Adjuvant Breast and Bowel Program. Pediatric Oncology Group. *J Clin Epidemiol* 1998; 51:69-79.
493. Roberts C. The implications of variation in outcome between health professionals for the design and analysis of randomized controlled trials. *Stat Med* 1999; 18:2605-2615.
494. Rothmann M, Li N, Chen G, Chi GY, Temple R, Tsou HH. Design and analysis of non-inferiority mortality trials in oncology. *Stat Med* 2003; 22:239-264.
495. Egger M, Jüni P, Bartlett C, CONSORT Group (Consolidated Standards of Reporting of Trials). Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001; 285:1996-1999.
496. Shuster JJ. Median follow-up in clinical trials. *J Clin Oncol* 1991; 9:191-192.
497. Altman DG, De Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. *Br J Cancer* 1995; 72:511-518.
498. Pocock SJ. When to stop a clinical trial. *BMJ* 1992; 305:235-240.
499. Senn SJ. Base logic: Tests of baseline balance in randomized clinical trials. *Clin Res Regul Aff* 1995; 12:171-182.
500. Altman DG. Adjustment for covariate imbalance. In: Armitage P, Colton T (eds). *Encyclopedia of Biostatistics*. John Wiley & Sons, Chichester, 1998, pp 1000-1005.
501. Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther* 1995; 57:6-15.
502. Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Stat Med* 2000; 19:1849-1864.
503. Ruiz-Canela M, Martınez-Gonzalez MA, de Irala-Estevez J. Intention to treat analysis is related to methodological quality. *BMJ* 2000; 320:1007-1008.
504. Armitage P, Colton T. *Encyclopedia of Biostatistics*, John Wiley & Sons, New York, 1998.
505. Committee for Proprietary Medicinal Products (CPMP). Points to Consider on Missing Data: The European Agency for the Evaluation of Medicinal Products. London, 2001.

506. Altman DG. Confidence intervals in practice. In Altman DG, Machin D, Bryant TN, Gardner MJ (eds). *Statistics with Confidence: Confidence Intervals and Statistical Guidelines,* 2nd ed. BMJ Books, London, 2000, pp 6-14.
507. Altman DG. Clinical trials and meta-analyses. In Altman DG, Machin D, Bryant TN, Gardner MJ (eds). *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*, 2nd ed. BMJ Books, London, 2000 pp 120-138.
508. Barron JP. Uniform requirements for manuscripts submitted to biomedical journals recommended by the International Committee of Medical Journal Editors. *Chest* 2006; 129:1098-1099.
509. Gardner MJ, Altman DG. Confidence intervals rather than *P* values: Estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)* 1986; 292:746-750.
510. Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication updated October 2004. *Mymensingh Med J* 2005; 14:95-119.
511. Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *JAMA* 1997; 277:927-934.
512. Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *N Engl J Med* 1997; 336:309-315.
513. Uniform requirements for manuscripts submitted to biomedical journals. Preface. *Lancet* 1979; 1:428-430.
514. Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Applied Statistic* 2000; 49:359-370.
515. Egger M, Dickersin K, Davey Smith G. Problems and limitations in conducting systematic reviews. In Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-Analysis in Context*. BMJ Publishing Group, London, 2001, pp 43-68.
516. Cook RJ, Sackett DL. The number needed to treat: A clinically useful measure of treatment effect. *BMJ* 1995; 310:452-454.
517. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999; 319:1492-1495.
518. Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Stat Med* 2000; 19:3325-3336.
519. Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials. An evaluation of 7 medical areas. *JAMA* 2001; 285:437-443.
520. Horton R. The rhetoric of research. *BMJ* 1995; 310:985-987.
521. *Annals of Internal Medicine*. Information for authors. Available at www.annals.org.
522. Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ* 1999; 318:1224-1225.
523. Purcell GP, Donovan SL, Davidoff F. Changes to manuscripts during the editorial process: Characterizing the evolution of a clinical paper. *JAMA* 1998; 280:227-228.
524. Kiviluoto T, Sire´n J, Luukkonen P, Kivilaakso E. Randomised trial of laparoscopic versus open cholecystectomy for acute and gangrenous cholecystitis. *Lancet* 1998; 351:321-325.
525. Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychol Bull* 1957; 54:297-312.
526. McAlister FA. Applying the results of systematic reviews at the bedside. In Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-Analysis in Context*. BMJ Books, London, 2001, pp 373-385.
527. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988; 318:1728-1733.