Pain Physician 2009; 12:73-108 • ISSN 1533-3159

Evidence-Based Medicine

Evidence-Based Medicine, Systematic Reviews, and Guidelines in Interventional Pain Management: Part 4: Observational Studies

Laxmaiah Manchikanti, MD[1], Vijay Singh, MD[2], Howard S. Smith, MD[3], and Joshua A. Hirsch, MD[4]

From: [1]Pain Management Center of Paducah, Paducah, KY; [2]Pain Diagnostic Associates, Niagara WI; [3]Albany Medical College, Albany, NY; [4]Massachusetts General Hospital and Harvard Medical School, Boston, MA.

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, Louisville, KY.
Dr. Singh is Medical Director, Pain Diagnostic Associates, Niagara, WI.
Dr. Smith is Associate Professor and Academic Director of Pain Management for Albany Medical College Department of Anesthesiology, Albany, NY.
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.

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

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

Manuscript received: 12/18/2008
Accepted for publication: 12/29/2008

Free full manuscript:
www.painphysicianjournal.com

Evidence-based medicine (EBM) stresses the examination of evidence from clinical research and describes it as a shift in medical paradigms, in contrast to intuition, unsystematic clinical experience, and pathophysiologic rationale. While the importance of randomized trials has been created by the concept of the hierarchy of evidence in guiding therapy, much of the medical research is observational. There is competition, contrast, and a feeling of inferiority and uselessness for observational studies, created by a lack of understanding of medical research. However, observational studies and randomized clinical trials (RCTs) can be viewed as the steps of observation and experimentation that form the basis of the scientific methodology. Further, rational healthcare practices require knowledge about the etiology and pathogenesis, diagnosis, prognosis, and treatment of disorders.

The reporting of observational research is often not detailed and clear enough with insufficient quality and poor reporting, which hampers the assessment of strengths and weaknesses of the study and the generalizability of the mixed results. Thus, design, implementation, and reporting of observational studies is crucial. The biased interpretation of results from observational studies, either in favor of or opposed to a treatment, and lack of proper understanding of observational studies, leads to a poor appraisal of the quality. Similar to the Consolidated Standards of Reporting Trials (CONSORT) statement for the reporting of randomized trials, the Strengthening of the Reporting of Observational Studies in Epidemiology (STROBE) statement was developed with recommendations to improve the quality of reporting observational studies. The STROBE statement consists of a checklist of 22 items, which relate to the title, abstract, introduction, methods, results, and discussion sections of articles.

Multiple types of observational studies are conducted; however, 3 types have been highlighted in the STROBE document and also in the present review, which include cohort studies, case-controlled studies, and cross-sectional studies.

This comprehensive review provides an introduction and rationale, types, design, and reporting of observational studies; outcomes assessment and data presentation and analysis; statistical analysis, results, and a discussion of observational studies.

Key words: Observational studies, cohort studies, case control studies, cross-sectional studies, allocation bias, sample size, Strengthening of the Reporting of Observational Studies in Epidemiology (STROBE)

Pain Physician 2009; 12:1:73-108

Evidence-based medicine (EBM) is described a shift in medical paradigms and about solving clinical problems (1-3). This is in contrast to the traditional paradigm of medical practice, which is based on intuition, unsystematic clinical experience, and pathophysiologic rationale, which are considered insufficient grounds for clinical decision-making. EBM stresses the examination of evidence from clinical research. Further, rational health care practices require knowledge about the etiology and pathogenesis, diagnosis, prognosis, and treatment of diseases (4). Randomized trials provide valuable evidence about treatments and other interventions. However, most of the research in clinical practice comes from observational studies (4,5). Randomized trials work by first assuming there is no difference between a new and an old or placebo treatment to prove the null hypothesis (6). Basically, it may be described that standard randomized controlled trials (RCTs) are in fact set up to show that treatments do not work, rather than to demonstrate that treatments do work (6). The 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.

Based on the lack of understanding or politics, EBM has been characterized as a stick by which policy-makers and academicians beat clinicians (7-11). In addition, it has been alleged that the research performed to test new treatments has often been of poor quality, leading to clinicians criticizing the research establishment for failing to provide answers to relevant clinical problems of everyday practice (12). Most questions in medical research are investigated in observational studies (4,5,13-17). Consequently, observational studies are more likely to provide an indication of daily medical practices (18). Thus, the proponents of observational studies describe that observational studies are just as effective as RCTs. However, from a methodologic perspective, the 2 types of studies are considered complementary rather than opposing (17). Thus, observational studies and RCTs can be viewed as expressions in the setting of modern clinical research of the steps of observation and experimentation that form the basis of the scientific methodology. Greene (17) describes that the observational step is used to uncover patterns and formulate hypothesis regarding cause-and-effect relationships, which is followed by the experimentation step in which the hypothesis formed in the observational setting are confirmed or refuted in an experiment in which the independent variables are controlled by the experimenter (19,20). Consequently, the argument about one or other evidence is misplaced since both observation and experimentation steps are required for scientific advancement (17). In fact, Guyatt and Drummond (1) in a description of the hierarchy of strength of evidence for treatment decisions provide significant strength to systematic reviews of observational studies and single observational studies.

Reporting of observational research is often not detailed and clear enough to assess the strengths and weaknesses of the investigations (4,5,19,20). Research should be reported transparently so that readers can follow what was planned, what was done, what was found, and what conclusions were drawn (5). The credibility of research is based on assessment of the strengths and weaknesses in the study design, conduct, and transparent reporting (21-37).

This review is intended to describe the nature of observational studies and their importance; differences from randomized trials; design, implementation, and reporting of observational studies (4,5,38).

1. An Introduction to Observational Studies

An observational study is defined as an etiologic or effectiveness study, a cross-sectional study, a case series, a case-control design, a design with historical controls, or a cohort design (27,39). These designs have long been used in the evaluation of exposures that may cause disease or injury (40). In addition, studies of risk factors generally cannot be randomized because they relate to inherent human characteristics or practices, and exposing subjects to harmful risk factors is unethical (41). Further, design of diagnostic studies is based on non-randomization (42-46). At times, clinical data may be summarized in order to design a randomized comparison (47). Observational data may also be needed to assess the effectiveness of an intervention in a community as opposed to the special setting of a control trial (48).

2. Why Observational Studies?

The World Health Organization (WHO) defines clinical trials 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" (49). Thus, 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. However, this is not in any way limited to randomized, double-blind, placebo-controlled trials. The majority of the studies in interventional pain management are observational (50-71) and treatments even in surgery are more likely to be based on observational studies compared to internal medicine, which are based on RCTs (72-75). Further, many studies have been ruled observational based on a lack of understanding of the design of the study (active control trials without a placebo) (76-85).

3. How Do Observational Studies Differ from Randomized Trials?

The basis for randomized trials arises from the evidence that many surgical and medical interventions recommended based on observational studies have later been demonstrated to be ineffective or even harmful (86-90). However, there also has been contradictory evidence demonstrated for RCTs (38,91). Further, not all questions can be addressed in an RCT and evidence shows that only 40% of treatment questions involving surgical procedures are amenable to evaluation by an RCT, even in an ideal clinical setting (92-95). 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 as shown in Table 1 (96). In placebo-controlled trials, multiple effects can occur to distort the results not only limited to placebo or Hawthorne effect (97). 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 (98-103). In addition, in so-called placebo-controlled trials, specifically in evaluation of interventional techniques, researchers and practitioners are not aware of the effects of solutions injected into closed spaces, joints, and over the nerve roots (76-85,104-112). Further, multiple authors have considered local anesthetic injection as placebo (85,106,107,110), even though evidence has been contrary (76-84,106-110).

Due to some of the disadvantages of placebo-controlled trials, physicians and other medical decision-makers should choose practical clinical trials to obtain high quality evidence-based, head-to-head comparisons of clinically relevant alternatives. MacPherson (113) described in detail pragmatic clinical trials, along with the differences between explanatory and pragmatic trials, as illustrated in Table 2.

Table 1. *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 (96).

Table 2. *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 (38,113).

4. Is Evidence from Observational Studies Viable?

In a health technology assessment, Deeks et al (23) concluded that results of observational studies sometimes, but not always, differ from results of randomized studies of the same intervention. They also added that observational studies, however may give seriously misleading results when treated and control groups appear similar in key prognostic factors. Thus, standard methods of case-mixed adjustment do not guarantee removal of bias. Residual confounding may be high even when good prognostic data are available, and in some situations adjusted results may appear more biased than unadjusted results (23). They concluded that all other issues remaining equal, lack of randomization introduces bias into the assessment of treatment effects. The bias may be systematic and appear on average to act in a particular direction if the non-random allocation mechanism leads to a consistent difference in case-mix or it can act in either direction, increasing uncertainty in outcome in ways that cannot be predicted. In addition, statistical methods of analysis cannot properly correct for inadequacies of study design, systematic reviews of effectiveness often do not adequately assess the quality of observational studies.

It is widely held that bias in patient selection may irretrievably weigh the outcome of historically controlled trials in favor of new therapies in observational studies. It is based on an evaluation showing 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 (114). Further, it was also reported that in comparing the effects in RCTs with observational studies in digestive surgery, one-fourth of the observational studies gave different results than randomized trials (115). Poor quality of reporting in observational intervention studies was reported as a potential factor for confounding bias in 98% of

the studies (116). In a 2005 publication, Hartz et al (117) assessed observational studies of medical treatments and concluded that reporting was often inadequate to compare the study designs or allow other meaningful interpretation of results. However, the concept that assignment of the subjects randomly to either experimental or control groups as the perfect science has been questioned (118). While researchers believe that randomization ensures that participating groups will differ only by chance, it does not guarantee that the balance will actually be achieved through randomization (87,119,120). In fact, in acomparison of randomized and observational samples, 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 to randomization showing significant differences in 7 parameters indicating that a randomized design may not be the best in interventional pain management settings always (119).

Benson and Hartz (121), 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. Further, Hartz et al (122), in assessing observational studies of chemonucleolysis, 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 studies or even RCTs; however, cautioning that the potential of comparative observational studies has not been realized because of concurrent inadequacies in their design, analysis, and reporting. Concato et al (123), 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 do not systematically overestimate the magnitude of the effects of treatment as compared with those in RCTs on the same topic. In fact, Shrier et al (124) found that the advantages of including both observational studies and randomized trials in a meta-analysis could outweigh the disadvantages in many situations and that observational studies should not be excluded a priori.

5. Can Methodologic Quality of Observational Studies Be Assessed?

Assessment of methodologic quality is crucial in all types of studies. There are several instruments for methodologic quality assessment of randomized trials. In addition, methodologic quality assessment of randomized trials (50,120,125-144) is quite frequently published in contrast to observational studies. Despite a paucity of the literature, numerous publications dealt with methodologic quality assessment of observational studies (38,42,50,56-58,62,145-148).

West et al (42) in the Agency for Healthcare Research and Quality (AHRQ) evidence report of technology assessment, entitled "Systems to Rate the Strength of Scientific Evidence," provided pertinent evidence to rating the quality of individual articles including observational studies, apart from systematic reviews, RCTs, and diagnostic tests. They assessed 19 systems relating to observational studies or investigations. Of these, they characterized 4 as scales (149-152), 8 as checklists (153-160), 5 as guidance documents (161-165), and 2 as Evidence-Based Practice Centers (EPCs) rating systems for evaluating observational studies identical to those used for RCTs (166,167). They considered 5 key domains to arrive at a set of high-performance scales or checklists pertaining to observational studies, which included comparability of subjects, exposure or intervention, outcome measurement, statistical analysis, and funding or sponsorship. Table 3 illustrates the important domains and elements for systems to rate the quality of observational studies, along with methodologic quality assessment criteria as utilized presently (42). These criteria have been used in multiple systematic reviews with or without weighted scoring (168-172).

Sanderson et al (173) in a systematic review of tools for assessing quality and susceptibility to bias in observational studies in epidemiology identified a number of useful assessment tools. They concluded that tools should be rigorously developed, evidence-based, valid, reliable, and easy to use. Further, they commented that there is a need to agree on critical elements for assessing susceptibility to bias in epidemiology and to develop appropriate evaluation tools. They identified a total of 86 tools comprised of 41 simple checklists, 33 scales, and 12 checklists with additional summary judgments (149-153,156,160,174-189). Among these a number of groups were designed to address specific

Table 3. *Modified AHRQ quality assessment criteria for observational studies.*

CRITERION	Weighted Score
1. Study Question	**2**
• Clearly focused and appropriate question	
2. Study Population	**8**
• Description of study population	5
• Sample size justification	3
3. Comparability of Subjects	**22**
• Specific inclusion/exclusion criteria for all groups	5
• Criteria applied equally to all groups	3
• Comparability of groups at baseline with regard to disease status and prognostic factors	3
• Study groups comparable to non-participants with regard to confounding factors	3
• Use of concurrent controls	5
• Comparability of follow-up among groups at each assessment	3
4. Exposure or Intervention	**11**
• Clear definition of exposure	5
• Measurement method standard, valid and reliable	3
• Exposure measured equally in all study groups	3
5. Outcome measures	**20**
• Primary/secondary outcomes clearly defined	5
• Outcomes assessed blind to exposure or intervention	5
• Method of outcome assessment standard, valid and reliable	5
• Length of follow-up adequate for question	5
6. Statistical Analysis	**19**
• Statistical tests appropriate	5
• Multiple comparisons taken into consideration	3
• Modeling and multivariate techniques appropriate	2
• Power calculation provided	2
• Assessment of confounding	5
• Dose-response assessment if appropriate	2
7. Results	**8**
• Measure of effect for outcomes and appropriate measure of precision	5
• Adequacy of follow-up for each study group	3
8. Discussion	**5**
• Conclusions supported by results with possible biases and limitations taken into consideration	
9. Funding or Sponsorship	**5**
• Type and sources of support for study	
TOTAL SCORE	100

Adapted and modified from West S et al. *Systems to Rate the Strength of Scientific Evidence*, Evidence Report, Technology Assessment No. 47. AHRQ Publication No. 02-E016 (42).

study design types: case-control studies alone (19%), cohort studies alone (27%), and cross-sectional studies alone (7%). Table 4 illustrates domains and criteria for evaluating each tool's content utilized by Sanderson et al (173). The majority of the tools included items relating to methods of selection of study participants, methods of measuring exposure and outcome variables, design specific sources of bias, methods of control confounding, and statistical methods. Only 3% of the tools included conflict of interest or funding sources as a quality measure. Sanderson et al (173) attempted to address weighting; however, they found little consistency among tools with considerable variability in the number of items across domains and across tool types. Consequently, they highlighted the lack of a single obvious candidate tool for assessing quality of observational epidemiological studies. In contrast, West et al (42) have presented the tool for assessment as illustrated in Table 3 and weighting has also been addressed in some reports (168-179).

6. Types of Observational Studies

Observational studies serve a wide range of purposes: from reporting a first hint of a potential cause of a disease to verifying the magnitude of previously reported associations (4). While there multiple types of study designs for observational reports, the main study designs include cohort, case-control, and cross-sectional designs. These designs represent different approaches of investigating the occurrence of health-related events in a given population and time period (4). These studies may address many types of health-related events including disease or disease remission, disability or complications, death or survival, and the occurrence of risk factors. While designs are clearly described, unfortunately terminology is often used incorrectly with authors using diverse terminology to describe these study designs (4,5,34,190).

Deeks et al (23) also described that there is inconsistent use of nomenclature when describing observational studies and other taxonomies may apply different definitions to the same study designs. A taxonomy of study designs that may be used to assess the effectiveness of an intervention has been provided (5,23,153).

To attempt to avoid the problems of inconsistent terminology, 6 features were identified that differentiate between these studies (23). First, some studies make comparisons between groups, while some simply describe outcomes in a single group (e.g., case series). Second, the comparative designs differ in the way that participants are allocated to groups, varying from the use of randomization (RCTs), quasi randomization, geographical or temporal factors (cohort studies), the decisions of healthcare professionals (clinical database cohorts), to the identification of groups with specific outcomes (case-control studies). Third, studies differ in the degree to which they are prospective (and therefore planned) or retrospective, for matters such as the recruitment of participants, collection of baseline data, collection of out-

Table 4. *Domains and criteria for evaluating each tool's content.*

DOMAIN	TOOL ITEM MUST ADDRESS	TOOLS MEETING CRITERIA
Methods for selecting study participants	Appropriate source population (cases, controls, and cohorts) and inclusion or exclusion criteria	92%
Methods for measuring exposure and outcome variables	Appropriate measurement methods for both exposure(s) and/or outcome(s)	86%
Design-specific sources of bias (excluding confounding)	Appropriate methods outlined to deal with any design-specific issues such as recall bias, interviewer bias, biased loss to follow-up or blinding	86%
Methods to control confounding	Appropriate design and/or analytical methods	78%
Statistical methods (excluding control of confounding)	Appropriate use of statistics for primary analysis of effect	78%
Conflict of interest	Declarations of conflict of interest or identification of funding sources	3%

Source: Sanderson S et al. 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 (173).

come data, and generation of hypotheses. Fourth, the method used to investigate comparability of the groups varies: in RCTs no investigation is necessary (although it is often carried out), in controlled before-and-after designs baseline outcome measurements are used, and in cohort and case-control studies, investigation of confounders is required. Fifth, studies differ in the level at which the intervention is applied: sometimes it is allocated to individuals, other times to groups or clusters. Finally, some studies are classified as experimental whereas others are considered as observational.

In experimental studies, the study investigator has some degree of control over the allocation of interventions. In contrast, in observational studies, the groups that are compared are generated according to variation in the use of interventions that occurs regardless of the study. Thus, when allocation is determined largely by health professionals, the treatment decision is based not only on "hard" data such as age, sex, and diagnostic test results, but also on "soft" data, including type and severity of symptoms, rate of development of the illness, and severity of any co-morbid conditions, which are rarely made explicit (191). Further, allocation in observational studies may also be based on factors such as availability of care or geographical location. Thus, there are likely to be systematic differences in the case-mix of patients allocated to the interventions and comparison groups in observational studies. In addition, allocation to groups can also be based on patient choice, as in patient preference trials (192), which may enhance the therapeutic effect of an intervention (193).

West et al (42) described the challenges of rating observational studies, emphasizing that observational study by its very nature "observes" what happens to individuals. Thus, to prevent selection bias, the comparison groups in an observational study are supposed to be as similar as possible except for the factors under study. For investigators to derive a valid result from their observational studies, they must achieve this comparability between study groups (and, for some types of prospective studies, maintain it by minimizing differential attrition).

Figure 1 illustrates a rough guide to a research design (194). Thus, most, but not all, studies fit into one of the main types of research designs: cross-sectional, case-control, or cohort studies as described by the STROBE statement (4,5). However, some studies incorporate more than one design. The most commonly therapeutic interventional studies include cohort or case-control designs (195).

7. How to Report Observational Studies

Reporting of observational research is often not detailed or clear enough to assess the strengths and weaknesses of the investigations (24,34). The STROBE statement (5) for reporting observational studies was developed to ensure clear presentation of what was planned, done, and found in an observational study. In addition, an explanation and elaboration of STROBE was also published (4). The STROBE checklist (Table 5) shows the items related to title, abstract, introduction, methods, results, and discussion sections. The STROBE statement (5) and explanation and elaboration (4) are similar to the revised Consolidated Standards of Reporting Trials (CONSORT) statement for reporting randomized trials, the extension of the CONSORT statement for reporting non-inferiority and equivalence randomized trials, and reporting of pragmatic trials (21,22,37). STROBE provides general reporting recommendations for descriptive observational studies and studies that investigate associations between exposures and health outcomes. STROBE addresses the 3 main types of observational studies: cohort, case-control, and cross-sectional studies.

Cochrane reviews (196) caution with regards to inclusion of observational studies in systematic reviews and meta-analysis due to the problems incorporated in observational studies compared to randomized trials. In an evaluation of 2,993 publications, it was shown that in the majority of studies (greater than 98%) the potential for confounding bias was reported (116). Details on the selection and inclusion of observed confounders were reported in 10% and 51%, respectively. It was also reported that the quality of reporting of confounding score was mediocre. However, these authors concluded that even though the quality of reporting of confounding in articles on observational medical intervention studies was poor, the STROBE statement for reporting of observational studies may considerably impact the reporting of such studies.

In an evaluation of effect of formal statistical significance on the creditability of observational associations (197), statistically significant results offered less than strong support to the creditability for 54% to 77% of the 272 epidemiologic associations for the diverse risk factors and 44% to 70% of the 50 associations from genetic meta-analyses. The analysis of observational studies published in general medical and specialty journals found that the rationale behind the

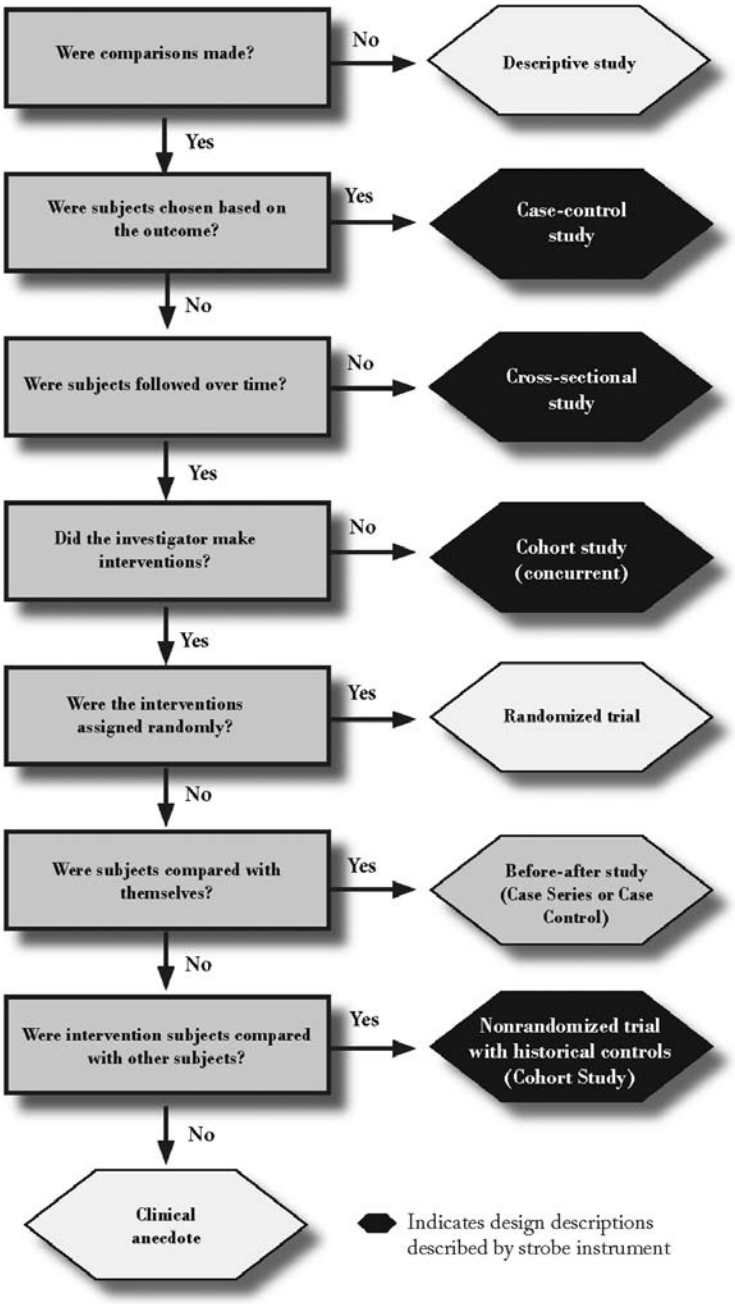

Fig. 1. *Illustration of rough guide to a research design.*

Adapted and modified from Browner WS. Methods. In: *Publishing and Presenting Clinical Research.* Lippincott Williams & Wilkins, Philadelphia, 2006, pp 27-44 (194).

Table 5. *The STROBE statement - checklist of items that should be addressed in reports of observational studies.*

Item number		Recommendation
TITLE and ABSTRACT		
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found
INTRODUCTION		
Background/ rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
METHODS		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(a) Cohort study—Give the eligibility criteria, sources and methods of selection of participants. Describe follow-up method Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls Cross-sectional study—Give the eligibility criteria and the sources and methods of selection of participants (b) Cohort study—For matched studies, give matching criteria and number of exposed and unexposed Case-control study—For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) Cohort study—If applicable, explain how loss to follow-up was addressed Case-control study—If applicable, explain how matching of cases and controls was addressed Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses
RESULTS		
Participants	13*	(a) Report the numbers of individuals at each stage of the study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analyzed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram
Descriptive data	14*	(a) Give study participants characteristics (e.g., demographic, clinical, social), information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) Cohort study—Summarize follow-up time (e.g., average and total amount)
Outcome data	15*	Cohort study—Report numbers of outcome events or summary measures over time Case-control study—Report numbers in each exposure category or summary measures of exposure Cross-sectional study—Report numbers of outcome events or summary measures
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions and sensitivity analyses
DISCUSSION		
Key results	18	Summarize key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalizability	21	Discuss the generalizability (external validity) of the study results
OTHER INFORMATION		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

Adapted from Vandenbroucke JP et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Ann Intern Med* 2007; 147:W163-W194 (4).

choice of potential confounding variables was often not reported (24). Only a few reports of case-control studies in psychiatry explained the methods used to identify cases and controls (32). Further, in a survey of longitudinal studies in stoke research 17 of the 49 articles (35%) did not specify the eligibility criteria (33).

Incomplete or poor reporting has been shown to exist in numerous evaluations (198-203). Consequently, there have been arguments that without sufficient clarity of reporting, the benefits of research might be achieved more slowly (34), and that there is a need for guidance in reporting observational studies (35,36).

The STROBE statement is a checklist of 22 items that are considered essential for good reporting of observational studies (Table 5). These items relate to the article title and abstract (item 1), the introduction (items 2 and 3), methods (items 4 to 12), results (items 13 to 17), and discussion sections (items 18 to 21), and other information with funding (item 22). Of these, 18 items are common to all 3 designs, while 4 (items 6, 12, 14, and 15) are design-specific, with different versions for all or part of the item. Further, STROBE provides appropriate instructions for some items indicated by asterisks, with necessity to provide information separately for cases and controls in case-control studies, or exposed and unexposed groups in cohort and cross-sectional studies. While Table 5 illustrates a single checklist, on the STROBE website (www.strobestatement.org/Checklist.html), separate checklists are available for each of the 3 study designs.

In the reporting of observational studies, quality assessment criteria also must be taken into consideration (42) (Table 3). Multiple investigators have reported on the issues of quality of reporting, both for RCTs and observational studies, and have showed significant improvements when methodologic quality of assessment criteria, as well as reporting statements were followed (204-209).

8. Observational Study Designs

Three major study designs considered are cohort, case-control, and cross-sectional. Prospective or retrospective are terms not well defined (210). These words have been controversial. Some have described cohort and prospective as synonymous and reserve the word retrospective for case-control studies (211). Others have described prospective and retrospective cohort studies to distinguish the timing of data collection relative to when the idea of the study was developed (212). Another usage distinguishes prospective and retrospective case-control studies depending on whether the data about the exposure of interest existed when cases were selected (213). It has been advised by some not to use these terms (214), whereas others advise adaptation of the alternatives concurrent and historical for describing cohort studies (215).

8.1 Sources of Bias in Observational Studies

Cochrane Reviewers' Handbook (216) has laid out the 4 main sources of systematic bias in trials of effect of healthcare as being selection bias, performance bias, attrition bias, and detection bias (Table 6). Of all the sources of bias, selection bias is the most important aspect.

8.1.1 Selection Bias

The greatest distinction between the results of randomized and observational studies is the risk of selection bias, where systematic differences in comparison groups arise at baseline. In observational studies, selection bias will be introduced when participants chosen for one intervention have different characteristics from those allocated to the alternative intervention. In observational studies, the choice of a given intervention is largely dependent on the discretion of the treating physician, similar to clinical practice. Thus, the choice of an intervention under these circum-

Table 6. *Sources of bias.*

Source of bias	RCTs	Cohort Studies (Observational)
Selection bias	Randomisation	Control for confounders
Performance bias	Blinding (of participants and/or investigators)	Measurement of exposure
Attrition bias	Completeness of follow-up	Completeness of follow-up
Detection bias	Blinded outcome assessment	Blinded outcome assessment

stances will be influenced not only by the clinician's own personal preference for one intervention over another, but also by the patient's preference, characteristics, and clinical history (23). However, it has been stated that sometimes a clinician's treatment decision will be influenced by subtle clues that are not easily identifiable, which may allegedly result in treatment groups that are incomparable, often with one intervention group "heavily weighted by the more severely ill," (212,217). Confounding by severity or by prognosis may be considered as a special form of confounding, which occurs where the severity of the extraneous condition influences both the introduction of the intervention and the outcome of the study (218), that is, any treatment reserved for the most ill will be associated with the worst outcomes.

Protopathic bias describes situations where the first symptoms of a given outcome are the reason for treatment initiation with a therapeutic intervention or agent that is inadvertently prescribed for an early manifestation of a disease that has not yet been diagnostically detected (219). However, selection bias based on severity and confounding by severity are not potential issues in interventional pain management settings, since chronic pain patients are not included for interventional techniques unless they have had pain for a certain duration of time (usually 6 months or longer) and have failed conservative modalities of treatment, and also have been investigated appropriately.

A restricted cohort design has been developed to improve protopathic bias (219,220). This approach involves restricting the eligibility criteria of cohort studies to those used in clinical trials, defining a "zero-time" point from which patients are followed up, and using an approximation of intention-to-treat analysis. This has been somewhat validated by using a study of β-blocker therapy after acute myocardial infarction, with similar results reported by RCTs (220).

One method for circumventing the problem of selection bias is to match individuals who are similar with respect to characteristics that might affect the study results. Matching is not limited to making the groups uniform, but may be used for any characteristic related to the probability of experiencing the outcome of the study. However, a disadvantage of matching groups is that the investigators cannot study the effect that the matching characteristic has on the outcome being measured.

8.1.2. Other Biases

Other biases include attrition, detection, and performance bias. Attrition bias occurs with drop-outs, detection bias occurs if the assessment of outcomes is not standardized and blinded, and performance bias occurs if there are errors and inconsistencies in the allocation, application, and recording of interventions. While all the biases can also occur in RCTs, there is perhaps potential for their impact to be greater in observational studies which are usually undertaken without protocols specifying standardized interventions, outcome assessments, and data recording procedures.

While matching is not a primary requirement in cohort studies, occasionally, matching is used to make groups comparable at the start of follow-up. Matching in cohort studies makes groups directly comparable for potential confounders and presents fewer intricacies than with case-control studies (4). However, it is not necessary to take the matching into account for the estimation of the relative risk, because matching in cohort studies may increase statistical precision, investigators might allow for the matching in their analysis and thus obtain narrower confidence intervals (CIs) (221).

"Cohort profiles" include detailed information on what was measured at different points in time in particular studies (222). In addition, all candidate variables are a requirement for utilization in statistical analysis (223).

8.2 Study Design and Methods Terminology

As per the reporting guidelines of STROBE statements (4,5), cohort, case-control, and cross-sectional studies are described here.

8.2.1 Cohort Studies

Cohort in Latin, "cohors," means a group of soldiers. The studies may be prospective or retrospective and sometimes 2 cohorts may be compared. These are the best methods for determining the incidence and natural history of the condition (224,225). Famous examples of cohort studies include the Framingham Heart study (226), the UK studies of doctors who smoke (227), studies on British children born in 1958 (228), adverse socioeconomic conditions in childhood (229), and the use of accident and emergency departments by patients with diabetes (230).

8.2.1.1 Prospective Design

In a prospective cohort study, a group of individuals who do not have the disease, but may or may not have the exposure, are selected and followed over a period of time. Subsequently, a variety of relevant variables are measured over a period of time.

8.2.1.2 Retrospective Design

In a retrospective cohort study, information on exposure and diseases is already collected, either as a part of another study or medical records. Consequently, the existing information is used to evaluate the relationship between exposure and disease over a period of time. However, the first part may involve retrospective study data collection, whereas the subsequent part may involve prospective follow-up of the same subjects over time to assess the occurrence of new outcomes.

8.2.1.3 Key Points

Cohort studies key points are as follows (224):

- Cohort studies describe incidence or natural history.
- They analyze predictors (risk factors) thereby enabling calculation of relative risk.
- Cohort studies measure events in temporal sequence thereby distinguishing causes from effects.
- Retrospective cohorts where available are cheaper and quicker.
- Confounding variables are the major problem in analyzing cohort studies.
- Subject selection and loss to follow-up is a major potential cause of bias.

8.2.2 Case-Control Studies

Case-control studies are very popular with clinical researchers as well as interventional pain physicians. In this, 2 groups of individuals are selected, either diseased or non-diseased. The exposure is then measured in both of the groups and the association of exposure to the disease is calculated. However, the cases and controls have to be selected from the same underlying population. In a case-control study, intervention may also be compared, either with 2 interventions or with no intervention.

8.2.2.1 Study Design

Case-control studies have been described as only retrospective by some (224). However, when measuring interventions, these can also be prospective. Case-control studies not only determine the relative importance of a predictor variable in relation to the presence or absence of the disease, but can be used to calculate odds ratio, which in turn is usually approximate to the relative risk. Case-control studies are very common and are particularly useful for studying infrequent events. An example is the study of atrial fibulation in middle aged men during exercise (231).

The controls are usually matched to the cases for a characteristic. This process of matching can be efficient in case-control studies because a comparison group similar to the index group is selected.

8.2.2.2 Key Points

Important features of case-control studies are as follows (224):

- Case-control studies are simple to organize.
- Compare 2 groups.
- Aim to identify predictors of an outcome.
- Permit assessment of the influence of predictors on outcome via calculation of an odds ratio.
- Useful for hypothesis generation.
- Can only look at one outcome.
- Bias is a major problem.

8.2.3 Cross-sectional Studies

Cross-sectional studies are primarily used to determine prevalence, which equals the number of cases in the population at a given point in time. Consequently, all the measurements on each person are made at one point in time. In addition, cross-sectional studies are used to infer causation. It is considered as a one-time snapshot study (225). Usually a sample is selected from a target population, thus, the exposure of interest, the disease of interest, and other covariates are measured in the selective population at one point of time.

8.2.3.1 Study Design

As an example, all patients suspected of lumbar facet joint pain with low back pain may be targeted. Subsequently, controlled comparative local anesthetic blocks may be applied and prevalences determined in this population or in the population of all patients with low back pain in that particular setting.

8.2.3.2 Key Points

Cross-sectional studies key points are as follows (224):

- Cross sectional studies are the best way to determine prevalence.
- Are relatively quick.
- Can study multiple outcomes.
- Do not themselves differentiate between cause and effect or the sequence of events.

9. Outcomes

Clearly defined primary and secondary outcome measures are essential for an appropriately conducted observational study to address the possible association between exposures (risk factors) and outcomes. Differences between groups in outcome variables are believed to be the results of different interventions. While the primary outcome is the outcome of greatest importance, data on secondary outcomes are used to evaluate additional effects of interventions. Outcomes are also important to compare the observational studies with randomized studies. Outcomes assessments are described in a previous publication (38).

Multiple instruments are available to assess the impact of chronic pain and subsequent interventions on quality of life (232-246). Multiple measures have been developed which evaluate the disease-specific disability (234), general pain measures (238), or other measures evaluating health and illness (233). Oswestry Disability Index (ODI), Roland-Morris Disability Questionnaire (RDQ), and the Neck Pain Disability Index (NDI) are the most commonly recommended condition-specific outcome measures for spinal disorders (234-236). Even though all these instruments are considered as objective evaluations, all of them depend on subjective information; consequently objective assessment is from subjective reporting.

Reduction in pain intensity is the most frequently employed primary outcome in RCTs as well as in observational studies. Pain assessment scales have been shown to have both face validity and intuitive appeal, despite multiple questions (240). 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 (242-244).

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" (247,248). It is crucial to distinguish between the responsiveness as a test property and the MCID as a quantity useful in interpreting study results. MCID shows the change of health status rather than differences between patients (249). In contrast, minimal clinically important change (MCIC) measures relevant outcome measures enabling a comparison between interventions on the patient level.

10. Data Presentation and Analysis

For transparent reporting of the research, appropriate research methodology and statistical techniques are important. However, the methods of statistical inference in current use are not evidence-based, leading to widespread misconceptions (250). It is generally perceived that statistical methods can provide a number that by itself reflects a probability of reaching erroneous conclusions. Thus, it is important to understand how the strength of evidence in a particular study can be related to, and combined with, strength of other evidence. Consequently, statistical methods are important in comparing groups for determination of sample size, outcomes, and additional analysis. The literature is replete with publications describing deficiencies of medical statistics (250,251).

10.1 Sample Size

Sample size is one of the important features of an observational study. This has to be planned carefully with a balance between clinical and statistical considerations. AHRQ criteria for methodologic assessment of observational studies provide significant weight for sample size consideration as shown in Table 3 (42). The study should be large enough to have a high probability (power) of detecting a statistically significant clinically important difference of a given size. The size of an effect deemed important is inversely related to the sample size necessary to detect it — large samples are necessary to detect small differences (21). Consequently, it has been widely believed 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, either in randomized trials or more so in observational studies (252). In fact, multiple reviews of published trials have consistently found that a high proportion of trials have a very low power to detect clinically meaningful treatment effects (253-255). While in interventional pain management studies, a sample size of 50 in the smallest group has been considered to be appropriate (132) for randomized trials, there are no such guidances available for observational trials, except that AHRQ criteria describes that sample size is essential to maintain methodologic quality assessment.

10.1.1 Determination of Sample Size

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.

10.1.2 Parameter Definition

An appropriate sample size generally depends on 5 study design parameters (256,257). These are 1) minimum expected difference or the ES, 2) estimated measurement variability, 3) desired statistical power, 4) significance criterion, and 5) whether a one- or two-tailed statistical analysis is planned.

10.1.2.1 Minimum Expected Difference

Minimum expected difference is the smallest measured difference between comparison groups that the investigator would like the study to detect (256). As the minimum expected difference is made smaller, the sample size needed to detect statistical significance increases. Estimation of reasonable minimum difference is based on clinical judgment and experience with the problem being investigated and the results of pilot studies or a literature review.

10.1.2.2 Estimated Measurement Variability

Estimated measurement variability is represented by the expected standard deviation (SD) in the measurements made within each comparison group (256). Even though it is estimated on the basis of subjective experience, 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.

10.1.2.3 Statistical Power

Statistical power is the probability of demonstrating statistical significance if the study hypothesis is true. In general, power increases as sample size increases. However, there is an obvious tradeoff 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 0.08% (256).

10.1.2.4 Significance Criterion

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, or the *P* value.

10.1.2.5 One- or Two-Tailed Statistical Analysis

One- or two-tailed statistical analysis may be performed. Generally it is not 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. However, two-tail analysis is most commonly performed.

10.1.2.6 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 number 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 (257). However, until the allocation ratio is allowed to exceed 2:1 with the same total sample size, the power falls very slowly.

10.1.2.7 Minimizing the Sample Size

Multiple strategies have been described for minimizing the sample size (258). These include use of continuous measurements instead of categories, more precise measurements, paired measurements, unequal group sizes, and expanding the minimum expected difference.

10.1.2.8 Importance of Sample Size

Large samples are needed to distinguish a small association from no association. Small studies often provide valuable information, but wide CIs may indicate that they contribute less to current knowledge in comparison with studies providing estimates with narrower CIs. The importance of sample size determination in observational studies depends on the context. If an analysis is performed on data that were already available for other purposes, the main question is whether the analysis of the data will produce results with sufficient statistical precision to contribute substantially to the literature, and sample size considerations will be informal (4). Formal, a priori calculation of sample size

may be useful when planning a new study (258-260). Even then, such calculations are associated with more uncertainty than implied by the single number that is generally produced. It has been stated that estimates of the rate of the event of interest or other assumptions central to calculations are commonly imprecise, if not guesswork (261). The precision obtained in the final analysis can often not be determined beforehand because it will be reduced by inclusion of confounding variables in multivariable analyses (262), the degree of precision with which key variables can be measured (263), and the exclusion of some individuals.

10.2 Statistical Methods

In general, there is no one correct statistical analysis but, rather, several possibilities that may address the same question, but make different assumptions. Additional analyses are needed, either instead of, or as well as, those originally envisaged, and these may sometimes be motivated by the data (4). The distinction between pre-specified and exploratory analyses may sometimes be blurred.

10.2.1 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, i.e., age, height, weight. Discrete variables deal only with whole numbers, they can take on only certain definite and separate values, i.e., 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.

10.2.2 Confounding Variables

If groups being compared are not similar with regard to some characteristics, adjustment should be made for possible confounding variables by stratification or by multivariable regression (264,265). Confounding, literally means confusion of effects. A study might seem to show either an association or no association between an exposure and the risk of a disease. In reality, the seeming association or lack of association is due to another factor that determines the occurrence of the disease but that is also associated with the exposure. The other factor is called the confounding factor or confounder. Confounding thus gives a wrong assessment of the potential "causal" association of an exposure (4).

Taking confounders into account is crucial in observational studies, but analyses adjusted for confounders do not automatically establish the caudal part of an association. Consequently, results may still be distorted by residual confounding, random sampling error, selection bias, and information bias (264).

Often, the study design determines which type of regression analysis is chosen. Cox proportional hazard regression is commonly used in cohort studies (265,266), whereas logistic regression is often the method of choice in case-control studies (267,268).

10.2.3 Parametric vs. Non-Parametric Statistics

Typically used parametric tests are t-tests and 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 a 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 clinical studies (269,270). However, it has been stated that, parametric methods are applicable if the sample size is 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" (271).

The explicit rationale for recommending non-parametric over parametric methods is not obvious (269). The empirical statistical research has clearly demonstrated that the t-test does not inflate type I (false-positive error) except 5% of the time (272). Thus, concern over the relative advantages of parametric and non-parametric methods is focused on type II errors or false-negative results (273-277).

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 liner regression, has been shown to be more powerful than the t-test (278,279).

10.2.4 The P Value

P value was proposed as an informed index to be used as a measure of discrepancy between the data and null hypothesis (280). 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 (250). It has been suggested that *P* value 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.

Generally it is interpreted a *P* value of 0.05 means that the null hypothesis has a probability of only 5%. In contrast, a *P* value of 0.05 represents that there is a 95% or greater chance that the null hypothesis is correct (280,281). Thus, misinterpretation may reinforce the mistaken notion that the data alone can tell the probability that a hypothesis is true (250).

In modern medicine, more emphasis is laid upon CIs than *P* values. The problem with the *P* value is that a small effect in a study with large sample size can have the same *P* value as a large effect in a small study. Consequently, when the *P* value was proposed, some scientists and statisticians criticized the logical basis and practical utility of *P* value (282).

10.2.5 Confidence Intervals (CIs)

The CIs, along with *P* values, are crucial to determine the likelihood that a difference in a study is due to chance. However, CIs are far from a panacea (250). In essence, CIs embody many of the same problems that afflict current methods, albeit in a subtler form (278,283,284). The most important drawback of CIs is that they offer no mechanism to unite external evidence with that provided by an experiment.

The level of certainty (power with which they can make a conclusion) is essential in all reports. If the power is 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. Even though, it only means the difference is likely to be real and not due to chance, but the question remains if the difference is clinically significant.

10.2.6 Odds Ratios

Because the number of cases and the number of controls is predetermined in a case-controlled study, the relative risk cannot be used (279,285). Thus, an alternative way of measuring risk is in terms of the odds ratio. The odds of a disease given a risk factor is the probability of having the disease with the factor divided by the probability of not having the disease with the factor present. Thus, the odds ratio is the ratio of the odds of the disease with the risk factor present, divided by the odds of the disease with the risk factor absent. For example, smoking in lung cancer patients, divided by the odds of smoking in the controls is equivalent to the odds ratio for the disease given the risk factor. If the condition or the disease is rare, the odds ratio and relative risk are almost the same.

10.2.7 Relative Risk

The relative risk is the ratio of the probability of the event with the factor present compared with (divided by) the probability of the event occurring with the factor absent. Thus, the relative risk is only determined over a period of a time frame for the event to occur. The relative risk factor is determined for 2 different levels of the risk factor. If the risk factor is continuous, the 2 levels must be chosen. However, if the risk factor were discrete, the relative risk is determined pair-wise.

10.3 Subgroup Analysis

In addition to the main analysis, other analyses are often done in observational studies (4). They may address specific subgroups, the potential interaction between risk factors, the calculation of attributable risks, or use alternative definitions of study variables in sensitivity analysis. There is continued debate on the safety and appropriateness of subgroup analysis, and multiplicity of analysis in general (24,33,286-290). While there is value in exploring whether an overall association appears consistent across several, preferably pre-specified subgroups, especially when a study is large enough to have sufficient data in each subgroup, there is too great a tendency to look for evidence of subgroup-specific associations, or effect measure modification, when overall results appear to suggest little or no effect. Thus, occasionally important findings may also arise by chance.

11. Design of Protocol and Reporting

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 Institutional Review Board (IRB) to evaluate the quality of the study design, the ethics of the conducting the study, and the safeguards provided for patients, including a review of the informed consent statement. In addition, 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. Further, 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.

The STROBE statement provides general reporting recommendations for descriptive observational studies and studies that investigate associations between exposures and health outcomes. Taking into account empirical evidence and theoretical consideration, a group of methodologists, researchers, and editors developed STROBE recommendations to improve the quality of reporting of observational studies. The STROBE statement provides guidance to authors about how to improve the reporting of observational studies and facilitates critical appraisal and interpretation of studies by reviewers, journal editors, and readers. Further, the explanatory and elaboration document (4) was also published to enhance the use, understanding, and dissemination of the STROBE statement. Table 5 illustrates the STROBE statement, a checklist of items that should be addressed in reports of observational studies. STROBE recommendations do not address observational studies that specifically address diagnostic tests. STARD (291) recommendations were separately developed for diagnostic studies.

11.1 Title and Abstract

The title should indicate the study's design with a commonly used term in the title or the abstract so that readers are able to easily identify the design that was used from the title or abstract. An explicit, commonly used term for the study design also helps ensure correct indexing of articles in electronic databases (121,196). The title should illustrate the type of study, either a cohort, case-control, or cross-sectional.

The structured abstract must provide a series of headings pertaining to the background, design, conduct, and analysis of a trial with standardized information appearing under each heading (4,5,21,22,292-294). It may be appropriate to describe limitations in the abstract (294). It has been shown that structured abstracts are of higher quality than the more traditional descriptive abstracts, and they also allow readers to find information more easily (295).

Typical components include a statement of the research question, a short description of methods and results, and a conclusion (296). Abstracts should summarize key details of studies and should only present information that is provided in the article. Further, key results should be provided in a numerical form that includes numbers of participants, estimates of associations, and appropriate measures of variability and uncertainty (e.g., odds ratios with CIs).

A series of headings pertaining to the background, design, conduct, limitations, and analysis of a study may provide easy to understand information (292).

11.2 Introduction

The introduction section should describe why the study was done and what questions and hypotheses it addresses (4). Further, the study should describe the context and potential contribution to current knowledge.

11.2.1 Background/Rationale

The scientific background of the study provides important context and sets the stage for the study and describes its focus. It provides an overview of what is known on a topic and what gaps in current knowledge are addressed by the study. Background material should note recent pertinent studies and any systematic reviews of pertinent studies.

11.2.2 Objectives

The study should clearly state specific objectives, including any prespecified hypotheses. Objectives are the detailed aims of the study. Well crafted objectives specify populations, exposures and outcomes, and parameters that will be estimated.

11.3 Methods

The methods section should describe what was planned and what was done in sufficient detail to allow others to understand essential aspects of the study, to judge whether the methods were adequate to provide reliable and valid answers, and to assess whether any deviations from the original plan were reasonable.

11.3.1 Study Design

The study should present key elements of the study design. The study design should describe clearly if it is a cohort study, case-control study, or cross-sectional study rather than retrospective or prospective (292). STROBE recommends not using the words "prospective" and "retrospective" nor alternatives such as "concurrent" and "historical." If these words are used, they should define what they mean. Most importantly, authors should describe exactly how and when data collection took place.

11.3.2 Setting

Setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection must be described. Information about setting includes recruitment sites or sources (i.e., hospital outpatient, clinic, ambulatory surgery center). Further, information about location may refer to the city, county, town, hospital, practices, or country where the investigation took place. Information about the dates is essential rather than length of time periods. It has been stated that almost 80% of 132 reports in oncology journals that used survival analysis included the starting and ending dates for accrual of patients, but only 24% also reported the date on which follow-up ended (198).

11.3.3 Participants

Participants are described based on the design.

For a cohort study, eligibility criteria and the sources and methods of selection of participants and the follow-up must be described. For a case-control study, the eligibility criteria and the sources and methods of case ascertainment and control selection, and the rationale for the choice of cases and controls must be described. For a cross-sectional study, eligibility criteria and the sources and methods of selection of participants must be described.

Generally, study population is restricted by defining clinical, demographic, and other characteristics of eligible participants. Typical eligibility criteria relate to age, gender, diagnosis, and comorbid conditions. In spite of the great importance, eligibility criteria often are not reported adequately (33).

Eligibility criteria may be presented as inclusion and exclusion criteria, even though this distinction is not always necessary or useful. Regardless, STROBE advises to report all eligibility criteria and also to describe the group from which the study population was selected, such as general population of a region or country or patient population attending a clinic, and the method of recruitment, whether by referral or self selection through advertisements, etc. The validity of results is facilitated by details about follow-up procedures, including whether procedures minimized non-response and loss to follow-up and whether the procedures were similar for all participants.

Matching is much more common in case-controlled studies, but occasionally, investigators use matching in cohort studies to make groups comparable at the start of the follow-up. Matching in cohort studies makes groups directly comparable for potential confounders and presents fewer intricacies than with case-controlled studies. However, it is not necessary to take the matching into account for the estimation of the relative risk (221), because matching in cohort studies may increase statistical precision, investigators might allow for the matching in their analyses and thus obtain narrower CIs.

In case-control studies matching is done to increase a study's efficiency by ensuring similarity in the distribution of variables between cases and controls, in particular the distribution of potential confounding variables (221,297). Since matching can be done in multiple ways, with one or more controls per case, the rationale for the choice of matching variables and the details of the method used should be described. Commonly used forms of matching are frequency matching or group matching and individual matching. In frequency matching, investigators choose controls so that the distribution of matching variables becomes identical or similar to that of cases (4). Individual matching involves matching one or several controls to each case. Even though intuitively appealing and sometimes useful matching case-control studies have a number of disadvantages, it is not always appropriate, and needs to be taken into account in the analysis.

11.3.4 Variables

All variables considered for, and included in, the analysis including outcomes, exposures, predictors, potential confounders, and potential effect modifiers should be defined. Clear definitions and steps taken to adhere to them are particularly important for any disease condition of primary interest in the study (4).

11.3.5 Data Sources/Measurement

The way exposures, confounders, and outcomes were measured affects the reliability and validity of a study. Measurement error and risk classification of exposures or outcomes can make it more difficult to detect cause-effect relationships, or may produce spurious relationships. Consequently, it has been reported that error in measurement of potential confounders can increase the risk of residual confounding (298,299). STROBE advises that, it is helpful, if authors report the findings of any studies of the validity or reliability of assessment or measurements, including details of the reference standard that was used.

11.3.6 Bias

Bias studies produce results that differ systematically from the truth. It is important to report what measures were taken during the conduct of a study to reduce the potential bias. Ideally, investigators must consider potential sources of bias when they plan their study. Further, at the stage of reporting, authors must assess the likelihood of relevant biases, specifically, the direction and magnitude of bias. For instance, in case-control studies information bias can occur, but may be reduced by selecting an appropriate control group (199,200).

In many cases, authors do not address important biases when reporting their results. In fact, among 43 case-control and cohort studies published from 1990 to 1994 that investigated the risk of second cancers in patients with a history of cancer, medical surveillance bias was mentioned in only 5 articles (201). Further, a survey of reports of mental health research published during 1998 in 3 psychiatric journals found that only 13% of 392 articles mentioned response bias (202). Finally, a survey of cohort studies in stroke research found that 14 of 49 (28%) of articles published from 1999 to 2003 addressed potential selection bias in the recruitment of study participants and 35 (71%) mentioned the possibility that any type of bias may have affected results (33).

11.3.7 Study Size

A study should be large enough to obtain a point estimate with a sufficiently narrow CI to meaningfully answer a research question. However, in observational studies large samples are needed to distinguish a small association from no association, whereas small studies often provide valuable information, but wide CIs may indicate that they contribute less to current knowledge in comparison with studies providing estimates with narrower CIs.

11.3.8 Quantitative Variables

Authors should explain how quantitative variables were handled in analysis and also must describe which groupings were chosen and why if applicable. Investigators generally make choices regarding how to analyze quantitative data about exposures, effect modifiers, and confounders. Further, grouping choices may have important consequences for later analysis (300,301). The STROBE statement (4) explains why and how grouping quantitative data, including the member of categories, the cut-points, and category mean or how median values were grouped.

11.3.9 Statistical Methods

Any and all methods used to examine subgroups and interactions should be described. Missing data are common in observational, as well as randomized studies; however, missing data must be reported appropriately and addressed. Few articles report in detail on the problem of missing data (33,200). Missing data may be handled in many ways, either by complete-case "analyses" or based on a model for the probability of an observation being missing (4,302-304). A clear description of the reasons for missing values should be provided, and indicate how many individuals were excluded because of missing data when describing the flow of participants through the study. For analyses that account for missing data, authors should describe the nature of the analysis (e.g., multiple imputation) and the assumptions that were made (e.g., missing at random).

In a cohort study, analysis is carried out using life table methods or other approaches that are based on the person-time of follow-up and time to developing the disease of interest. Among individuals who remain free of the disease at the end of their observation period, the amount of follow-up time is assumed to be unrelated to the probability of developing the outcome.

In case-control studies, a matched analysis is often necessary, because in individually matched case-control studies, a crude analysis of the odds ratio, ignoring the matching, usually leads to an estimation that is biased towards unity. This is understood as a stratified analysis: each case is seen as one stratum with his or her set of matched controls. In individually matched studies, the most widely used method of analysis is conditional logistic regression, in which each case and their controls are considered together. The conditional method is necessary when the number of controls varies among cases, and when, in addition to the matching variables, other variables need to be adjusted for.

Most cross-sectional studies use a pre-specified sampling strategy to select participants from a source population. Sampling may be more complex than taking a simple random sample, however. This may include several stages and clustering of participants. Proportionate stratifications may ensure that subgroups with a specific characteristic are correctly represented. Disproportionate stratification may be useful to oversample a subgroup of particular interest.

11.3.9.1 Sensitivity Analyses

Sensitivity analyses are useful to investigate whether or not the main results are consistent with those obtained with alternative analysis strategies or assumptions (305). Issues that may be examined include the criteria for inclusion in analyses, the definitions of exposures or outcomes (306), which confounding variables merit adjustment, the handling of missing data (307,308), possible selection bias or bias from inaccurate or inconsistent measurement of exposure, disease and other variables, and specific analysis choices, such as the treatment of quantitative variables. However, in the modern era, sophisticated methods are increasingly used to simultaneously model the influence of several biases or assumptions (309-311).

11.3.9.2 Intention-to-Treat Analysis

One of the commonly recommended strategies to handle such issues as protocol violations and withdrawals to analyze all participants according to their original group assignment, regardless of what subsequently occurred, is intention-to-treat analysis. While it is not always straightforward to implement, due to reasons known and unknown, it is common for some patients to not complete a study. Thus, these participants cannot be included in the analysis; they are customarily referred in the analysis of all available participants as an intention-to-treat analysis. The term is sometimes inappropriately used when some participants for whom data are available are excluded to improve analysis. Conversely, analysis can be restricted to only participants who fulfill the protocol in terms of eligibility, interventions, and outcome assessment. Such an 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 (312,313). The scientific community feels that studies reporting an intention-to-treat analysis are also associated with some other aspects of good study design and reporting, such as describing a sample size calculation (314). Finally, subjects included in intention-to-treat analyses regardless of their follow-up status requires investigators to deal with the resulting missing data.

11.3.9.2.1 Last Observation Carried Forward (LOCF)

Last observation carried forwards (LOCF) is the most common approach in the replacement of each subject's missing data with his or her last non-missing observation (315). This method works best if the observations are expected to remain at the same level or if there are only a few missing values. However, if the observations in a test are expected to increase or decrease over time this method may not provide appropriate and clinically reliable data.

11.3.9.2.2 Best or Worst Case Imputation

Two other methods when dealing with missing data are best case and worst case imputation. This essentially leads to either an under or over evaluation of the data. Best or worse case imputation may be used to assess a lower bound of efficacy as a demonstration of robustness (316).

11.3.9.2.3 Mean Value Methods

A natural method of imputation is to use the mean value of the recorded observations. This method leads to lower variance; however, a logical concern here is that the dropouts might be more likely to be patients with more extreme values (e.g., 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.

11.3.9.2.4 Regression Methods

Linear regression methods can be used for imputation. Calculations need to control for factors studied which are not being investigated for association (38).

11.4 Results

The results section should give a factual account of what was found, from the recruitment of study participants and the description of the study population to the main results and ancillary analyses. It should be free of interpretations and discursive text reflecting the authors' views and opinions (4).

11.4.1 Participants

Detailed information on the process of recruiting study participants must be described. Ideally, it is recommended that, investigators provide an account for numbers of individuals considered at each stage of recruiting study participants, from the choice of a target population to the inclusion of participants' data in the analysis (4). It is recommended that in case-control studies, authors describe the flow of participants separately for case and control groups (317). It has been shown that among epidemiologic studies published in 10 general epidemiology, public health, and medical journals, some information regarding participants was provided in 47 of 107 case-control studies (59%), 49 of 154 cohort studies (32%), and 51 of 86 cross-sectional studies (59%) (203). Incomplete or absent reporting of participation and non-participation in epidemiological studies was also documented in 2 other surveys of the literature (24,33). The reasons why people no longer participated in a study or why they were excluded from statistical analysis should be explained.

An informative and well-structured flow diagram can readily and transparently convey information that might otherwise require a lengthy description (21,318). The diagram may include the main results, such as the number of events for the primary outcome. While the STROBE document recommends a flow diagram particularly for complex observational studies, they do not propose a specific format for the diagram. In such cases, a CONSORT diagram may be utilized (Fig. 2) (21).

11.4.2 Descriptive Data

Summary of continuous variables for each study group should be provided by giving the mean and SD, or when the data have an asymmetrical distribution, as often is the case, the median and percentile range (i.e., 25th and 75th percentiles) (4). In studies that compare groups, the descriptive characteristics and numbers should be given by group. Inferential measures such as standard errors and CIs should not be used to describe the variability of characteristics, and significance tests should be avoided in descriptive tables (4). *P* values are not considered as appropriate criterion for selecting which confounders to adjust for in analysis; even small differences in a confounder that has a strong effect on the outcome can be important (319,320).

In cohort studies, documentation must be provided as to how an exposure relates to other characteristics and potential confounders. In contrast, in case-control studies, potential confounders cannot be judged by comparing cases and controls. Control persons represent the source population and will usually be different from the cases in many respects. However, in case-control studies, the equivalent of comparing exposed and non-exposed for the presence of potential confounders (as is done in cohorts) can be achieved by exploring the source population of the cases: if the control group is large enough and represents the source population, exposed and unexposed controls can be compared for potential confounders (305,321).

It is also essential to describe the extent of follow-up for the available outcome data. The data may be presented as the summary of the average follow-up with either mean or median follow-up time or both. The mean provides information about the total number of person-years by multiplying it with the number of study participants. It may also be beneficial to present minimum and maximum times or percentiles of the distribution to show the spread of follow-up times. It has been shown that almost half of the 132 articles in cancer journals did not give any summary of length of follow-up (198).

11.4.3 Outcome Data

Outcome data is reported differently for different types of studies. For a cohort study, outcome data is reported by numbers of outcome events or summary measures over time. For a case-control study, outcome data is reported by numbers in each exposure category, or summary measures of exposure, whereas for a cross-sectional study, outcome data is reported by numbers of outcome events or summary measures (322-326). However, before addressing the possible association between exposures (risk factors) and outcomes, relevant descriptive data must be provided. Further, association also may be reported.

11.4.4 Main Results

The STROBE documents (4,5) recommend that unadjusted analysis together with the main data should be presented, essentially providing information about the number of cases and controls that were exposed or not. Thus, in many situations, authors may present results of unadjusted or minimally adjusted analysis and those from fully adjusted analysis. For adjusted analysis, the number of persons in the analysis must be reported, as this number may differ because of miss-

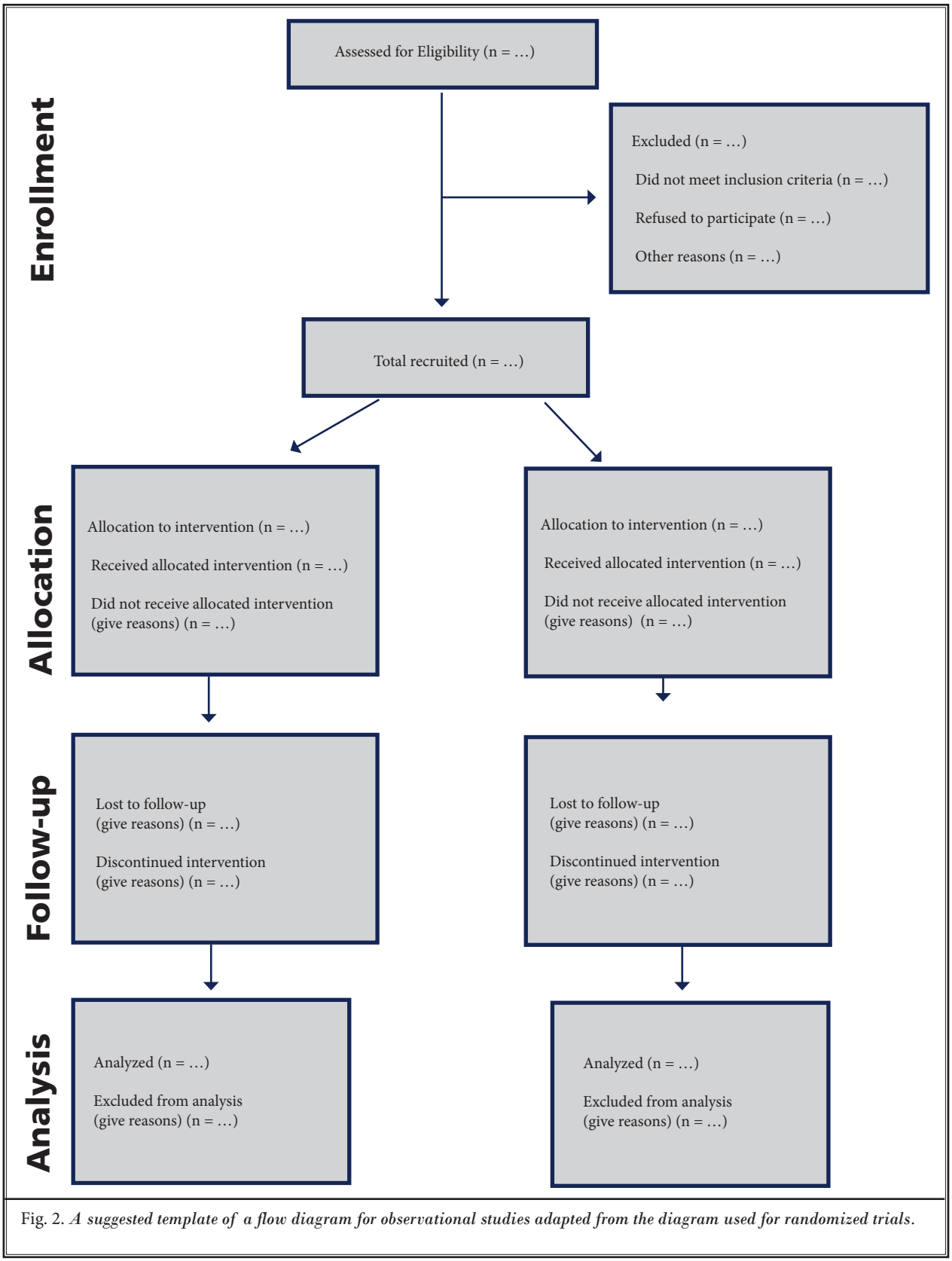

Fig. 2. *A suggested template of a flow diagram for observational studies adapted from the diagram used for randomized trials.*

ing values in covariates. Further, estimates should be given with CIs.

All potential confounders should be considered and explained, along with the criteria for excluding or including variables in statistical models. The STROBE statement and multiple others do not advise selecting confounders based solely on statistical significant testing (4,5,33,320,327,328). While CIs are reported in most articles, very few authors explain their choice of confounding variables (24,33).

As a minimum, the STROBE document (4) recommends that authors should report the category boundaries and the range of the data and the mean or median values within categories. In tables, outcomes should be given for each exposed category, for example as counts of persons at risk, person-time at risk, if relevant, separately for each group (e.g., cases and controls).

Further, if relevant, authors should consider translating relative risk into absolute risk for a meaningful time period (4). Relative measures captured the strength of the association between an exposure and disease. Relative effects or associations tend to be more consistent across studies and populations than absolute measures, but what often tends to be the case may be irrelevant in a particular instance. In contrast, the absolute risk associated with an exposure is of greater interest than the relative risk. Authors should not only be aware but clearly report the methodology used to calculate attributable risks, ideally giving the formula used in determining a causal relationship between the risk factor and the disease due to the semantic ambiguity and complexities involved (329,330). In fact, a survey of abstracts of 222 articles published in leading medical journals found that in 62% of abstracts of randomized trials, including a ratio measure, absolute risks were given, but only in 21% of abstracts of cohort studies (331). In another evaluation with a free text search of Medline from 1966 to 1997, 619 items mentioned attributable risks in the title or abstract, compared to 18,955 using relative risk or odds ratio, for a ratio of 1 to 31 (332).

11.4.5 Other Analyses

In addition to the main analysis, other analyses are often performed to address specific subgroups, the potential interaction between risk factors, the calculation of attributable risks, or use alternative definitions of study variables in sensitivity analyses. There is debate about the dangers associated with subgroup analyses, and multiplicity of analyses in general (24,286-290). The STROBE statements (4,5) similar to CONSORT statements (21,22), believe that there is too great a tendency to look for evidence of subgroup-specific associations, or effect-measure modification, when overall results appear to suggest little or no effect. However, there is value in exploring whether an overall association appears consistent across several, preferably pre-specified subgroups, especially when a study is large enough to have sufficient data in each subgroup.

11.5 Discussion

The discussion section addresses the central issues of validity and meaning of the study (333). Many of the discussion sections have been found to be dominated by incomplete or biased assessments of the study's results and their implications, and the rhetoric supporting the authors' findings (334-336). Structuring the discussion helps avoid unwarranted speculation and other over-interpretation of results (337,338).

The Annals of Internal Medicine (335) recommends that authors structure the discussion section with systematic description of the following:

- A brief synopsis of the key findings.
- Consideration of possible mechanisms and explanation.
- Comparison with relevant findings from other published studies.
- Limitations of the present study and methods used to minimize and compensate for those limitations.
- 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 (339,340). 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.

Finally, the difference between statistical significance and clinical importance must be described. The section on research recommendations and the section on limitations of the study should be closely linked to each other (4). Further, it would be beneficial if investigators suggest ways in which subsequent research can improve on their studies rather than blindly stating more research is needed (341,342).

11.5.1 Key Results

The discussion should begin with a short summary of the main findings of the study. This short summary provides understanding subsequent interpretation and implications offered in the study supported by the findings.

11.5.2 Limitations

The identification and discussion of the limitations of a study are an essential part of the scientific reporting. It is important not only to identify the sources of bias and confounding that could have affected results, but also to discuss the relative importance of different biases, including the likely direction and magnitude of any potential bias (4). In discussion of limitations, the present study may compare the results with other studies in the literature in terms of validity, generalizability, and precision (343). Thus, each study can be viewed as a contribution to the literature, not as a stand-alone basis for inference and action. Surprisingly, a survey of authors who had published original research articles in *The Lancet* found that important weaknesses of the study were reported by the investigators in the survey questionnaires, but not in the published article (344).

11.5.3 Interpretation

One of the most important parts of the discussion section is the interpretation of a study's results. Overinterpretation is common and human. Even when one tries hard to give an objective assessment, reviewers often rightly point out that the authors went too far in some respects. Thus, when interpreting results, authors should carefully consider the nature of the study on the discovery to verification continuum and potential sources of bias, including loss to follow-up and non-participation. Further, due consideration should be given to confounding, the results of relevant sensitivity analyses, and to the issue of multiplicity and subgroup analyses. In this discussion, the existing external evidence from different types of studies, should always be included, but may be particularly important for studies reporting small increases in risk. Further, authors should put their results in context with similar studies and explain how the new study affects the existing body of evidence, ideally by referring to a systematic review.

11.5.4 Generalizability

Generalizability, also called external validity or applicability, is the extent to which the results of a study can be applied to other circumstances (344). However, there is no external validity *per se*, the term is meaningful only with regard to clearly specified conditions (345). Generalizability or external validity generally means results can be applied to an individual, groups, or populations that differ from those enrolled in the study with regard to age, sex, ethnicity, severity of disease, and co-morbid condition. Further, the nature and level of exposures and the definitions of outcomes are applicable to another setting or population. Data collected in longitudinal studies many years ago are still relevant today. Finally, results from health services research in one country are applicable to health systems in other countries.

The question of generalizability is often a matter of judgment that depends on the study setting, the characteristics of the participants, the exposures examined, and the outcomes assessed. Thus, it is a crucial opportunity for authors to provide adequate information about the setting and locations, eligibility criteria, the exposures and how they were measured, the definition of outcomes, and the period of recruitment and follow-up.

11.6 Funding

Authors should provide the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based (4). Some journals require authors to disclose the presence or absence of financial and other conflicts of interest (346,347). It is also essential in evaluation of methodologic quality assessment of observational studies to evaluate the role of funding and conflicts of interest (42). However, strong associations and conflicts of interest have been demonstrated, not only in conclusions of research articles, but also in preparation of guidelines, systematic reviews, etc. (50,56-59,62,128,139,348-353). Authors or funders may have conflicts of interest that influence any of the following: the design of the study (354), choice of exposures (354,355), outcomes (356), statistical methods (357), and selective publication of outcomes (245) and studies (358). Thus, if the funding was provided, the role of the funders should be described in detail: in what part of the study they took direct responsibility (i.e., design, data collection, analysis, drafting of manuscript, decision to publish) (346). Further, other sources of undue influence include employers such as university administrators for academic researchers, advisory committees, litigants, and special interest groups.

12. Discussion

This manuscript describes multiple concepts of observational studies, including advantages, disadvantages, design, and reporting. Even though, randomized trials are considered to be the gold standard and N of 1 RCT is considered to be at the top of the hierarchy of strength of evidence, observational evidence continues to play an important role. As Greene (17) proposes, instead of wondering whether observational studies are just as effective as RCTs, we should consider observational studies and RCTs as expressions in the setting of modern clinical research of the steps of observation and experimentation that form the basis of the scientific methodology. Thus, they are complementary rather than contradictory. Since both observation and experimentation steps are required for scientific advancement, it seems misplaced to argue that one is more effective than the other (17). Much of the literature in interventional pain management and surgery is based on observational studies. Even then, the reporting of observational studies is often of insufficient quality with poor reporting which hampers the assessment of the strengths and weaknesses of the study and the generalizability of its results. In interventional pain management settings, results from clinical trials, both randomized and observational, with substantial impact on patient care, have been proven ineffective based on flawed methodology and evidence synthesis. There is also empirical evidence that some RCTs have biased results and in some cases, there was no difference between observational studies and randomized trials. The poorly executed observational studies tend to exaggerate treatment effects and to have important biases, which is not limited to only observational studies but extends to randomized trials also. Consequently, it is of paramount importance to produce high-quality research, which consistently eliminates bias and shows significant ES in interventional pain management. The design, implementation, and report of observational studies requires methodologic as well as clinical expertise and discipline, 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 minimize the bias and association of effect and risk. Sound reporting encompasses adequate reporting and the conduct of ethical studies rests on the footing of the sound signs, which will not subject readers to speculation. To improve the reporting of observational research, the STROBE statement (5) and explanation and elaboration (4) have been developed which relate to title, abstract, introduction, methods, results, and discussion sections of articles. STROBE provides general reporting recommendations for descriptive observational studies and studies that investigate associations between exposures and health outcomes. The STROBE statement tends to provide helpful recommendations for reporting observational studies in epidemiology. Consequently, interventional pain specialists must focus on the differences between multiple types of trials — cohort studies, case-control studies, and cross-sectional studies. Further, it is essential to describe the rationale, objectives, study design, outcomes, and results without creativity.

13. Conclusion

In conclusion, observational studies and randomized trials are complementary and both are required for the development of scientific evidence, specifically in interventional pain management.

Acknowledgments

The authors wish to thank Vidyasagar Pampati, MSc, statistician; Sekar Edem for assistance in search of literature; and Tonie M. Hatton and Diane E. Neihoff, transcriptionists, for their assistance in the preparation of this manuscript.

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. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Ann Intern Med* 2007; 147: W163-W194.
5. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Ann Intern Med* 2007; 147:573-577.
6. Healy D. Randomized Controlled Trials: Evidence Biased Psychiatry. Alliance for Human Research Protection, 2002. www.ahrp.org/COI/healyo802.php
7. Hotopf M. The pragmatic randomized controlled trial. *Adv Psychiatr Treat* 2002; 8:326-333.
8. Williams DD, Garner J. The case against "the evidence": A different perspective on evidence-based medicine. *Br J Psychiatry* 2002; 180:8-12.
9. 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.
10. Thornley B, Adams C. Content and quality of 2,000 controlled trials in schizophrenia over 50 years. *BMJ* 1998; 317:1181-1184.
11. Hotopf M, Churchill R, Lewis G. Pragmatic randomized controlled trials in psychiatry. *Br J Psychiatry* 1999; 175:217-223.
12. 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.
13. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004; 328:39-41.
14. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; 312:1215-1218.
15. Funai EF, Rosenbush EJ, Lee MJ, Del Priore G. Distribution of study designs in four major US journals of obstetrics and gynecology. *Gynecol Obstet Invest* 2001; 51:8-11.
16. Scales CD Jr., Norris RD, Peterson BL, Preminger GM, Dahm P. Clinical research and statistical methods in the urology literature. *J Urol* 2005; 174:1374-1379.
17. Greene T. Are observational studies "just as effective" as randomized clinical trials? *Blood Purif* 2000; 18:317-322.
18. Papanikolaou PN, Christidi GD, Ioannidis JP. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ* 2006; 174:635-641.
19. Hempel CG. *Philosophy of Natural Science*. Prentice-Hall, Englewood Cliffs, 1966.
20. Popper KR. *The Logic of Scientific Discovery*. Cambridge University Press, London, 1958.
21. 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.
22. 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.
23. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG; International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003; 7:1-173.
24. Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, Kalish LA, Kasten LE, McCormack VA. Issues in the reporting of epidemiological studies: A survey of recent practice. *BMJ* 2004; 329:883.
25. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001; 323:42-46.
26. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998; 316:140-144.
27. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; 283:2008-2012.
28. Bero LA, Jadad AR. How consumers and policy makers can use systematic reviews for decision making. *Ann Intern Med* 1997; 127:37-42.
29. Berlin JA, Rennie D. Measuring the quality of trials. *JAMA* 1999; 282:1083-1085.
30. Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol* 1999; 28:1-9.
31. Radford MJ, Foody JM. How do observational studies expand the evidence base for therapy? *JAMA* 2001; 286:1228-1230.
32. Lee W, Bindman J, Ford T, Glozier N, Moran P, Stewart R, Hotopf M. Bias in psychiatric case-control studies: Literature survey. *Br J Psychiatry* 2007; 190:204-209.
33. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. *Am J Epidemiol* 2005; 161:280-288.
34. Bogardus ST Jr., Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research: The need for methodological standards. *JAMA* 1999; 281:1919-1926.
35. Verhalen RD. Guidelines for documentation of epidemiologic studies. *Am J Epidemiol* 1981; 114:609-613.
36. Rennie D. CONSORT revised—improving the reporting of randomized trials. *JAMA* 285:2006-2007.
37. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D; CONSORT group; Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: An extension of the CONSORT statement. *BMJ* 2008; 337:1223-1226.
38. Manchikanti L, Hirsch JA, Smith HS. Evidence-based medicine, systematic reviews, and guidelines in interventional

pain management: Part 2: Randomized controlled trials. *Pain Physician* 2008; 11:717-773.

39. Peipert JF, Phipps MG. Observational studies. *Clin Obstet Gynecol* 1998; 41:235-244.
40. Ioannidis JP, Lau J. Pooling research results. *Jt Comm J Qual Improv* 1999; 25:462-469.
41. Lipsett M, Campleman S. Occupational exposure to diesel exhaust and lung cancer: A metaanalysis. *Am J Public Health* 1999; 89:1009-1017.
42. 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
43. 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.
44. 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.
45. 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.
46. 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.
47. Vickers A, Cassileth B, Ernst E, Fisher P, Goldman P, Jonas W, Kang SK, Lewith G, Schulz K, Silagy C. How should we research unconventional therapies? A panel report from the Conference on Complementary and Alternative Medicine Research Methodology, National Institutes of Health. *Int J Technol Assess Health Care* 1997; 13:111-121.
48. Mann CC. Can meta-analysis make policy? *Science* 1994; 266:960-962.
49. World Health Organization. *International Clinical Trials Registry Platform (ICTRP)*. 2007. www.who.int/entity/ictrp/en/
50. 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.
51. Manchikanti L, Pampati V, Rivera JJ, Beyer CD, Damron KS, Barnhill RC. Caudal epidural injections with Sarapin or steroids in chronic low back pain. *Pain Physician* 2001; 4:322-335.
52. 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.
53. 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.
54. 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.
55. 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.
56. 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.
57. 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.
58. 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.
59. 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.
60. Barnsley L. Percutaneous radiofrequency neurotomy for chronic neck pain: Outcomes in a series of consecutive patients. *Pain Med* 2005; 6:282-286.
61. Sapir D, Gorup JM. Radiofrequency medial branch neurotomy in litigant and nonlitigant patients with cervical whiplash. *Spine* 2001; 26:E268-E273.
62. Manchikanti L, Singh V, Vilims BD, Hansen HC, Schultz DM, Kloth DS. Medial branch neurotomy in management of chronic spinal pain: Systematic review of the evidence. *Pain Physician* 2002; 5:405-418.
63. Dreyfuss P, Halbrook B, Pauza K, Joshi A, McLarty J, Bogduk N. Efficacy and validity of radiofrequency neurotomy for chronic lumbar zygapophysial joint pain. *Spine* 2000; 25:1270-1277.
64. Lutz GE, Vad VB, Wisneski RJ. Fluoroscopic transforaminal lumbar epidural steroids: An outcome study. *Arch Phys Med Rehabil* 1998; 79:1362-1366.
65. Buttermann GR. The effect of spinal steroid injections for degenerative disc disease. *Spine J* 2004; 4:495-505.
66. Buttermann GR. Treatment of lumbar disc herniation: Epidural steroid injection compared with discectomy. A prospective, randomized study. *J Bone Joint Surg Am* 2004; 86-A:670-679.
67. Botwin KP, Gruber RD, Bouchlas CG, Torres-Ramos FM, Sanelli JT, Freeman ED, Slaten WK, Rao S. Fluoroscopically guided lumbar transformational epidural steroid injections in degenerative lumbar stenosis: An outcome study. *Am J Phys Med Rehabil* 2002; 81:898-905.
68. Bush K, Hillier S. Outcome of cervical radiculopathy treated with periradicular/epidural corticosteroid injections: A prospective study with independent clinical review. *Eur Spine J* 1996; 5:319-325.
69. Gerdesmeyer L, Lampe R, Veihelmann A, Burgkart R, Gobel M, Gollwitzer H, Wagner K. Chronic radiculopathy. Use of minimally invasive percutaneous epidural neurolysis according to Racz. *Der Schmerz* 2005; 19:285-295.
70. Igarashi T, Hirabayashi Y, Seo N, Saitoh K, Fukuda H, Suzuki H. Lysis of adhesions and epidural injection of steroid/local anesthetic during epiduroscopy potentially alleviate low back and leg pain in elderly patients with lumbar spine stenosis. *Br J Anaesth* 2004; 93:181-187.

71. Geurts JW, Kallewaard JW, Richardson J, Groen. Targeted methylprednisolone acetate/hyaluronidase/clonidine injection after diagnostic epiduroscopy for chronic sciatica: A prospective, 1-year follow-up study. *Reg Anesth Pain Med* 2002; 27:343-352.
72. McLeod RS. Issues in surgical randomized controlled trials. *World J Surg* 1999; 23:1210-1214.
73. Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized controlled trials in surgery. *Surgery* 1994; 115:707-712.
74. Hardin WD Jr, Stylianos S, Lally KP. Evidence-based practice in pediatric surgery. *J Pediatr Surg* 1999; 34:908-912.
75. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: Problems and possible solutions. *BMJ* 2002; 324:1448-1451.
76. Manchikanti L, Cash KA, McManus CD, Pampati V, Smith HS. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 1. Discogenic pain without disc herniation or radiculitis. *Pain Physician* 2008; 11:785-800.
77. Manchikanti L, Singh V, Cash KA, Pampati V, Damron KS, Boswell MV. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 2. Disc herniation and radiculitis. *Pain Physician* 2008; 11:801-815.
78. Manchikanti L, Singh V, Cash KA, Pampati V, Datta S. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 3. Post surgery syndrome. *Pain Physician* 2008; 11:817-831.
79. Manchikanti L, Cash KA, McManus CD, Pampati V, Abdi S. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 4. Spinal stenosis. *Pain Physician* 2008; 11:833-848.
80. 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.
81. 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.
82. 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.
83. 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.
84. Manchikanti L, Manchikanti K, Manchukonda R, Cash KA, Damron KS, Pampati V, McManus CD. Evaluation of lumbar facet joint nerve blocks in the management of chronic low back pain: A preliminary report of a randomized, double-blind controlled trial. Clinical Trial NCT00355914. *Pain Physician* 2007; 10:425-440.
85. 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].
86. 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.
87. 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.
88. 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.
89. 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.
90. 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.
91. 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.
92. Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? *Surgery* 1995; 118:459-467.
93. 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.
94. 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.
95. 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.
96. 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.
97. Kao LS, Tyson JE, Blakely ML, Lally KP. Clinical research methodology I: Introduction to randomized trials. *J Am Coll Surg* 2008; 206:361-369.
98. 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.
99. 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.
100. 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.

101. 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.
102. 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.
103. 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.
104. 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.
105. 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.
106. 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.
107. 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.
108. 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.
109. 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.
110. 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.
111. 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.
112. 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.
113. MacPherson H. Pragmatic clinical trials. *Complement Ther Med* 2004; 12:136-140.
114. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982; 72:233-240.
115. 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.
116. 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; 18:746-751.
117. 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.
118. Kane RL. Approaching the outcomes question. In: Kane RL (ed). *Understanding Health Care Outcomes Research*. Aspen Publications, Gaithersburg, 1997, pp 1-15.
119. Manchikanti L, Pampati V. Research designs in interventional pain management: Is randomization superior, desirable or essential? *Pain Physician* 2002; 5:275-284.
120. 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.
121. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000; 342:1878-1886.
122. Hartz A, Benson K, Glaser J, Bentler S, Bhandari M. Assessing observational studies of spinal fusion and chemonucleolysis. *Spine* 2003; 28:2268-2275.
123. 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.
124. 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.
125. American College of Occupational and Environmental Medicine. Low Back Disorders Chapter. 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, Elk Grove Village, 2007.
126. American College of Occupational and Environmental Medicine. Chronic Pain Chapter (revised 2008). 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, Elk Grove Village, Epublished August 14, 2008.
127. 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.
128. 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.
129. Letter to Robert K. McLellan, MD, President of the American College of Occupational & Environmental Medicine (ACOEM) from the Honorable Bart Stupak and Ed Whitfield, U.S. House of Representatives, January 25, 2008.
130. Chou R. Using evidence in pain practice: Part I: Assessing quality of systematic reviews and clinical practice guidelines. *Pain Med* 2008; 9:518-530.
131. Chou R. Using evidence in pain practice: Part II: Interpreting and applying systematic reviews and clinical practice guidelines. *Pain Med* 2008; 9:531-541.
132. Koes BW, Scholten RJ, Mens JM, Bout-

er LM. Efficacy of epidural steroid injections for low-back pain and sciatica: A systematic review of randomized clinical trials. *Pain* 1995; 63:279-288.

133. 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.
134. 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.
135. Nelemans PJ, Debie RA, DeVet HC, Sturmans F. Injection therapy for subacute and chronic benign low back pain. *Spine* 2001; 26:501-515.
136. 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.
137. 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.
138. 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.
139. 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.
140. 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.
141. 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.
142. 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.
143. Staal JB, de Bie R, de Vet HC, Hildebrandt J, Nelemans P. Injection therapy for subacute and chronic low-back pain. *Cochrane Database Syst Rev* 2008; 3: CD001824.
144. Peloso PMJ, Gross A, Haines T, Trinh K, Goldsmith CH, Burnie SJ; Cervical Overview Group. Medicinal and injection therapies for mechanical neck disorders. *Cochrane Database Syst Rev* 2007; 3:CD000319.
145. Manchikanti L, Singh V, Helm II S, Hirsch JA. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 3: Systematic reviews and meta-analysis of randomized trials. *Pain Physician* 2009; 12:35-72.
146. Manchikanti L, Boswell MV, Giordano J. Evidence-based interventional pain management: Principles, problems, potential, and applications. *Pain Physician* 2007; 10:329-356.
147. Manchikanti L. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 1: Introduction and general considerations. *Pain Physician* 2008; 11:161-186.
148. Manchikanti L, Heavner J, Racz GB, Mekhail NA, Schultz DM, Hansen HC, Singh V. Methods for evidence synthesis in interventional pain management. *Pain Physician* 2003; 6:89-111.
149. Cho MK, Bero LA. Instruments for assessing the quality of drug studies published in the medical literature. *JAMA* 1994; 272:101-104.
150. Goodman SN, Berlin J, Fletcher SW, Fletcher RH. Manuscript quality before and after peer review and editing at Annals of Internal Medicine. *Ann Intern Med* 1994; 121:11-21.
151. 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.
152. Corrao G, Bagnardi V, Zambon A, Arico S. Exploring the dose-response relationship between alcohol consumption and the risk of several alcohol-related conditions: A metaanalysis. *Addiction* 1999; 94:1551-1573.
153. Khan KS, Ter Riet G, Popay J, Nixon J, Kleijnen J. Undertaking systematic reviews of research effectiveness. CRD's guidance for those carrying out or commissioning reviews. CRD Report number 4 (2nd edn). 2001. The University of York Centre for Reviews and Dissemination.
154. New Zealand Guidelines Group. Tools for Guideline Development & Evaluation. www.nzgg.org.nz/
155. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001; 323:334-336.
156. Reisch JS, Tyson JE, Mize SG. Aid to the evaluation of therapeutic studies. *Pediatrics* 1989; 84:815-827.
157. Spitzer WO, Lawrence V, Dales R, Hill G, Archer MC, Clark P, Abenhaim L, Hardy J, Sampalis J, Pinfold SP. Links between passive smoking and disease: A best-evidence synthesis. A report of the Working Group on Passive Smoking. *Clin Invest Med* 1990; 13:17-42.
158. National Health and Medical Research Council (NHMRC). How to Review the Evidence: Systematic Identification and Review of the Scientific Literature. NHMRC, Canberra, Australia, 2000.
159. Zaza S, Wright-De Agüero LK, Briss PA, Truman BI, Hopkins DP, Hennessy MH, Sosin DM, Anderson L, Carande-Kulis VG, Teutsch SM, Pappaioanou M. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Task Force on Community Preventive Services. *Am J Prev Med* 2000; 18:44-74.
160. Ariens GA, van Mechelen W, Bongers PM, Bouter LM, van der Wal G. Physical risk factors for neck pain. *Scand J Work Environ Health* 2000; 26:7-19.
161. Lohr KN, Carey TS. Assessing "best evidence": Issues in grading the quality of studies for systematic reviews. *Jt Comm J Qual Improv* 1999; 25:470-479.
162. Carruthers SG, Larochelle P, Haynes RB, Petrasovits A, Schiffrin EL. Report of the Canadian Hypertension Society Consensus Conference: 1. Introduction. *CMAJ*

1993; 149:289-293.

163. Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. *JAMA* 1994; 272:234-237.
164. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. *JAMA* 1994; 271:1615-1619.
165. Angelillo IF, Villari P. Residential exposure to electromagnetic fields and childhood leukaemia: A meta-analysis. *Bull World Health Organ* 1999; 77:906-915.
166. 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.
167. 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.
168. Atluri S, Datta S, Falco FJE, Lee M. Systematic review of diagnostic utility and therapeutic effectiveness of thoracic facet joint interventions. *Pain Physician* 2008; 11:611-629.
169. Conn A, Buenaventura R, Datta S. Abdi S, Diwan S. Systematic review of caudal epidural injections in the management of chronic low back pain. *Pain Physician* 2009; 12:109-135.
170. Parr AT, Diwan S, Abdi S. Lumbar interlaminar epidural injections in managing chronic low back and lower extremity pain: A systematic review. *Pain Physician* 2009; 12:163-188.
171. Benyamin R, Singh V, Parr AT, Conn A, Diwan S, Abdi S. Systematic review of the effectiveness of cervical epidurals in the management of chronic neck pain. *Pain Physician* 2009; 12:137-157.
172. Helm S, Hayek S, Benyamin R, Manchikanti L. Systematic review of effectiveness of thermal annular procedures in treating discogenic low back pain. *Pain Physician* 2009; 12:207-232.
173. 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.
174. Avis M. Reading research critically. II. An introduction to appraisal: Assessing the evidence. *J Clin Nurs* 1994; 3:271-277.
175. DuRant RH. Checklist for the evaluation of research articles. *J Adolesc Health* 1994; 15:4-8.
176. Elwood M. Forward projection—using critical appraisal in the design of studies. *Int J Epidemiol* 2002; 31:1071-1073.
177. Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the quality of evidence for clinical practice guidelines. *J Clin Epidemiol* 1996; 49:749-754.
178. Zola P, Volpe T, Castelli G, Sismondi P, Nicolucci A, Parazzini F, Liberati A. Is the published literature a reliable guide for deciding between alternative treatments for patients with early cervical cancer? *Int J Radiat Oncol Biol Phys* 1989; 16:785-797.
179. Borghouts JA, Koes BW, Bouter LM. The clinical course and prognostic factors of non-specific neck pain: A systematic review. *Pain* 1998; 77:1-13.
180. Loney PL, Chambers LW, Bennett KJ, Roberts JG, Stratford PW. Critical appraisal of the health research literature: Prevalence or incidence of a health problem. *Chronic Dis Canada* 2000; 19:170-177.
181. Littenberg B, Weinstein LP, McCarren M, Mead T, Swiontkowski MF, Rudicel SA, Heck D. Closed fractures of the tibial shaft. A meta-analysis of three methods of treatment. *J Bone Joint Surg Am* 1998; 80:174-183.
182. Macfarlane TV, Glenny AM, Worthington HV. Systematic review of population-based epidemiological studies of orofacial pain. *J Dent* 2001; 29:451-467.
183. Margetts BM, Thompson RL, Key T, Duffy S, Nelson M, Bingham S, Wiseman M. Development of a scoring system to judge the scientific quality of information from case-control and cohort studies of nutrition and disease. *Nutr Cancer* 1995; 24:231-239.
184. Rangel SJ, Kelsey J, Colby CE, Anderson J, Moss RL. Development of a quality assessment scale for retrospective clinical studies in pediatric surgery. *J Pediatr Surg* 2003; 38:390-396.
185. van der Windt DAWM, Thomas E, Pope DP, de Winter AF, Macfarlane GJ, Bouter LM, Silman AJ. Occupational risk factors for shoulder pain: A systematic review. *Occup Environ Med* 2000; 57:433-442.
186. Cowley DE. Prostheses for primary total hip replacement. A critical appraisal of the literature. *Int J Technol Assess Health Care* 1995; 11:770-778.
187. Fowkes FG, Fulton PM. Critical appraisal of published research: introductory guidelines. *BMJ* 1991; 302:1136-1140.
188. Gyorkos TW, Tannenbaum TN, Abrahamowicz M, Oxman AD, Scott EA, Millson ME, Rasooly I, Frank JW, Riben PD, Mathias RG. An approach to the development of practice guidelines for community health interventions. *Can J Public Health* 1994; 85:S8-S13.
189. Steinberg EP, Eknoyan G, Levin NW, Eschbach JW, Golper TA, Owen WF, Schwab S. Methods used to evaluate the quality of evidence underlying the National Kidney Foundation-Dialysis Outcomes Quality Initiative Clinical Practice Guidelines: Description, findings, and Implications. *Am J Kidney Dis* 2000; 36:1-11.
190. Mihailovic A, Bell CM, Urbach DR. Users' guide to the surgical literature. Case-control studies in surgical journals. *Can J Surg* 2005 48:148-151.
191. Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine." *Am J Med* 1997; 103:529-535.
192. Bradley C. Designing medical and educational intervention studies: A review of some alternatives to conventional randomized controlled trials. *Diabetes Care* 1993; 16:509-519.
193. Torgerson D, Klaber-Moffett J, Russell I. Patient preferences in randomized trials: Threat or opportunity. *J Health Serv Res Policy* 1996; 1:194-197.
194. Browner WS. Methods. In: *Publishing and Presenting Clinical Research*. Lippincott Williams & Wilkins, Philadelphia, 2006, pp 27-44.
195. Riegelman RK. Types of studies and the M.A.A.R.I.E. Framework. In: *Studying a Study & Testing a Test. How to Read the Medical Evidence,* 5th edition. Lippincott Williams & Wilkins, Philadelphia, 2005, pp 7-15.
196. Including non-randomized studies. In: Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.1 Cochrane Collaboration, Oxford, 2008.
197. Ioannidis JP. Effect of formal statistical significance on the credibility of observational associations. *Am J Epidemiol* 2008; 168:374-383.
198. Altman DG, De Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. *Br J Cancer* 1995; 72:511-518.
199. Phillips MR, Yang G, Zhang Y, Wang L,

Ji H, Zhou M. Risk factors for suicide in China: A national case-control psychological autopsy study. *Lancet* 2002; 360:1728-1736.

200. Pasquale LR, Kang JH, Manson JE, Willett WC, Rosner BA, Hankinson SE. Prospective study of type 2 diabetes mellitus and risk of primary open-angle glaucoma in women. *Ophthalmology* 2006; 113:1081-1086.
201. Craig SL, Feinstein AR. Antecedent therapy versus detection bias as causes of neoplastic multimorbidity. *Am J Clin Oncol* 1999; 22:51-56.
202. Rogler LH, Mroczek DK, Fellows M, Loftus ST. The neglect of response bias in mental health research. *J Nerv Ment Dis* 2001; 189:182-187.
203. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: A survey of practice. *Am J Epidemiol* 2006; 163:197-203.
204. 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.
205. Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol* 2000; 136:381-385.
206. 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.
207. 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.
208. 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.
209. 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.
210. Vandenbroucke JP. Prospective or retrospective: What's in a name? *BMJ* 1991; 302:249-250.
211. Last JM. *A Dictionary of Epidemiology*, 4th Ed. Oxford University Press, New York, 2000.
212. Miettinen OS. *Theoretical Epidemiology: Principles of Occurrence Research in Medicine*. John Wiley & Sons, New York, 1985.
213. Rothman KJ, Greenland S. Types of Epidemiologic Studies. In: Rothman KJ, Greenland S (eds). *Modern Epidemiology*, 2nd ed. Lippincott Williams & Wilkins, Philadelphia, 1998, pp 74-75.
214. MacMahon B, Trichopoulos D. *Epidemiology, Principles and Methods,* 2nd ed. Little Brown and Company, Boston, 1996.
215. Lilienfeld AM. *Foundations of Epidemiology.* Oxford University Press, New York, 1976.
216. Clarke M, Oxman AD, editors. Cochrane Reviewers Handbook 4.1.4 [updated October 2001]. In: The Cochrane Library, Issue 4, 2001. Oxford: Update Software. Updated quarterly.
217. Green SB, Byar DP. Using observational data from registries to compare treatments: The fallacy of omnimetrics. *Stat Med* 1984; 3:361-373.
218. Salas M, Hofman A, Stricker BH. Confounding by indication: An example of variation in the use of epidemiologic terminology. *Am J Epidemiol* 1999; 149:981-983.
219. Horwitz RI, Feinstein AR. The application of therapeutic trial principles to improve the design of epidemiologic research. *J Chron Dis* 1981; 34:575-583.
220. Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990; 89:630-638.
221. Costanza MC. Matching. *Prev Med* 1995; 24:425-433.
222. Walker M, Whincup PH, Shaper AG. The British Regional Heart Study 1975-2004. *Int J Epidemiol* 2004; 33:1185-1192.
223. Wieland S, Dickersin K. Selective exposure reporting and Medline indexing limited the search sensitivity for observational studies of the adverse effects of oral contraceptives. *J Clin Epidemiol* 2005; 58:560-567.
224. Mann CJ. Observational research methods. Research design II: Cohort, cross sectional, and case-control studies. *Emerg Med J* 2003; 20:54-60.
225. Setia MS. Observational studies: How to go about them? *Indian J Dermatol Venereol Leprol* 2008; 74:288-291.
226. Lerner DJ, Kannel WB. Patterns of coronary heart disease morbidity and mortality in the sexes: A 26 year follow-up of the Framingham population. *Am Heart J* 1986; 111:383-390.
227. Doll R, Peto H. Mortality in relation to smoking. 40 years observation on female British doctors. *BMJ* 1989;208:967-973.
228. Alberman ED, Butler NR, Sheridan MD. Visual acuity of a national sample (1958 cohort) at 7 years. *Dev Med Child Neurol* 1971;13:9-14.
229. Smith GD, Hart C, Blane D, Hole D. Adverse socioeconomic conditions in childhood and cause specific mortality: Prospective observational study. *BMJ* 1998; 316:1631-1635.
230. Goyder EC, Goodacre SW, Botha JL, Bodiwala GG. How do individuals with diabetes use the accident and emergency department? *J Accid Emerg Med* 1997; 14:371-374.
231. Karjalainen J, Kujala U, Kaprio J, Sarna S, Viitasalo M. Lone atrial fibrillation in vigorously exercising middle aged men: case-control study. *BMJ* 1998; 316:1784-1785.
232. 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.
233. 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.
234. Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine* 2000; 25:2940-2952.
235. 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.
236. Vernon H, Mior S. The Neck Disability Index: A study of reliability and validity. *J Manipulative Physiol Ther* 1991; 14:409-415.
237. 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.
238. Melzack R. The McGill Pain Questionnaire: Major properties and scoring methods. *Pain* 1975; 1:277-299.

239. 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.
240. 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.
241. 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.
242. Turk DC. Statistical significance and clinical significance are not synonyms! *Clin J Pain* 2000; 16:185-187.
243. 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.
244. 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.
245. 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.
246. Turk DC. Clinical effectiveness and cost-effectiveness of treatments for patients with chronic pain. *Clin J Pain* 2002; 18:355-365.
247. 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.
248. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10:407-415.
249. 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.
250. Goodman SN. Toward evidence-based medical statistics. 1: The *P* value fallacy. *Ann Intern Med* 1999; 130:995-1004.
251. 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.
252. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; 311:485.
253. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994; 272:122-124.
254. 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.
255. Scott PA, Kingsley GH, Smith CM, Choy EH, Scott DL. Non-steroidal anti-inflammatory drugs and myocardial infarctions: Comparative systematic review of evidence from observational studies and randomised controlled trials. *Ann Rheum Dis* 2007; 66:1296-1304.
256. Eng J. Sample size estimation: How many individuals should be studied? *Radiology* 2003; 227:309-313.
257. Woodward M. Formulas for sample-size, power and minimum detectable relative risk in medical studies. *Statistician* 1992; 41:185-196.
258. 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.
259. 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.
260. Rigby AS, Vail A. Statistical methods in epidemiology. II: A common sense approach to sample size estimation. *Disabil Rehabil* 1998; 20:405-410.
261. Schulz KF, Grimes DA. Sample size calculations in randomized trials: Mandatory and mystical. *Lancet* 2005; 365:1348-1353.
262. Drescher K, Timm J, Jockel KH. The design of case-control studies: The effect of confounding on sample size requirements. *Stat Med* 1990; 9:765-776.
263. Devine OJ, Smith JM. Estimating sample size for epidemiologic studies: The impact of ignoring exposure measurement uncertainty. *Stat Med* 1998; 17:1375-1389.
264. Olsen J, Basso O. Re: Residual confounding. *Am J Epidemiol* 1999; 149:290.
265. Slama R, Werwatz A. Controlling for continuous confounding factors: non- and semiparametric approaches. *Rev Epidemiol Sante Publique* 2005; 53:2S65-S280.
266. Greenland S. Introduction to regression modelling. In: Rothman KJ, Greenland S (eds). *Modern Epidemiology*. 2nd ed. Lippincott Raven, Philadelphia 1998, pp 401-432.
267. Thompson WD. Statistical analysis of case-control studies. *Epidemiol Rev* 1994; 16:33-50.
268. Schlesselman JJ. *Logistic Regression for Case-Control Studies. Case-Control Studies Design, Conduct, Analysis.* Oxford University Press, New York, 1982, pp 235-241.
269. Vickers AJ. Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. *BMC Med Res Methodol* 2005; 5:35.
270 Altman DG. *Practical Statistics for Medical Research*. Chapman and Hall, London, 1991.
271. Jekel JF, Katz DL, Elmore JG. *Epidemiology, Biostatistics and Preventive Medicine*. WB Saunders Company, Philadelphia, 2001.
272. 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.
273. 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.
274. 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.
275. 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 ap-

plied research. *J Clin Epidemiol* 1999; 52:229-235.

276. Senn S. *Statistical Issues in Drug Development*. John Wiley & Sons Ltd, Chichester, 1997.
277. 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.
278. Feinstein AR. *P*-values and confidence intervals: Two sides of the same unsatisfactory coin. *J Clin Epidemiol* 1998; 51:355-360.
279. Hoffman RG, Lim JH. Observational study design. In: Ambrosius WT (ed). *Methods in Molecular Biology, vol 404: Topics in Biostatistics*. Humana Press Inc., Totowa, NJ, 2007, pp 19-31.
280. Fisher R. *Statistical Methods and Scientific Inference*. 3rd ed. Macmillan, New York, 1973.
281. Freeman PR. The role of *p*-values in analysing trial results. *Stat Med* 1993; 12:1443-1552.
282. 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.
283. Altman DG. Confidence intervals in research evaluation. *Ann Intern Med* 1992; 116:A28-A29.
284. Simon R. Confidence intervals for reporting results of clinical trials. *Ann Intern Med* 1986; 105:429-435.
285. Kahn HA, Sempos CT. *Statistical Methods in Epidemiology*. Oxford University Press, Oxford, 1990.
286. Gøtzsche PC. Believability of relative risks and odds ratios in abstracts: Cross sectional study. *BMJ* 2006; 333: 231-234.
287. 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.
288. Rothwell PM. Treating individuals 2. Subgroup analysis in randomized controlled trials: Importance, indications, and interpretation. *Lancet* 2005; 365:176-186.
289. 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.
290. Lagakos SW. The challenge of subgroup analyses — reporting without distorting. *N Engl J Med* 2006; 354:1667-1669.
291. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003; 138:40-44.
292. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med* 1990; 113:69-76.
293. Hartley J, Sydes M. Which layout do you prefer? An analysis of readers' preferences for different typographic layouts of structured abstracts. *J Inf Sci* 1996; 22:27-37.
294. 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.
295. Hartley J, Sydes M, Blurton A. Obtaining information accurately and quickly: Are structured abstracts more efficient? *J Inf Sci* 1996; 22:349-356.
296. American Journal of Epidemiology. Information for authors. www.oxfordjournals.org/aje/for_authors/index.html
297. Sturmer T, Brenner H. Flexible matching strategies to increase power and efficiency to detect and estimate gene-environment interactions in case-control studies. *Am J Epidemiol* 2002; 155:593-602.
298. Becher H. The concept of residual confounding in regression models and some applications. *Stat Med* 1992; 11:1747-1758.
299. Brenner H, Blettner M. Controlling for continuous confounders in epidemiologic research. *Epidemiology* 1997; 8:429-434.
300. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 1994; 86:829-835.
301. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: A bad idea. *Stat Med* 2006; 25:127-141.
302. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol* 1991; 134:895-907.
303. Little RJ, Rubin DB. A taxonomy of missing-data methods. In: *Statistical Analysis with Missing Data*, 2nd Edition. John Wiley & Sons, Hoboken, 2002, pp 19-23.
304. Rubin DB. Inference and missing data. *Biometrika* 1976; 63:581-592.
305. Rothman KJ, Greenland S. Basic methods for sensitivity analysis and external adjustment. In: Rothman KJ, Greenland S (eds). *Modern Epidemiology*, 2nd ed. Lippincott Williams & Wilkins, Philadelphia, 1998, pp 343-357.
306. Custer B, Longstreth WT Jr, Phillips LE, Koepsell TD, Van Belle G. Hormonal exposures and the risk of intracranial meningioma in women: A population-based case-control study. *BMC Cancer* 2006; 6:152.
307. Dunn NR, Arscott A, Thorogood M. The relationship between use of oral contraceptives and myocardial infarction in young women with fatal outcome, compared to those who survive: Results from the MICA case-control study. *Contraception* 2001; 63:65-69.
308. Wakefield MA, Chaloupka FJ, Kaufman NJ, Orleans CT, Barker DC, Ruel EE. Effect of restrictions on smoking at home, at school, and in public places on teenage smoking: cross sectional study. *BMJ* 2000; 321:333-337.
309. Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: A case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc* 2003; 98:47-54.
310. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003; 14:451-458.
311. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 2003; 14:459-466.
312. Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther* 1995; 57:6-15.
313. 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.
314. Ruiz-Canela M, Martınez-Gonzalez MA, de Irala-Estevez J. Intention to treat analysis is related to methodological quality. *BMJ* 2000; 320:1007-1008.
315. Armitage P, Colton T. *Encyclopedia of Biostatistics*. John Wiley & Sons, New York, 1998.
316. Committee for Proprietary Medicinal

Products (CPMP). Points to Consider on Missing Data: The European Agency for the Evaluation of Medicinal Products. London, 2001.

317. Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet* 2002; 359:431-434.
318. Hay AD, Wilson A, Fahey T, Peters TJ. The duration of acute cough in pre-school children presenting to primary care: A prospective cohort study. *Fam Pract* 2003; 20:696-705.
319. Dales LG, Ury HK. An improper use of statistical significance testing in studying covariables. *Int J Epidemiol* 1978; 7:373-375.
320. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. *Am J Epidemiol* 1993; 138:923-936.
321. Rothman KJ, Greenland S. Precision and validity in epidemiologic studies. In: Rothman KJ, Greenland S (eds). *Modern Epidemiology,* 2nd ed. Lippincott Williams & Wilkins, Philadelphia, 1998, pp 120-125.
322. Mastrangelo G, Fedeli U, Fadda E, Valentini F, Agnesi R, Magarotto G, Marchi T, Buda A, Pinzani M, Martines D. Increased risk of hepatocellular carcinoma and liver cirrhosis in vinyl chloride workers: Synergistic effect of occupational exposure with alcohol intake. *Environ Health Perspect* 2004; 112:1188-1192.
323. Salo PM, Arbes SJ Jr, Sever M, Jaramillo R, Cohn RD, London SJ, Zeldin DC. Exposure to Alternaria alternata in US homes is associated with asthma symptoms. *J Allergy Clin Immunol* 2006; 118:892-898.
324. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: Good practice and pitfalls. *Lancet* 2002; 359:1686-1689.
325. Sasieni P. A note on the presentation of matched case-control data. *Stat Med* 1992; 11:617-620.
326. Lee GM, Neutra RR, Hristova L, Yost M, Hiatt RA. A nested case-control study of residential and personal magnetic field measures and miscarriages. *Epidemiology* 2002; 13:21-31.
327. Greenland S, Neutra R. Control of confounding in the assessment of medical technology. *Int J Epidemiol* 1980; 9:361-367.
328. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology* 2001; 12:313-320.
329. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J Public Health* 1998; 88:15-19.
330. Uter W, Pfahlberg A. The application of methods to quantify attributable risk in medical practice. *Stat Methods Med Res* 2001; 10:231-237.
331. Schwartz LM, Woloshin S, Dvorin EL, Welch HG. Ratio measures in leading medical journals: Structured review of accessibility of underlying absolute risks. *BMJ* 2006; 333:1248.
332. Nakayama T, Zaman MM, Tanaka H. Reporting of attributable and relative risks, 1966-97. *Lancet* 1998; 351:1179.
333. Hess DR. How to write an effective discussion. *Respir Care* 2004; 49:1238-1241.
334. Horton R. The hidden research paper. *JAMA* 2002; 287:2775-8277.
335. *Annals of Internal Medicine.* Information for authors. www.annals.org.
336. Horton R. The rhetoric of research. *BMJ* 1995; 310:985-987.
337. Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ* 1999; 318:1224-1225.
338. Perneger TV, Hudelson PM. Writing a research article: Advice to beginners. *Int J Qual Health Care* 2004; 16:191-192.
339. 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.
340. Kiviluoto T, Sirén J, Luukkonen P, Kivilaakso E. Randomised trial of laparoscopic versus open cholecystectomy for acute and gangrenous cholecystitis. *Lancet* 1998; 351:321-325.
341. Maldonado G, Poole C. More research is needed. *Ann Epidemiol* 1999; 9:17-18.
342. Phillips CV. The economics of "more research is needed." *Int J Epidemiol* 2001; 30:771-776.
343. Poole C, Peters U, Il'yasova D, Arab L. Commentary: This study failed? *Int J Epidemiol* 2003; 32:534-535.
344. Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychol Bull* 1957; 54:297-312.
345. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999; 130:515-524.
346. Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *N Engl J Med* 1997; 336:309-315.
347. Krimsky S, Rothenberg LS. Conflict of interest policies in science and medical journals: Editorial practices and author disclosures. *Sci Eng Ethics* 2001; 7:205-218.
348. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *JAMA* 2003; 289:454-465.
349. Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med* 1986; 1:155-158.
350. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. *BMJ* 2003; 326:1167-1170.
351. Barnes DE, Bero LA. Industry-funded research and conflict of interest: An analysis of research sponsored by the tobacco industry through the Center for Indoor Air Research. *J Health Polit Policy Law* 1996; 21:515-542.
352. Boyd EA, Bero LA. Improving the use of research evidence in guideline development: 4. Managing conflicts of interest. *Health Res Policy Syst* 2006; 4:16.
353. Higgins P, Orris P. Providing employer arranged occupational medical care: Conflicting interests. *Occup Med* 2002; 17:601-606.
354. Safer DJ. Design and reporting modifications in industry-sponsored comparative psychopharmacology trials. *J Nerv Ment Dis* 2002; 190:583-592.
355. Aspinall RL, Goodman NW. Denial of effective treatment and poor quality of clinical information in placebo controlled trials of ondansetron for postoperative nausea and vomiting: A review of published trials. *BMJ* 1995; 311:844-846.
356. Chan AW, Hrobjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *JAMA* 2004; 291:2457-2465.
357. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: Review of studies in new drug applications. *BMJ* 2003; 326:1171-1173.
358. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2007; MR000005.