A Review Paper

Bias in Research

Steven S. Agabegi, MD, and Peter J. Stern, MD

Abstract

Bias is a systematic inconsistency in research that contaminates the primary comparison. There are several forms of bias, and there are specific methods of minimizing them in different study designs. The randomized controlled trial (RCT) is the gold standard to which all other study designs are compared. However, errors can be made at various stages of a RCT that introduce bias. Furthermore, not all questions can be addressed by a RCT, and in some cases another study design may be more appropriate. Observational studies are more prone to bias, but, when properly conducted with rigorous methods to minimize bias, these studies can be valuable in clinical research.

Bias is a systematic inconsistency in research studies that contaminates a primary comparison and affects the internal validity of the study. Bias is defined as "any trend in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth."[1] While it cannot be totally eliminated, some study designs are more susceptible than others to particular forms of bias, and there are methods of minimizing bias. An understanding of the concept of bias and its deleterious effects on a study's validity serves as the building blocks for critically appraising the literature.

Studies are classified according to "levels of evidence" on the basis of research design, using internal validity (ie, the correctness of the results) as the criterion for hierarchical rankings. Randomized controlled trials (RCTs) receive the highest grade, descriptive studies (case series, expert opinion) the lowest grade, and observational studies (cohort and case-control studies) are intermediate.[2] Each category is considered methodologically superior to those below it,[3] primarily because it is less susceptible to bias.

Studies can be categorized as therapeutic, diagnostic, prognostic, and economic analyses. The discussion of bias in this article focuses on therapeutic studies, which are the most common type in the orthopedic literature.[4] The types of bias in such studies include selection bias, performance bias, attrition bias and detection bias.[5,6] Other terms have been used to describe the types of bias (eg, information bias, recall bias, ascertainment bias), but the above 4 categories are the most common and will be discussed. In this article, we will describe the forms of bias and how they can be minimized in different study designs.

Dr. Agabegi is Resident, and Dr. Stern is Chairman, Department of Orthopaedic Surgery, University of Cincinnati College of Medicine, Cincinnati, Ohio.

Address correspondence to: Steven S. Agabegi, MD, University of Cincinnati, Department of Orthopaedic Surgery, 231 Albert Sabin Way, Cincinnati, OH 45267 (tel, 513-558-4592; fax, 513-558-2220; e-mail, agabegis@yahoo.com).

Am J Orthop. 2008;37(5):242-248.

Types of Studies

The advantages and disadvantages of 4 types of study designs are summarized in Table I.

Case reports and *case series* provide anecdotal information for rare conditions or new treatments, determine long-term outcomes of a procedure, and describe the natural history of a condition or the complication rates after a surgical procedure. The size of a case series can range from 2 or 3 to thousands of patients.[7]

In case series, consecutive cases should be reported so that all outcomes, including those that are unfavorable, are included. The criteria for case selection should be clearly defined with inclusion and exclusion criteria such that readers can reliably compare their patients with those in the case series.[7] If a surgical procedure is being studied, it should be described in such detail that the reader can replicate that same operation.[7] A major drawback of case series is the lack of a comparison group. Historical controls (from previous case series) should be used with caution because of differences in inclusion and exclusion criteria. Treatment techniques may have improved with time, and the results may be more a reflection of this than the actual treatment described.[7] Furthermore, the comparison of case series may be misleading because it is often unknown whether patient factors and treatments were similar.[8]

In *case-control studies,* patients are selected because they have a certain outcome (eg, a complication, development of a disease, mortality). The investigator then retrospectively reviews records to identify exposures or risk factors associated with development of that outcome. Case-control studies are always retrospective. For example, a case-control design would take a group of patients with cervical degenerative disc disease (outcome) and a group without degenerative disc disease to determine whether a history of whiplash injury (exposure) is associated with this outcome.

In *cohort studies,* subjects are chosen based on a certain exposure and are followed over time to observe the development of an outcome. This "exposure" may take the form of surgery, any other intervention (therapy, injections), injury, patient characteristic (obesity, smoking), or any other variable that is being investigated. Cohort stud-

Table I. Study Designs

Study Type	Advantages	Disadvantages
Case series	Useful for determining long-term outcomes of a given treatment or procedure Valuable in describing natural history of a condition, or rates of complications after surgical procedure or treatment	No comparison group Often anecdotal Prone to many biases Difficult to compare with other series Should not be used to assess treatment efficacy
Cohort study/longitudinal study	Good for measuring incidence Can study rare exposure Can demonstrate temporal association between exposure and outcome (necessary for proving causation) Bias in exposure measurement less likely than in case–control studies	Time-consuming, expensive No randomization, so selection bias inevitable Blinding difficult Attrition bias a problem: loss to follow-up common
Case–control study	Time-efficient, inexpensive Effective design for study of rare diseases (may be only feasible method for rare conditions) Best design for diseases with long latent periods Can evaluate multiple exposures or risk factors	Susceptible to confounders and bias (particularly selection bias, detection bias); both exposure and disease occurred before study Appropriate control group may be difficult to identify Dependent on reliable, thorough medical records Difficult to demonstrate temporal association between exposure and disease
Randomized controlled trial	Best study for proving causation Randomized assignation theoretically distributes both known and unknown confounders evenly between groups Highest level of evidence when properly conducted	Expensive, time-consuming Ethically problematic at times Some surgical questions cannot be addressed with these trials

ies may be prospective or retrospective. Using the above example, a prospective cohort study may take a cohort of patients who have sustained a whiplash injury (exposure) and follow these patients over many years to determine the incidence of cervical degenerative disc disease (outcome). In contrast, a retrospective cohort study may identify a group of patients who report a history of whiplash injury (either by recall or clinical records) many years ago, and obtain current cervical radiographs of these patients to determine the incidence of degenerative disc disease. The disadvantage of a retrospective cohort study is the investigators' reliance on patient records (which are often incomplete) for exposure measurement.[9] As is evident from the examples above, the focus of patient selection in a cohort study is the presence of exposure. This is in contrast to the case-control study in which patients are selected based on presence of the outcome. In cohort studies, it is important that subjects with similar characteristics be assembled at a common point in their disease course.[10]

As is evident from the above discussion, the method of patient selection in cohort studies and case-control studies is reversed. In the former, patients are selected because they have been exposed to a factor or intervention and are followed to determine whether these patients develop a particular outcome. In the latter, patients are selected because they have a particular outcome, and the goal is to determine whether these patients had a certain exposure in the past that would be a risk factor for the outcome.

The *randomized controlled trial* (RCT) is a type of cohort study in which subjects are allocated to either the experimental or the control group based on chance. It is the only means of controlling for both known and unknown confounding variables (see below). The basic difference between a RCT and a prospective cohort study is the manner in which subjects are allocated. In a RCT, subject allocation is left to chance, whereas in cohort studies, the treatment decision is not random and is determined by the recommendations of the physician as well as the wishes of the patient. Simply stated, patient assignment to either the experimental or control group is left to chance in the RCT, while it reflects personal judgment, decisions, and beliefs (of both physician and patient) in the prospective cohort study.[11] Therefore, selection bias is inherent in the design of the cohort study, as will be discussed in more detail below.

When conducted flawlessly under ideal conditions, the RCT should theoretically eliminate bias. Practically, however, the RCT is the most difficult to conduct and errors can be made at a number of stages that introduce bias and weaken the validity of the results, as described below.

Blinding

The term *blinding* refers to keeping one or more groups involved in a research study unaware of the assigned intervention or exposure, so that they will not be influenced by that knowledge.[12] Traditionally, there have been 3 levels of

blinding: *single-*, *double-*, and *triple-blinding*, referring to the blinding of study subjects only; subjects and clinicians (physician, nurses or anyone evaluating the patient) only; and subjects, clinicians, and those assessing outcome, respectively. The higher the level of blinding, the lower the risk of bias. Devereaux and colleagues[13,14] showed that physicians and textbooks vary greatly in their interpretations of single, double and triple blinding and recommended abandoning these terms in favor of a precise description of which groups were actually blinded to allocation. These groups include the patients, physicians, data collectors, and individuals assessing outcomes, as well as data analysts and authors.[13]

The effects of unblinded patients, physicians, and data collectors are discussed below in the section on performance bias, and the effects of unblinded investigators assessing outcomes is discussed under detection bias. Unblinded data analysts can also introduce bias through decisions on patient withdrawals and selection of outcomes to analyze and report as well as other decisions such as choice of analytical strategies.[13,15,16]

Confounding Variables

A *confounding variable* is a factor other than the intervention under investigation that obscures the primary comparison. In clinical research, common confounders include gender, age, socioeconomic status, and comorbidities.[10,17] When systematic, confounding is a form of bias and often overlaps the forms of bias described below.

To be a true confounder, a variable must meet 2 criteria: It must be a risk factor for the outcome of interest and it must be associated with the explanatory variable.[10,17] Consider, for example, a hypothetical study designed to assess fracture healing after intramedullary nailing of the tibia with radiographic union/nonunion as endpoints in 2 groups of patients, one of which is allowed partial weight-bearing postoperatively and the other is allowed full weight-bearing. Thus, weight-bearing status is the explanatory variable under investigation. Two other possible risk factors for nonunion include smoking and age (both meeting the first criterion for being a confounder). For smoking and age to be considered confounders, they must also meet the second criterion, ie, be associated with the explanatory variable (weight-bearing status). Otherwise, they are likely to be equally distributed between comparison groups. Age is more likely to be associated with weight-bearing status (ie, older patients may not be able to comply with weight-bearing restrictions) and therefore be a true confounder. Although smoking may be a risk factor for nonunion (first criteria), it is no more likely that patients in the full weight-bearing group are smokers than patients in the partial weight-bearing group, or vice versa. Therefore, smokers are more likely to be distributed evenly between the 2 groups, minimizing the effect of smoking on the primary comparison. In this example, age is a confounding variable because it meets both criteria above. Smoking, on the other hand, only meets one of the 2 criteria and is unlikely to be a confounding variable.

Confounders can be controlled either in the design of the study and/or during the analysis stage such that their effect on the results is minimized. It is always preferable to control for confounders in the design of the study. Control methods include randomization, matching, and restriction (Table II).

Randomization is the only means of controlling for both known and unknown (or unrecognized) confounders.[17] This is one of the main reasons the RCT is considered methodologically superior to other study designs. The fundamental criticism of observational studies is that unrecognized confounders may lead to invalid results.[18] The random allocation process of the RCT minimizes confounding and therefore selection bias, leading to enhanced internal validity.[19]

Forms of Bias

Selection Bias

Selection bias refers to differences in the characteristics of subjects included and those excluded for a given study or between selected comparison groups of the study.[10] Selection bias is intimately related to the concept of confounding, but it is a broader term and describes various sources of differences between the groups. Ideally, comparison groups should be identical in all respects other than the factor under investigation (eg, exposure to a risk factor or an investigational treatment), but in reality such a comparison group does not exist.[11] So the goal is to eliminate any element of human intrusion from the process of patient allocation. Randomization accomplishes this by leaving allocation completely to chance. Any degree of human intrusion into the randomization of subjects introduces selection bias. It is for this reason that the process of allocation concealment (described below) is critical to the integrity of the randomization process.

Selection bias is inevitable to some degree in observational studies because patients are not randomized. The challenge, then, is to design observational studies with a comparison group that is, in all respects, as similar as possible to the study group, apart from the fact that this group did not receive the exposure or treatment under investigation. This can be accomplished by methods that control for confounders in the design stage (preferred) as well as in the analysis stage (Table II).

The randomization process in RCTs hinges on adequate allocation concealment, which reduces selection bias.[20] Allocation concealment, a process distinct from blinding, ensures that the random assignment sequence is concealed from the investigator and patient before and until allocation to treatment.[21] Blinding, however, refers to the process of concealing the treatment from the patient, investigator, and other participants (see above) throughout the remainder of the study after the patient has been assigned to either the experimental or control group.

To illustrate the importance of allocation concealment, consider an investigator enrolling participants in a trial

Table II. Methods to Control for Confounding Variables

Stage	Method	Description
Design	Randomization	Only means of controlling for all confounders (unknown and known) by balancing groups
	Matching	Goal is equal representation of subjects with certain confounder(s) among comparison groups. *Example:* For each group A patient who smokes, have a group B patient who also has this "confounding" variable. Disadvantage of matching patients on a specific variable (eg, age, smoking) is that it excludes that variable from analysis. Matching is most commonly done in case–control studies.
	Restriction	Exclusion of patients with a known confounder from both study groups—only way to totally eliminate *known* confounders. *Example:* Exclude smokers from both groups. *Disadvantage:* Sample size is reduced.
Analysis	Stratification	Divides variable into "stratum" and ensures that an equal number of subjects from each stratum is represented in each group. *Example:* If age is a confounder, determine "age strata" (20-30, 31-40, 41-50, >51) and ensure that comparison groups have equal numbers from each stratum. Disadvantage is that sample sizes are reduced; therefore, chances that a treatment effect will be statistically significant are reduced as well.
	Frequency matching	Ensures that proportion of people with the confounder are the same in the 2 groups—with no concern that a specific person in one group is paired with a person in the other group
	Statistical methods	Use the confounder as a covariate (eg, regression analysis)

Table III. Summary of Methods to Minimize Bias in Different Study Designs[a]

	Study Design		
Type of Bias	Randomized Controlled Trial	Cohort	Case–Control
Selection	Randomization	Control for confounders by matching or restricting	Control for confounders by matching or restricting
Performance	Blinding; avoid contamination and cointervention	Blinding if possible; measurement of exposure should be objective and reliable	Measurement of exposure should be objective and reliable; recall bias often a problem
Attrition	Completeness of follow-up; analyze results on intent-to-treat basis	Completeness of follow-up	Completeness of follow-up
Detection	Blinding of outcome assessor	Blinding of outcome assessor	Method of information gathering should be similar for cases and controls

[a]Adapted from http://www.cochrane.dk/cochrane/handbook/hbook.[26]

who has knowledge of the next intervention "assignment." He may intentionally exclude certain patients based on their prognosis because they would have been allocated to the perceived inappropriate group or, conversely, he may delay the enrollment of patients until the next "assignment" is the perceived appropriate group.[20] When those making the decisions about patient eligibility are aware of the treatment arm to which patients will be allocated, they may systematically enroll sicker, or less sick, patients in either the treatment or the control group.[22] Intentionally or unintentionally, a researcher may assign subjects to the treatment arm of a study that may have a different baseline prognosis than those assigned to the placebo arm.

Intentional violation of strict allocation concealment protocols does occur, sometimes without the knowledge of the primary investigator. For example, an assistant, resident, or student who deciphers the allocation scheme and makes decisions on patient assignment based on this knowledge may be unaware of the ramifications of these decisions and their effect on the validity of a trial.[20]

There are several methods of allocation concealment and the method chosen should be clearly described to allow the reader to critically assess a trial. The investigator must be aware that any mechanism of allocation concealment, no matter how meticulous and well-designed, may be potentially vulnerable to being deciphered. Thus, proper safeguards must be in place to preserve the integrity of the process. Note that the term randomization is sometimes used inappropriately in the literature. Any process that potentially allows a researcher to figure out which patients received which treatment (eg, alternate allocation or allocation by date of birth, hospital number, or day of week attending clinic) is not a true randomization process because the treatment allocation can be predicted and is not left to chance.[23] Such studies are controlled trials rather than randomized controlled trials.

Preferred methods of allocation concealment include remote allocation (in which the individual recruiting the patient makes a call to a methods center to discover the treatment arm to which the patient is allocated),[22] centralized allocation (assignment of subjects performed by a separate department or institution, computerized allocation, or using coded identical containers/envelopes).[24] If envelopes are used, they should be sealed and opaque and sequentially numbered.[25] Invalid methods include alternate allocation and allocation by date of birth, hospital number, or day attending clinic. All of these methods allow the person enrolling subjects to predict which group the subject will be assigned to, and this violates the integrity of randomization. If a study reports that audit checks were done and revealed no tampering, this increases a reader's confidence in the integrity of the process.[25]

Performance Bias

Performance bias is introduced during the treatment or exposure phases of a study and occurs when subjects in comparison groups are systematically given different care in ways other than the intervention under investigation.

To minimize performance bias in randomized controlled trials, the subjects, physicians, and those collecting the data should be "blinded" to the designated intervention status of each group (Table III [26]). Unblinded patients who know that they are in the control or placebo group may be inclined to seek other forms of treatment. Conversely, unblinded patients who know that they are in the treatment arm of a study may be more likely to report placebo effects or have more favorable expectations.

Likewise, physicians who are not blinded and have knowledge of the subject's group assignment may decide to withdraw a participant from a study, provide treatments other than those under study (cointervention or contamination—see below), and may influence patient compliance or reporting of symptoms.[13,27] Of course, in surgical trials, it is impossible to blind the surgeon. It is sometimes reported in RCTs that surgeons are "blinded" to the randomization schedule, but this is misleading because this actually represents allocation concealment rather than blinding. Use of the term blinding implies that the surgeon was unaware of the treatment the patient received throughout the course of the study, which is impossible in a surgical trial. It must be kept in mind that the purpose of allocation concealment is to eliminate or minimize selection bias, whereas blinding serves to minimize performance and detection biases (see detection bias below).[23] So one should distinguish between blinding up to the point of patient assignment into either the control or experimental groups and blinding that occurs after patient allocation. The former refers to allocation concealment (which should always be done) while the latter represents the proper use of the term blinding (which may not always be possible).

Furthermore, if individuals collecting the data are not blinded, results may be distorted because of varying intensity of examination, the possibility of repeating a test for an unexpected finding, the recording of outcomes, or differential encouragement during performance testing.[13,16,28] Unfortunately, many RCTs simply use the terms single-blinded or double-blinded without clarifying which group was actually blinded and without explicitly reporting the blinding status of all groups involved in the trial.[14]

An important principle in minimizing performance bias in RCTs is to ensure that additional treatment other than the intervention be avoided. Additional care (whether invasive or noninvasive such as counseling or physical therapy) for any of the patients becomes a potential confounding factor. Cointervention refers to the provision of additional treatment to either the experimental or control group.[29,30] For example, consider a study aimed at comparing the efficacy of a bone graft substitute with the efficacy of autogenous iliac crest graft in inducing spinal fusion. If the patients in whom the bone graft substitute was used also received electrical bone stimulation postoperatively, this is a cointervention and introduces performance bias. Likewise, contamination refers to provision of the intervention under investigation in the control group.[29,30] For example, using a bone graft substitute in patients randomized to the autogenous bone graft group obviously invalidates the results.

To minimize performance bias in observational studies, the measurement of exposure (to the intervention or factor under investigation) should be objective and consistent (Table III). For example, in a study to determine risk factors for plantar fasciitis, one factor under investigation may be time (number of hours) spent standing during an average day. Proper measurement of this "exposure" (time standing) may prove problematic because it hinges on an honest and accurate estimate on the part of the subject and may be difficult to quantify with precision. Another risk factor under investigation may be the patient's weight, which can be measured more objectively and consistently. Therefore, the nature of the exposure or risk factor itself may pose a problem because some factors (eg, weight, smoking status, diabetes) are more objectively measured than others.

In addition, assessing the degree of exposure from personal recall (as done in many case-control studies) is problematic because recall bias can be introduced.[10] For example, subjects with a positive diagnosis are more likely to recall exposures to risk factors when completing a questionnaire than are subjects without the disease or condition.[30] A patient who suffers from the pain of osteoarthritis is more likely to recall prior episodes of injury than an asymptomatic patient, simply because the former is more likely to have ruminated about the possible origin of the pain. Assessing exposure from medical records may be more objective in certain cases, but medical records may be incomplete, and, often, histories in the medical record are also obtained by personal recall.

One advantage of cohort studies over case-control studies is that bias in exposure measurement is minimized. For example, with a cohort design, one may select a group of subjects who spend more than 12 hours a day standing at work and another group who stands less than 4 hours

a day and follow both groups over time to determine the incidence of plantar fasciitis in each group. With such a cohort study design, measurement of the exposure is more accurate because it is the basis of patient selection. With a case-control design, patients who have already developed plantar fasciitis are reviewed to determine whether these patients have a history of prolonged standing. The retrospective recall of this exposure is difficult to reliably quantify for reasons mentioned above. Practically, using a cohort design to investigate the above problem is time consuming and loss to follow-up is problematic. On the other hand, while a case-control study may be more practical, it is also more susceptible to performance bias because exposure measurement (time spent standing in a day) can be unreliable.

Attrition Bias

Attrition bias relates to patient dropout or exclusion from a study. High dropout rates or systematic differences in the number of dropouts between comparison groups introduces bias, because participants who drop out of a study may differ systematically from those that remain.[31] Ignoring dropouts in the analysis typically skews the results in favor of the intervention under investigation. For example, the results of a study are biased in favor of an experimental treatment when patients drop out due to adverse reactions from this treatment, unless these patients are taken into account in the analysis.[28]

The reader should discern whether the investigators made every reasonable attempt to follow up with dropouts. Unfortunately, there is no recognized dropout rate that is considered acceptable, and any designation is likely to be arbitrary. In general, higher dropout rates are expected (and tolerated) in long-term cohort studies because of the nature of the longitudinal design.

It is critical that the results of comparative studies be analyzed on an *intent-to-treat* basis. This means that the data from patients who withdrew from the study should be analyzed along with the data from the patients who completed the study, regardless of their designation in the intervention or control arm of the study.[22]

Failure to apply the intent-to-treat principle threatens the balance that randomization achieves. To illustrate, consider a hypothetical study to determine the efficacy of a new chemotherapeutic medication to treat osteosarcoma. In a RCT, this medication may be shown to be absolutely effective and safe, resulting in remission in 100% of treated patients when the results are not analyzed on an intent-to-treat basis. However, suppose this medication was so distasteful and caused such severe gastrointestinal side effects that 95% of patients were noncompliant and dropped out of the study prior to completion. In the 5% of patients who completed the study, all experienced remission. In this situation, excluding the dropouts obviously creates a bias because doing so would lead to the erroneous conclusion that the drug is 100% effective. However, by an intent-to-treat analysis of every patient treated with this medication (including those that dropped out), one can show that in a practical "real world" situation, the drug actually fails in 95% of patients.

Detection Bias

Detection bias (also called *ascertainment bias* and *information bias*) refers to systematic inconsistency in outcome assessment. Subjective measures of outcome (eg, pain, loose implants on radiographs) are more likely to contribute to detection bias than objective measures (eg, negative cultures, strength measurements). Outcome measures should be clearly defined, and their measurement should be reliable and standardized for both groups of comparison. For example, consider a hypothetical study to determine the effect of the number of daily dressing changes on the healing rate of a wound. In this study, measurement of the outcome (healing rate) may take the form of subjective assessment of healing (by gross inspection of wound appearance) by 1 observer or by objectively measuring the change in diameter of the wound at weekly intervals by 2 or more independent observers trained to measure wound size in a standardized fashion. Certainly, the latter is more reliable and reduces the possibility of bias. Another example of an unreliable outcome measure may be the assessment of spinal fusion on plain radiographs, which is often inaccurate. A more accurate means of assessing fusion may include a CT scan, or if an animal model is used, histological analysis and biomechanical testing.

Detection bias is minimized in RCTs and cohort studies if the individuals assessing outcome are blinded to treatment allocation, because this ensures that the method of outcome assessment is identical in the comparison groups. Blinding is especially important when outcome measures are subjective.[24] If those assessing outcome are unblinded, the interpretation of marginal findings, or those that require judgment, may be biased.[13,27,28] The observer may be more likely to ascribe a more favorable outcome to the treatment that he prefers.[8]

While the outcome assessor can almost always be blinded, there are situations in which this is not possible. Consider a study to determine the time to union of distal radius fractures treated with either surgery or cast immobilization. If radiographic union is used as the outcome measure, it is impossible to blind the individual(s) assessing outcome, because the surgical implants are visible on radiographs. In this instance, if the individual assessing radiographic union has a preference for surgical treatment, he may be more likely to describe the fracture treated surgically as being healed. Likewise, if tenderness at the fracture site is used as the outcome measure of union, the presence of a surgical incision also precludes blinding of the individual assessing outcome, unless steps are taken to hide the incision or if "sham surgery" is performed. When the outcome assessor cannot be blinded, the use of more than one investigator to independently assess outcome would help minimize detection bias.

In case-control studies, the method of gathering information about exposure should be similar between the cases and the controls. For example, an investigator may use a questionnaire to gather information about exposure from a case but may use

a telephone interview for a control. Furthermore, the search for exposure history may be more intensive for a case than for a control subject. To minimize this form of bias, detail about exposures should be obtained by an investigator who is unaware of whether the subject is a case or a control.

Conflict of Interest

Funding of research studies from industry sources can create bias due to obvious financial incentives. Studies have shown that research funded by pharmaceutical and medical device industries is more likely to reach a positive conclusion.[32-34] Recently, Shah and colleagues[34] showed that in the spine literature, industry-funded studies were more than 3 times more likely to report a positive result than studies with other funding sources. The authors offered possible explanations for this finding, including biases in study design, interpretation, and publication.[34] The concern over conflict of interest in and of itself does not invalidate or dispute the findings of industry-funded research, which is frequently of high methodological quality. However, the reader should always note the funding source of a study and perhaps more closely scrutinize the methodology for bias if a conflict of interest does indeed exist.

Summary

Identifying bias in research studies is critical, because a study that suffers from bias lacks internal validity. Although bias cannot be totally eliminated from studies, the goal is to minimize it. Several pertinent questions can be asked to identify bias. First, were the groups being compared similar in all respects other than the intervention or exposure of interest? Put simply, are we comparing apples with apples? Second, were methods used to control for confounding variables? Although several methods exist to control for confounders, randomization is the only method that theoretically distributes both known and unknown confounders evenly between comparison groups. Furthermore, were subjects, investigators, and outcome assessors blinded to the treatment allocation? While it may not be possible to blind subjects and investigators, outcome assessors can usually be blinded. Finally, was the analysis made on an intent-to-treat basis?

The randomized controlled trial represents the highest level of evidence but it is difficult to conduct and may not be appropriate for some surgical questions. For many orthopedic questions, a well-designed observational study can be a valuable alternative. Regardless of study design, the reader should systematically search for bias and critically determine the degree to which a study correctly represents the relationships being assessed.

Authors' Disclosure Statement

The authors report no actual or potential conflict of interest in relation to this article.

References

1. Last, JM, ed. *A Dictionary of Epidemiology*. 2nd ed. New York, NY: Oxford University Press;1988.
2. Evidence-Based Medicine Working Groups. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA*. 1992; 268: 2420-2425.
3. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342(25): 1887-1892.
4. Obremskey WT, Pappas N, Attallah-Wasif E, et al. Level of evidence in orthopedic journals. *J Bone Joint Surg Am*. 2005;87(12):2632–2638.
5. Feinstein AR. *Clinical Epidemiology: The Architecture of Clinical Research*. Philadelphia, PA: Saunders;1985: 39-52.
6. Greenhalgh T. How to read a paper: assessing the methodological quality of published papers. *BMJ*. 1997;315(7103): 305-308.
7. Carey TS, Boden SD. Critical guide to case series reports. *Spine*. 2003;28(15):1631-1634.
8. Hartz A, Marsh JL. Methodological issues in observational studies. *Clin Orthop*. 2003;413: 33-42.
9. Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about prognosis. *J Bone Joint Surg Am*. 2001;83(10):1555-1564.
10. Szabo RM. Current concepts review—principles of epidemiology for the orthopaedic surgeon. *J Bone Joint Surg Am*. 1998;80(1):111-120.
11. Rochon PA, Gurwitz, JH, Sykora K. Reader's guide to critical appraisal of cohort studies: 1. Role and design. *BMJ*. 2005;330(7496):895-897.
12. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357(9263):1191-1194.
13. Devereaux PJ, Manns BJ, Ghali WA. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA*. 2001;285(15):2000-2003.
14. Montori VM, Bhandari M, Devereaux PJ. In the dark: the reporting of blinding status in randomized controlled trials. *J Clin Epidemiol*. 2002;55(8):787-790.
15. Gøtzsche PC. Blinding during data analysis and writing of manuscripts. *Control Clin Trials*. 1996;17(4):285-290.
16. Jadad A. Bias in RCTs: beyond the sequence generation. In: *Randomised Controlled Trials*. London, England: BMJ Publishing Group;1998:20-36.
17. Kocher MS, Zurakowski D. Clinical Epidemiology and biostatistics: an orthopaedic primer. *J Bone Joint Surg Am*. 2004;86(3):607-620.
18. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000;342(25):1878-1886.
19. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet*. 2002;359(9302):248-52.
20. Schulz KF, Grimes DA. Allocation concealment in randomized trials: defending against deciphering. *Lancet*. 2002; 359(9306):614–18.
21. Schulz KF. Subverting randomization in controlled trials. *JAMA*.1995;274(18):1456-1458.
22. Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about a surgical therapy. *J Bone Joint Surg Am*. 2001;83(6):916 – 926.
23. Altman DG, Schulz KF, Moher D. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134(8):663-94.
24. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273(5):408-412.
25. Devereaux PJ, McKee MD, Yusuf S. Methodological issues in randomized controlled trials of surgical interventions. *Clin Orthop*. 2003;413:25-32.
26. Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.0 [updated February 2008]. The Cochrane Collaboration, 2008. Available from www.cochrane-handbook.org.
27. Guyatt GH, Sackett DL, Cook DJ; the Evidence-Based Medicine Working Group. Users' guides to the medical literature, II: how to use an article about therapy and prevention, A: are the results of the study valid? *JAMA*.1993;270(21):2598-2601.
28. Greenhalgh T. Assessing methodological quality. In: *How to Read a Paper: The Basics of Evidence-Based Medicine*. London, England: BMJ Publishing Group; 1997:39,62-63.
29. The Canadian Cooperative Study Group. The Canadian trial of aspirin and sulfinpyrazone in threatened stroke. *N Engl J Med*. 1978;299:53-9.
30. Sackett DL. Bias in analytic research. *J Chron Dis*. 1979;32(1-2):51-63.
31. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ*. 2001;323(7303):42–46.
32. Friedberg M, Saffran B, Stinson TJ, et al. Evaluation of conflict of interest in economic analyses of new drugs used in oncology. *JAMA*.1999;282(15):1453–7.
33. Rochon PA, Gurwitz JH, Simms RW, et al. A study of manufacturer-supported trials of nonsteroidal anti-inflammatory drugs in the treatment of arthritis. *Arch Intern Med*. 1994;154(2):157–63.
34. Shah RV, Albert TJ, Bruegel-Sanchez V. Industry support and correlation to study outcome for papers published in *Spine*. *Spine*. 2005;30(9):1099-1104.

This paper will be judged for the Resident Writer's Award.