ISAKOS Scientific Committee Report

Research Methodology

Warren R. Dunn, M.D., Stephen Lyman, Ph.D., and Robert Marx, M.D., F.R.C.S.C.

Abstract: Research aims to reach valid conclusions through scientific enquiry. Valid conclusions can only be reached if bias is minimized or eliminated. Bias can potentially take place in the design, implementation, or analysis of a study. Various study designs reduce bias, and this article reviews some of the more common study designs in orthopaedic sports medicine. We also discuss bias and confounding factors as they relate to these studies. **Key Words:** Research—Epidemiology—Outcome—Methodology—Trial.

The primary goal of research is to arrive at valid conclusions through scientific enquiry. Valid conclusions can only be reached in observational or experimental research if bias can be eliminated. Bias is defined as a systematic deviation from the truth which can potentially take place in the design, implementation, or analysis of a study. In most cases, bias cannot be completely eliminated, but it can be minimized. There are various study designs that reduce bias and can be used to ascertain the best treatment options for a disease, as well as those that assess the natural history of disease. In this article, we review some of the more common study designs in orthopaedics and, more specifically, orthopaedic sports medicine, as well as discuss bias and confounding factors as they relate to these studies.

CASE SERIES

A case series is an uncontrolled, retrospective review of a group of patients. No comparison is made with an untreated, or alternate treatment group. By definition, there are no controls. It is cross-sectional in nature, as the treatment and outcome are determined simultaneously at a single point in time. These studies are relatively inexpensive and the results can be obtained quickly. Results from case series are often the first indication of a new diagnosis or treatment modality (for example, the first papers describing the AIDS epidemic). Case series are most useful as hypothesis generators to stimulate more powerful investigations to determine causation; for example, the case series by Hawkins and Hawkins[1] that found an association between shoulder arthrosis following anterior shoulder stabilization and loss of external rotation. This study led to other more confirmatory studies. Although this study design can be very useful in stimulating further study, it is likely the weakest of the commonly used study designs because it is impossible to make valid causal or temporal inferences. Also, patients are often selected based on a characteristic that is linked to the outcome of interest. For example, patients undergoing arthroscopic rotator cuff repair may be more likely to have smaller tears, while large tears may be more likely to be treated by open surgery.

CASE CONTROL STUDY

In a case control study, two groups are compared with one another at a given point in time. To study the

From the Sports Medicine and Shoulder Service, Hospital for Special Surgery, New York, New York, U.S.A.

Address correspondence and reprint requests to Robert G. Marx, M.D., F.R.C.S.C., Sports Medicine and Shoulder Service, Hospital for Special Surgery, 535 East 70th St, New York, NY 10021, U.S.A. E-mail: marxr@hss.edu

0749-8063/03/1908-3886$30.00/0
doi:10.1053/S0749-8063(03)00705-9

possible association of a potentially causative factor and a disease or outcome of interest, a group with that disease or specific outcome is identified (cases), and a group of subjects without the disease or outcome are assembled for comparison (controls). Past exposure to the potential risk factor is measured and compared between cases and controls. An example is given in our own case control study.[2] To determine the potential relationship between shoulder arthrosis and previous shoulder dislocation, a group of patients that had undergone shoulder replacement for arthrosis (cases) were asked about past shoulder dislocation. A control group, which consisted of patients who had undergone knee replacement for gonarthrosis, was also questioned about any past shoulder dislocations. If there are any important differences other than exposure history (previous shoulder dislocation) between these two groups, the ability to make inferences can be limited. The controls should arise from the same population as the cases. In other words, if one of the control subjects were diagnosed with the disease of interest before study initiation, that subject would be eligible for inclusion in the case group.

Matching is a method that can be used to assure that cases and controls are similar apart from the exposures of interest. Cases are matched to one or more controls based on characteristics that the researchers are not interested in studying, but that may influence the outcome of interest. It is important to limit matching to factors that are not of interest to the investigation, because once a factor is matched it ensures equal distribution in both groups. This precludes that factor from potential analysis. Relative risk is a measure of the strength of association. However, the relative risk cannot be directly calculated in case control studies. Instead, the odds ratio is used as an estimate of relative risk.

The advantages of this design are that it is relatively inexpensive and the time needed is relatively short because there is no wait for follow-up. This design is best used when the outcome or disease is rare and the exposure is frequent among the diseased. However, it can be difficult to establish a temporal relationship between exposure and disease, and it is often complex to select the appropriate controls. Potential bias is introduced with the assessment of previous exposure. If exposure data are obtained from old records, the information is often incomplete. Completeness may be related to disease state (i.e., reporting bias), which can greatly influence the study findings. If exposure data are acquired from patient interviews, the outcome may affect the recollection of exposure (i.e., recall bias).

COHORT STUDY

The cohort and case control designs differ in that the starting point of one is the end point of the other. The case control study design starts with subjects with the disease or outcome and then assesses past exposure, whereas the cohort study design begins with exposed and nonexposed groups, which are followed for the development of a disease or outcome. The terms prospective and retrospective indicate the timing of data collection, but "prospectively gathered data" does not mean that the study is prospective unless the study was proposed before data collection. The terms cohort study, follow-up study, and longitudinal study have all been used to describe the same design where a study population is identified at the beginning of a study and followed through calendar time until the point at which the outcome of interest develops or does not develop. Since new (incident) cases are identified, a temporal relationship can be determined. However, this can take a long time. In a retrospective cohort study, the patients may be tracked over time or identified at a later date, but the hypothesis generation and data gathering occur after the period of study. This may require less time to complete than a prospective study. It is similar to a case series, in that the cohort is assembled retrospectively. However, data are available at different points in time. As mentioned earlier, the measure of the strength of association used in cohort studies is relative risk.

CONFOUNDING

Confounding is perhaps the most significant dilemma in well designed and analyzed observational studies.[3] A confounder is a factor that can cause or preclude the outcome of interest. Consider a study on dislocation[2] where the presence of previous surgery on the shoulder is considered a confounder. Had the investigators not accounted for this factor, the study would have been weakened by confounding because the influence of previous surgery on arthrosis could not be distinguished from the factor under study (dislocation).

BIAS

Bias is the systematic deviation of inferences or results from the truth, or processes that bring about

such deviation. Numerous types of bias have been described,[4] some of which are discussed below. Bias in assessing outcome can occur when the person making decisions on disease status is aware of the exposure status as well. Information bias can occur when the quality and extent of information is different for the exposed and unexposed groups. This can occur with historical control groups where data from medical records are used.

Analytic bias arises if those analyzing the data have strong preconceptions about exposure and outcome. Assembly bias is a form of selection bias. This can occur if the patients assembled differ in ways other than the factors under study. This has been termed susceptibility bias.[5] For example, to study the association between anterior cruciate ligament (ACL) tears and subsequent gonarthrosis, a study is conducted using patients with previous ACL disruption in a surgeon's practice to determine whether they develop gonarthrosis. This surgeon's practice may focus on a subgroup of the population that is at high risk of gonarthrosis, which may result in assembly or susceptibility bias.

Migration bias occurs when patients leave the study or move to the other cohort. If the characteristics of the patients leaving one cohort are similar to those leaving the other, there is no bias. However, this is rarely the case. Usually, the reason for dropping out or changing cohorts is related to prognosis. For instance, patients who are doing extremely well or extremely poorly may elect not to return for follow-up. Similar are nonresponse and loss of follow-up bias. Measurement bias arises if one group has a better chance of having the outcome measured. For example, the numbers of pivoting episodes following ACL reconstruction may be followed more closely in a postoperative group compared with a nonoperative group and, therefore, are erroneously thought to have more such episodes. Also, a phenomenon worthy of mention is the Hawthorne effect, wherein subjects change their behavior because they are the target of special interest and attention in a study. This can be best managed with the use of a control group.

There are many strategies that can be used to reduce bias. If the study goal and purpose are established prior to gathering the data, bias is reduced. As discussed earlier, matching controls is also useful. Probably the best technique to diminish bias is, unfortunately, very difficult to do in surgery: randomization.

RANDOMIZED CLINICAL TRIALS

Randomized clinical trials are the gold standard for research involving the health of human populations. In a randomized controlled clinical trial (also called an experimental study as opposed to an observational study), a defined population is randomly allocated to either the treatment group or the control group. The treatment group receives the treatment under question and the control group receives no treatment or the standard treatment, depending on the situation.

Randomization is the ultimate method for arriving at comparable groups,[5] because it indirectly matches for all prognostic variables, recognized and unrecognized, and consequently minimizes selection bias. Any difference in outcome can then be attributed to the intervention. External and internal validity are often used to describe clinical trials. External validity refers to the generalizability of a study, i.e., how representative of the reference population the study sample is. Internal validity is the extent to which the observed differences in outcome in the study can be attributed exclusively to the hypothesized factor under investigation. These concepts have been clarified by Grimes and Shulz.[6] They describe internal validity simply as the degree to which the study measured what it set out to investigate, and external validity as the extent to which the reader can generalize the results to their patients. An indication for a randomized controlled trial is equipoise. Equipoise is a state of uncertainty in the orthopaedic or medical community at large regarding the benefits or troubles associated with each of two or more treatments. Without equipoise, it would be unethical to allocate patients to a treatment group suspected of being inferior to another. For this reason, this type of study design is rarely used to compare surgical interventions.

An alternative to randomization that is not uncommon in sports medicine is the use of simultaneous nonrandomized controls. This study design may be subject to significant bias. Some examples of nonrandom methods of treatment group allocation are even/odd days, the clinician's preference, or by patient selection of a particular surgeon in a group of surgeons participating in a study. Selection bias compromises the validity of these studies. Take the following hypothetical situation for example, in which a surgeon participating in a study investigating ACL reconstruction versus conservative management after ACL disruption uses even/odd days for treatment group allocation. The surgeon, acting on behalf of what he perceives to be the patient's best interests, could easily

alter the treatment group of a patient by having the patient return on a different day. Hence, young, active individuals may have ACL reconstruction and older, less active patients may be treated conservatively. The results of such a study would be questionable given the selection bias that led to noncomparable treatment groups.

HEALTH SERVICES RESEARCH

Health services research denotes studies comparing the effects of two or more health care interventions or treatments, types of health care delivery, or form of insurance coverage or compensation on health or economic outcomes. Often these studies use data from large administrative or financial databases that were not constructed for research purposes in an attempt to obtain useful information without a lengthy and costly prospective study. The advantages of such studies are that they typically make use of large, population-based data, which enhances generalizability. Analysis is relatively quick because the data have already been collected and problems associated with small sample size are avoided. The disadvantages of these studies stem from the fact that the data are often not collected for research purposes. Hence, coding may be inconsistent, disease severity is difficult or impossible to quantify, and the data may be incomplete. Researchers have no control over quality-control procedures and, therefore, the internal and external validity of the study are often unclear. For example, many health services researchers use national Medicare data. These data only apply to Americans over the age of 65 who have participated in the Medicare program. These people may not be representative of younger Americans or older Americans who do not use Medicare.

INJURY SURVEILLANCE

In its simplest form, injury surveillance reports the number of injuries that arise in a given population. However, in practice, surveillance systems make up some of the most useful databases available for study. Similar to a prospective cohort study, surveillance systems collect information on all cases (or a sample of cases) of the outcome of interest as the outcome occurs. These systems typically are used to track trends in diseases or injuries that are very common or severe in the population of interest. For example, the National Highway Traffic Safety Administration collects information on all fatal crashes on the public roads in the United States (Fatal Analysis Reporting System) and a sample of all tow-away property damage and injury crashes (National Automotive Sampling System). These surveillance systems differ from large administrative databases often used in health services research because they were designed with epidemiologic analysis in mind. They often have hundreds of variables and are carefully monitored for methodologic flaws and errors that are found during quality control efforts are corrected.

Injury surveillance systems of particular interest in sports medicine and orthopaedics include the National Electronic Injury Surveillance System (NEISS) administered by the Consumer Product Safety Commission. NEISS is a sample of all emergency room visits with an injury diagnosis. Another interesting injury registry is the National Collegiate Athletic Association's Injury Surveillance System (NCAA-ISS). NCAA-ISS collects information on a population of athletes playing college athletics. All injuries arising in these athletes are closely evaluated.

REFERENCES

1. Hawkins RH, Hawkins RJ. Failed anterior reconstruction for shoulder instability. *J Bone Joint Surg Br* 1985;67:709-714.
2. Marx RG, McCarty EC, Montemurno TD, Altchek DW, Craig EV, Warren RF. Development of arthrosis following dislocation of the shoulder: A case-control study. *J Shoulder Elbow Surg* 2002;11:1-5.
3. Gordis L. *Epidemiology*. Ed 2. Philadelphia: WB Saunders, 2000;308.
4. Sackett DL. Bias in analytic research. *J Chronic Dis* 1979;32: 51-63.
5. Rudicel S. Sports injury research. How to avoid bias. *Am J Sports Med* 1988;16:S48-S52 (Suppl 1).
6. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002;359:248-252.