Understanding Study Design

Kristin L. Sainani, PhD, Rita A. Popat, PhD

Medical studies use several common designs. Correctly naming a study's design is not just an exercise in terminology but helps readers to evaluate the study's level of evidence, limitations and biases, and statistical analyses. Analytic studies evaluate the association between a risk factor (eg, intervention, exposure) and the outcome. Studies fall into 2 overarching categories: observational and experimental. This article reviews the major analytic study designs, including their advantages and disadvantages, and their relevance for statistics (summarized in Table 1).

OBSERVATIONAL STUDIES

Confounding: extraneous variables that are related to both the risk factor and the outcome may create a spurious association or mask an association between them.

In observational studies, participants self-select to different risk factors. Observational studies usually are less costly and easier than experimental studies and thus make up the bulk of the literature. However, well-conducted observational studies provide a lower level of evidence than well-conducted experimental studies because of confounding. Confounding occurs because risk factors tend to cluster, which makes it difficult to isolate the causative agent. For example, people who drink heavily also tend to smoke, eat poorly, and avoid exercise; thus, it is difficult to be sure that drinking causes a particular outcome rather than some related factor (or factors).

Cross-sectional Studies

Prevalence: the proportion of subjects who have a particular risk factor or outcome; for disease outcomes, includes both old and new cases.

Lifetime prevalence: the proportion of subjects who have ever had the outcome, even if they have since recovered.

Recall bias: reporting of risk factors by people with the outcome may differ from people who do not have the outcome.

Nonresponse bias: people who participate in a study may differ from people who refuse to participate, which could bias the results.

Cross-sectional studies measure risk factors and outcomes at a single time point. They aim to describe the prevalence of risk factors and outcomes in a particular population, and to explore associations between them. They are relatively quick and easy; sometimes they just involve a survey. For example, Keilani et al [1] surveyed 100 Viennese skateboarders about their injury history and found that 92% had had at least one injury, 45% had had a severe injury, and only 13% used protective equipment. They also tested whether injury frequency was associated with weekly training time, terrain, occupation, stance, or other risk factors.

Cross-sectional studies have several limitations. Because risk factors and outcomes are measured simultaneously, it is impossible to ascertain whether a given risk factor actually preceded the outcome. For example, a cross-sectional association between low training time and sports injuries could arise if reduced training caused injuries (because of inexperience) or if injuries reduced training time (because of forced rest). Thus, temporal relationships will always be uncertain (unless the risk factor is something fixed, such as gender or genes). Because outcomes have already occurred, this also may bias self-reported measures (recall bias); for example,

K.L.S. Division of Epidemiology, Department of Health Research and Policy, Stanford University, HRP Redwood Bldg, Stanford, CA 94305. Address correspondence to: K.L.S.; e-mail: kcobb@stanford.edu
Disclosure: nothing to disclose

R.A.P. Division of Epidemiology, Department of Health Research and Policy, Stanford University, HRP Redwood Bldg, Stanford, CA
Disclosure: nothing to disclose

Disclosure Key can be found on the Table of Contents and at www.pmrjournal.org

Submitted for publication April 4, 2011; accepted April 18, 2011.

Table 1. *Summary of study designs*

Study Design	Description	Advantages	Limitations and biases
Observational			
Cross-sectional	Measures prevalence of risk factors (eg, interventions, exposures) and outcomes at one time point	Relatively inexpensive, easy, and quick; often generalizable; provides valid estimates of prevalence for risk factors and outcomes	Cannot determine whether risk factors preceded outcomes; may uncover risk factors associated with duration (or survival) rather than cause of outcome; inefficient for rare risk factors and outcomes; may be subject to nonresponse bias, recall bias, and confounding
Case-control	Selects participants based on outcome status (case or control) and asks them about past risk factors	Efficient for rare outcomes; relatively inexpensive, easy, and quick	Getting comparable control subjects is often tricky; temporal relationship between risk factors and the outcome may be uncertain; inefficient for rare risk factors; cannot be used to estimate rates or risks of the outcome or risk ratios; may be subject to recall bias and confounding
Prospective cohort	Measures risk factors in an outcome-free cohort and observes them until they develop the outcome	Temporality is certain: risk factors preceded outcomes; prevents bias that may occur after a person develops an outcome; provides valid estimates of rates and risks of outcomes; can be used to study multiple outcomes	Can be lengthy and costly; inefficient for rare outcomes; may be affected by loss to follow-up and confounding
Retrospective cohort	Similar to a prospective cohort study, but the cohort is assembled after outcomes have already occurred, by using stored records	Similar benefits as a prospective cohort study but also faster and cheaper	Risk factor data were not collected specifically for the study, and thus certain variables and confounders may be unavailable; data quality may be low; requires stored or electronic records; may be affected by loss to follow-up (may not be able to get outcome information on everyone in the retrospectively assembled cohort) and confounding
Experimental			
Randomized controlled trial	Participants are randomly assigned to interventions and then followed up over time	Criterion standard for showing cause-and-effect relationships; randomization minimizes confounding; blinding and placebos help minimize bias	Expensive; not practical for showing long-term effects; not always generalizable; not always ethical or feasible

athletes who have had an injury may be more likely to recall the use (or lack) of protective equipment. Cross-sectional studies also are limited by nonresponse bias: Persons who are willing to participate may differ from those who refuse to participate. For example, in Keilani et al [1], only 75 of 100 skaters completed the survey (a good response rate for surveys); it is possible that injured skaters were more motivated to respond (given their vested interest in the topic) than noninjured skaters, which may have led injury-prevalence estimates to be overestimated.

Cross-sectional studies select participants randomly rather than on the basis of their outcomes or risk factors. This helps to make the results generalizable to a wide population (by assuming that participation rates are high). However, it also makes cross-sectional studies inefficient for studying rare risk factors or outcomes, which will occur infrequently in a random sample.

Case-control Studies

Case-control studies select participants on the basis of outcomes and are thus the most efficient design for rare outcomes. In fact, for very rare diseases and for outbreak situations, the case-control study may be the only feasible design for exploring etiology. Researchers select participants based on their outcomes: They find persons who have had the outcome (called case subjects) and compare them with persons who have not had the outcome but who are similar to the case subjects in other ways (called control subjects). For

example, Faulk et al [2] studied risk factors for critical gastrointestinal (GI) bleeding in an inpatient rehabilitation center; this outcome occurs in only about 0.3% of patients. By using computer records, the researchers identified 35 cases that occurred in an 11-year period (1997-2008); for each case subject, they randomly selected a control subject, in this instance, someone who was admitted to the rehabilitation facility at the same time as the case subject but who did not develop a critical GI bleed. By using data extracted from the patients' medical records, they found that case subjects were more likely than control subjects to be diabetic and to have been treated with steroids or anticoagulants (before their bleeding event).

The choice of control subjects is arguably the most challenging aspect of a case-control study. If systematic differences exist in how case and controls subjects are selected, then this can create spurious associations. For example, if Faulk et al [2] had failed to match control subjects at admittance time and instead chose 35 control subjects from 2008, they may have found differences between case and control subjects that were solely related to evolving treatment practices at the rehabilitation facility.

As with cross-sectional studies, case-control studies collect data after an outcome has already occurred, which can lead to bias. Case subjects may be searching for reasons why they developed a disease and thus may be more likely than control subjects to remember and report risk factors (recall bias). They also may have changed their behavior as a result of their diagnosis and may not remember their prediagnosis habits. The onset of disease may also affect objective measurements that are taken as part of the study. For example, cancer patients may have lower blood levels of vitamin D, but it is impossible to tell whether this is a cause of, or a result of, their disease (eg, lower vitamin D could result from chemotherapy treatments, spending more time indoors, or changes in diet). Faulk et al [2] used medical records to obtain data on risk factors, which helps to avoid these issues.

Incidence: the rate at which new cases of disease are developing, expressed as the number of new cases per person-year.

Cumulative incidence or risk: the chance of developing an outcome over a specified time period, expressed as a percentage.

Case-control studies are also limited in the statistics that they can generate. They cannot be used to estimate the prevalence, incidence, or cumulative incidence of an outcome, because the percentage of case subjects in the sample does not reflect the percentage in the population under study. The sample in the study by Faulk et al [2] contains 35 case subjects and 35 control subjects but, clearly, critical GI bleeds do not occur in 50% of patients in rehabilitation facilities. (In a separate analysis, Faulk et al [2] estimated the risk of critical GI bleeds to be 0.3%.) Case-control studies are also inefficient for studying rare risk factors, which will occur infrequently in the sample.

Prospective Cohort Studies

Cohort studies select participants on the basis of their risk factors rather than on their outcomes. In a prospective cohort study, investigators assemble a cohort of persons who are free of the outcome at baseline and then observe them over time (longitudinally) until some persons develop outcomes. For example, Segal et al [3] assembled a cohort of 1540 older adults who were free of knee symptoms (pain and stiffness) at baseline but were at risk for developing symptoms because of excess weight or other factors. The investigators measured participants' knee strength and alignment at baseline, and then observed these individuals for 30 months to determine whether these variables predicted incident knee symptoms.

One of the greatest advantages of cohort studies is that risk factors are collected before outcomes have occurred, which makes temporal relationships certain and avoids recall bias. For example, cross-sectional studies detected an association between reduced knee strength and increased knee pain but could not establish which came first. Segal et al [3] failed to find a longitudinal relationship between baseline knee strength and incident knee pain, which suggests that knee pain may reduce knee strength (because of disuse) rather than vice versa.

Cohort studies have several additional advantages. They can be used to assess rare exposures (because investigators can target particular risk groups) and multiple outcomes (because sampling is not based on outcomes). These studies also provide valid estimates of incidence or cumulative incidence in a population. For example, in the study by Segal et al [3], 307 of 2275 knees (13%) developed symptoms; thus, the risk of developing knee symptoms in at-risk older adults can be estimated to be about 13% per knee over a 2.5-year period. Cohort studies often sample people from the general community, which makes the results generalizable to a wide population.

Prospective cohort studies, however, can be costly and time consuming. Large sample sizes and long follow-up periods may be necessary to ensure that sufficient outcomes occur. Segal et al [3] enrolled 2480 knees and observed participants for 30 months to detect 307 knees with incident symptoms. Losses to follow-up are also common. Segal et al [3] collected complete follow-up data on 2275 knees, losing only 9% to follow-up, an extremely high retention rate. Losses to follow-up can bias a study's results if these losses are related to both risk factors and outcomes.

Retrospective Cohort Studies

Retrospective cohort studies are conceptually similar to prospective cohort studies. The difference is the timing of when

the cohort is assembled. In retrospective cohort studies, investigators assemble the cohort after outcomes have already occurred, by using stored data that were collected in the past (before the development of outcomes). The investigators link these data to outcomes data. Retrospective cohort studies are much faster than prospective cohort studies, because the follow-up time has already transpired. However, retrospective cohort studies may lack key information because the stored data were not collected expressly for the study. Data quality also may be an issue. These studies otherwise are similar in regard to advantages and limitations.

For example, Meiner et al [4] performed a retrospective cohort study that examined the relationship between tissue plasminogen activator (tPA) use and functional outcomes in stroke patients. They used data from a stroke registry to identify 37 stroke patients treated with tPA and 37 patients (of similar ages and stroke severities) who had not been treated with tPA and then measured the participants' functional outcomes. Although identified as a case-control study, this study, in fact, is a retrospective cohort study because the participants were selected on the basis of their exposures (tPA treatment) rather than on their outcomes (functional status).

Meiner et al [4] found that patients treated with tPA were significantly more likely to have good functional outcomes than patients not treated with tPA (73% versus 46%, respectively). As with any observational study, the association could be explained by confounding. For example, those patients who get to an emergency room sooner are more likely to receive tPA, and prompt medical attention may be associated with improved outcomes for reasons other than tPA. The retrospective nature of the study may make it more difficult to address confounding; because the data were not collected expressly for a study on tPA, important confounders may not have been recorded (such as the exact contraindications in the non-tPA group).

EXPERIMENTAL STUDIES

Experimental studies typically offer a higher level of evidence than observational studies, because randomization minimizes confounding. In experimental studies, the investigator controls the selection of interventions. Well-conducted randomized controlled trials are considered the criterion standard for establishing a causal link between a risk factor and an outcome.

Randomized controlled trials

In randomized controlled trials, investigators randomly assign participants to different intervention arms and then observe these individuals over time to assess outcomes. For example, Boon et al [5] conducted a double-blind randomized controlled trial of botulinum toxin type A (BoNT-A) for the treatment of painful knee osteoarthritis. In this pilot study, 60 patients were randomly assigned to receive a single injection of steroids, low-dose BoNT-A, or high-dose BoNT-A. Pain outcomes were measured at 4, 8, 12, and 26 weeks after injection.

The use of blinding and placebos can help to strengthen the conclusions from randomized trials, especially when the outcome measures are subjective (eg, pain). In the BoNT-A trial, both the subjects and the investigators who made assessments were blinded to study treatment. However, the BoNT-A trial did not use a placebo (just 3 active treatments); thus, it is impossible to determine whether the treatment responses observed exceeded a placebo effect. It is not always feasible or ethical to use a placebo. However, controversial placebos, such as sham surgery, have revealed important findings; for example, a recent study found that vertebroplasty was no more effective than sham surgery for the treatment of osteoporotic spinal fractures [6].

Randomized trials may be limited by noncompliance and losses to follow-up. Noncompliance was not an issue in the study by Boon et al [5], because treatment involved only a single injection. However, in many studies, patients will stop taking treatments, will take them incorrectly, or will start alternative treatments. These added factors will reduce statistical power, which makes it more difficult to see effects. As with prospective cohort studies, losses to follow-up can be severe and can potentially bias the results. The study by Boon et al [5] had perfect retention through 8 weeks (the primary end point), but nearly half the participants were lost by 26 weeks, which makes longer-term end points difficult to interpret. Large losses to follow-up also create large amounts of missing data that investigators must grapple with if they are analyzing their data according to the intention-to-treat principle [7].

Persons who sign up for randomized trials are not always representative of the general population. Randomized trials may have extensive exclusion and inclusion criteria. Also, persons who are willing to leave their treatment up to fate (i.e., randomization) may be different than people who are not willing. Therefore, results sometimes have limited generalizability. Randomized trials are also expensive and are usually used only to measure short-term effects. Finally, randomized trials are not always ethical or practical; for example, it is not ethical to randomize participants to smoke.

CONCLUSIONS

Generalizability: the extent to which findings about the study sample can be applied to the population at large

When reviewing a study, readers should carefully identify the study design. Observational studies are more common in the literature but are limited by confounding. Cross-sectional and case-control studies assess risk factors after outcomes have occurred, which may introduce bias and make temporal relationships uncertain. Cohort studies

avoid this problem, but prospective cohort studies can be lengthy and costly, and retrospective cohort studies may be limited by incomplete and poor quality data. Well-conducted randomized trials are the criterion standard for establishing causal relationships, because randomization minimizes confounding. However, randomized trials are expensive and difficult to conduct; they may also be impractical or unethical for testing certain hypotheses.

REFERENCES

1. Keilani M, Krall C, Lipowec L, Posch M, Komanadj TS, Crevenna R. Skateboarding injuries in Vienna: location, frequency, and severity. PM R 2010;2:619-624.
2. Faulk CE, Sandoval A, Draughon M, Moore D, Whitted T, Wilford R Jr. Critical gastrointestinal bleeding at an inpatient rehabilitation center: incidence, risk factors, and the role of gastrointestinal prophylaxis. PM R 2010;2:1104-1112.
3. Segal NA, Torner JC, Felson DT, et al. Knee extensor strength does not protect against incident knee symptoms at 30 months in the Multicenter Knee Osteoarthritis (MOST) cohort. PM R 2009;1:459-465.
4. Meiner Z, Sajin A, Schwartz I, et al. Rehabilitation outcomes of stroke patients treated with tissue plasminogen activator. PM R 2010;2:698-702.
5. Boon AJ, Smith J, Dahm DL, et al. Efficacy of intra-articular botulinum toxin type A in painful knee osteoarthritis: a pilot study. PM R 2010;2: 268-276.
6. Kallmes DF, Comstock BA, Heagerty PJ, et al. A randomized trial of vertebroplasty for osteoporotic spinal fractures. N Engl J Med 2009;361: 569-579.
7. Sainani KL. Making sense of intention-to-treat. PM R 2010;2:209-213.