Overview of clinical research design

DANIEL M. HARTUNG AND DANIEL TOUCHETTE

Purpose. Basic concepts and terminology of clinical research design are presented for new clinical investigators.

Summary. Clinical research, research involving human subjects, can be described as either observational or experimental. The findings of all clinical research can be threatened by issues of bias and confounding. Biases are systematic errors in how study subjects are selected or measured, which result in false inferences. Confounding is a distortion in findings that is attributable to mixing variable effects. Uncontrolled observation research is generally more prone to bias and confounding than experimental research. Observational research includes designs such as the cohort study, case–control study, and cross-sectional study, while experimental research typically involves a randomized controlled trial (RCT). The cohort study, which includes the RCT, defines subject allocation on the basis of exposure interest (e.g., drug, disease-management program) and follows the patients to assess the outcomes. The case–control study uses the primary outcome of interest (e.g., adverse event) to define subject allocation, and different exposures are assessed in a retrospective manner. Cross-sectional research evaluates both exposure and outcome concurrently. Each of these design methods possesses different strengths and weaknesses in answering research questions, as well as underlying many study subtypes.

Conclusion. While experimental research is the strongest method for establishing causality, it can be difficult to accomplish under many scenarios. Observational clinical research offers many design alternatives that may be appropriate if planned and executed carefully.

Index terms: Clinical studies; Methodology; Nomenclature

Am J Health-Syst Pharm. 2009; 66:398-408

While the use of observation to disprove theories, the basis for modern scientific thought, has its origin in the 17th century, clinical research is a relatively new endeavor. At its core, clinical research involves the analyses of empirical data (information observed in nature) in an attempt to answer some discrete clinical questions. All research activities should begin with the development of a specific research question. For most questions, it is impossible to make observations from the entire population of interest; therefore, clinical research is conducted within a sample. While the distinction and relationship between the population and a study sample are straightforward concepts, they are fundamental to understanding issues related to study validity and statistical inference. Internal validity is the degree to which observations accurately reflect what they are intended to measure. A research design is never perfect, and internal validity of observations can be threatened in two major ways: systematic error and random error. Proper planning, design, and analysis are essential in minimizing the impact of these threats to the validity of scientific observation. The extent to which internally valid research findings can be inferred back to the population of interest and other populations is termed "external validity." The key to correctly translating the results of a research study to the population of interest rests in the ability to assure that the study was designed and executed in a manner minimizing internal errors (internal validity) and that the research was conducted in a sample that accurately reflects the population (external validity) of interest.

The relationship among the population and sample and validity of inferences is depicted in Figure 1. Suppose an investigator is interested in knowing whether a new drug or intervention is effective at

DANIEL M. HARTUNG, M.P.H., PHARM.D., is Assistant Professor, Pharmacy Practice, College of Pharmacy, Oregon State University, Portland. DANIEL TOUCHETTE, M.A., PHARM.D., is Assistant Professor, Pharmacy Practice, and Core faculty, Center for Pharmacoeconomic Research, College of Pharmacy, University of Illinois at Chicago, Chicago.

Address correspondence to Dr. Hartung at the College of Pharmacy, Oregon State University, 3303 Southwest Bond Avenue, CH12C, Portland, OR 97239 (hartungd@ohsu.edu).

The authors have declared no potential conflicts of interest.

Copyright © 2009, American Society of Health-System Pharmacists, Inc. All rights reserved. 1079-2082/09/0202-0398$06.00.
DOI 10.2146/ajhp080300

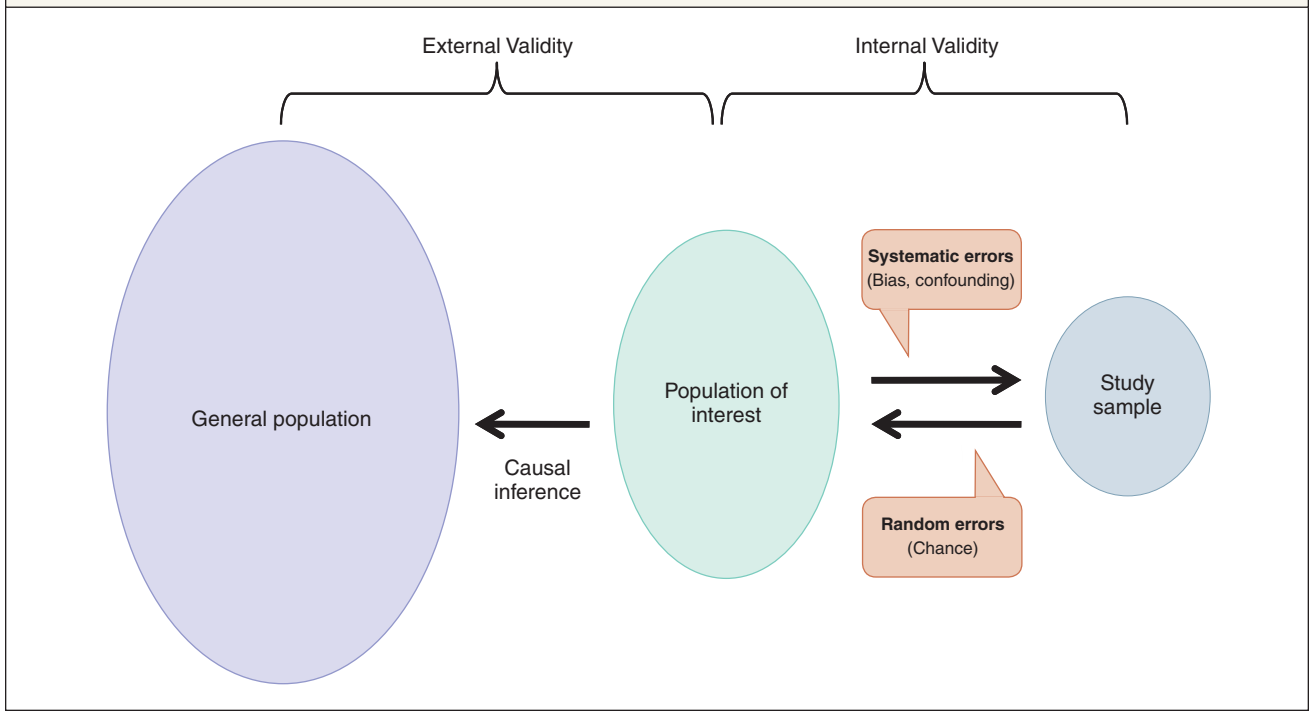
The Research Fundamentals *section comprises a series of articles on important topics in pharmacy research. These include valid research design, appropriate data collection and analysis, application of research findings in practice, and publication of research results. Articles in this series have been solicited and reviewed by guest editors Lee Vermeulen, M.S., and Almut Winterstein, Ph.D.*

reducing low-density-lipoprotein (LDL) cholesterol in patients with coronary artery disease (CAD) using a randomized controlled trial (RCT) methodology. Because it would not be feasible to enroll the entire population of patients with CAD, the investigator must implement the research plan in a sample of patients. Ideally, the study sample would be randomly drawn from the population of interest. However, this is almost never possible, and study samples are often enrolled from a logistically convenient population (e.g., men with CAD, age 55–75 years, who are recruited from an academic medical center). Nevertheless, observations from the sample are then extrapolated back to the population of interest using statistical inference to quantify random variation and the role of chance. From the previous example, the investigator observes that subjects receiving the new drug have their LDL cholesterol reduced more than those who do not receive this new drug. Using a statistical test, the investigator is able to probabilistically quantify the likelihood that the observation could have been attributable to random error. If the role of random variation can be ruled out and the study is relatively free of systematic errors (not biased), the investigator can then make inferences about causation to the general population. That is, the finding that a new drug is able to reduce LDL cholesterol levels among men with CAD, age 55–75 years, can be further generalized to the ability to reduce LDL cholesterol in everyone. It is important to note that the external validity, or generalization of study findings, is predicated on sound internal validity. This article describes the basic study designs available for clinical researchers, discusses their advantages and disadvantages with respect to minimizing threats to internal validity, and explains how to balance research validity requirements and study feasibility.

Clinical research design

Clinical research can generally be classified into two distinct categories: observational and experimental.[1] Experimental research entails any research in which the investigator intervenes in a population for the sole purpose of evaluation. The RCT is the predominant type of experimental clinical research. Research that is observational makes no attempt to intervene in the study sample for exclusively investigational purposes. It includes a wide variety of study

Figure 1. Relationship among the general population, the study population of interest, and the study sample as it relates to internal validity and external validity.



types, ranging from traditional epidemiologic cohort and case–control studies to quasi-experimental time series methods in which natural experiments are conducted. Natural experiments are conducted in situations in which the factor under study cannot be controlled, and those who experience the factor are compared to those who do not (e.g., a study comparing individuals living near a natural disaster epicenter compared to those who do not). Drawing from an epidemiologic framework, the variable of inquiry is often termed an exposure and can include clinical programs, environmental hazards or conditions, or consumption of prescription drugs.[2,3] Observational research can also be classified as retrospective or prospective with respect to the perspective of the investigator. For example, an observational cohort study is prospective if study subjects are followed in real time and data about exposures and outcomes are collected. In contrast, in a retrospective cohort study, all exposures and outcomes have occurred and the investigator recreates the time sequence using existing data, such as the medical record or administrative claims. All experimental research by nature is prospective.

Causal inference

Regardless of study design, results can only be inferred to the population if they are internally valid and bias, confounding, and chance can be reasonably minimized. Typically, clinical research involves analyzing the association between two or more variables. One or more variable is termed an outcome (dependent) variable, and the others are termed predictor (independent) variables. Association or correlation between the dependent and independent variables is often interpreted as a cause and effect relationship. However, establishing true causality requires more than just an observed association. Sir Austin Bradford Hill,[4] in a classic epidemiology article, laid out several criteria that, if present, strongly argue in favor of causation between two variables. Hill's criteria included the following:

- Strength of association—the validity of observation as well as the magnitude of the association estimate,
- Consistency—the reproducibility of the association in independent studies,
- Temporality—the time sequence consistent with cause and effect,
- Biological gradient—the dose or exposure response,
- Biological plausibility and coherence—the association biologically believable and consistent with current scientific thought,
- Specificity—the concept stipulating that one cause leads to a single effect,
- Analogy—the cause–effect relationship that exists between similar variables, and
- Experimentation—the removal of the variable in question, which decreases effect.

While the merit of each of these criteria has been debated (e.g., specificity implies that one exposure leads to one effect, which is not always true) in modern clinical research, the criteria remain useful in framing discussions about the relative strengths and weaknesses of different study designs.[5] This may be especially true for practice-based research settings where conducting true experimental research is not practical and observational methods predominate. Specifically, establishing the causal link between a predictor and outcome variable involves evaluating the strength of association and the validity of the observation, which means ensuring that the relationship is not attributable to bias, confounding, or chance.

Chance. Chance is the likelihood that the studied observation is the result of random variation. The role of chance in research is quantified through statistical tests (inferential statistics), and different study designs and data types require different statistical approaches. While detailed discussion of inferential statistics is beyond the scope of this paper, it is worth mentioning that ruling out chance as an explanation for an observed association is rarely the most vital issue related to study validity. General causal inference, as opposed to statistical inference, is a nonquantitative process that involves consideration of study observations as well as Hill's criteria, such as biological plausibility and consistency. This article examines specific research design methods and describes practical aspects of conduct and the relative advantages and disadvantages in minimizing bias, confounding, and chance.

Bias. Bias is a broad term for diverse scenarios in which systematic error in measurement of observations causes a nonrandom distortion of the true association, or lack thereof, between a predictor variable and an outcome variable.[6] While there are many different types, biases are generally categorized one of two ways: selection bias and information bias.[7] Selection biases occur when subjects are selected for a study in a way that creates a false association. One common type of selection bias encountered in observational studies is termed "healthy user bias." This bias was prominently highlighted as the reason for the disparate findings between observational studies demonstrating cardiovascular benefit of hormone replacement therapy (HRT) and the contrasting results of the Women's Health Initiative and the Heart and Estrogen/progestin Replacement Study trials.[8-11] Before the publication of these two landmark RCTs, data from methodologically strong observational studies had suggested that use of HRT protected against heart disease. However, when the question was finally studied in a randomized setting, the opposite was observed—HRT did not protect

against, and may have increased, the risk of heart disease. The health-user bias is based on the premise that, in observational studies, subjects are not randomized to treatment groups and are, in a sense, self-selected; therefore, important differences in comorbidity and lifestyle may also be selected. For example, it is likely that women in the original observational studies who were taking HRT also engaged in other healthy dietary and lifestyle activities compared to women who were not taking replacement therapy. Additionally, it also may be plausible that the women in these observational studies who were adherent with HRT were also more adherent with their other drugs known to reduce the risk of heart disease (e.g., antihypertensives) or were more actively engaged in other preventive measures.

The other primary type of bias is called information bias. Information bias occurs when the method of data collection is systematically different between study groups. In contrast to selection bias, information bias relates to systematic errors in how a variable is measured. One commonly cited type of information bias is recall bias, which arises when subjects have differential recollection of key variables that are related to their study group allocation. An example would be a case–control investigation of drug exposure in patients having a suspected adverse drug reaction. Those experiencing the adverse reaction may have a more heightened recall of everything they had ingested in the immediate past compared with those who did not have an adverse reaction. It is critical to note that, unlike chance and confounding, biases cannot be quantified easily and can only be eliminated or minimized in the design stage.

Confounding. Confounding occurs when the study association is partially or entirely mediated by a third factor.[2,6] Confounding is classically depicted as the interplay among three variables—A, B, and C—where A is the exposure of interest, B is the outcome, and C is the potential confounder. For C to be a confounder, it must be associated with the exposure, be a true cause of B, and not be in the causal pathway (an intermediary) between A and B. An observational study examining exposure to thiazolidenediones (A) among individuals with or without heart failure (B) might be confounded by the third factor of severity of disease (C). Because diabetes is an independent risk factor for development of heart failure and is associated with being prescribed a thiazolidenedione, it is reasonable to speculate that an observed association between thiazolidenedione use and heart failure might be explained by the confounder of poorly controlled diabetes.[12] Control of confounding is critically important in observational research, but it may also be problematic in nonrandomized, experimental research (quasi-experimental). Unlike biases, confounding is not necessarily introduced through a faulty study design but rather is a result of complex relationships between observed and unobserved factors. Additionally, unlike bias, confounding can be analytically managed and adjusted for using different design or analytic strategies. The most powerful method for controlling confounding in experimental research is randomization. In theory, randomization equally distributes known, unmeasured, and unknown subject characteristics (e.g., disease severity, age) that may confound associations of interest. While randomization greatly increases the likelihood of producing an equal distribution of potential confounders, it is not a guarantee and study group differences may persist. Because randomization is not always feasible, accounting and control of confounding in observational research are critical. There are four generally accepted methods to control for potential confounders in observational research—restriction, matching, stratification, and multivariate analysis.[3] The first two methods must be implemented during the design stage, and the latter two can be handled during analysis.

Restriction, or exclusion, is a technique in which individuals with variables suspected to be confounders are excluded or restricted to a particular value from the study with the aim of producing a more homogeneous study sample. For example, if one were to conduct an observational study examining the impact of a new drug for hyperlipidemia on cardiovascular endpoints, it would be important to control for the difference in smoking status between groups. One way to achieve this would be to restrict your study sample to either all smokers or all nonsmokers. Matching is a procedure where study patients are matched to controls on one or more potentially confounding variables.[13-15] Confounding can also be minimized by stratifying by the confounding variable. This is similar to exclusion, except each level of the confounding variable is analyzed separately. From the example above, instead of excluding one group contingent on smoking status, the investigator analyzes the association between lipid-lowering drug use and cardiovascular outcomes among smokers and nonsmokers separately. Finally, multivariate analysis is one of the most commonly used techniques to control for confounders in observational research because it can statistically adjust for a multitude of variables simultaneously using regression techniques.

Observational research designs

Because of the time and expense required for experimental clinical research, observational research offers the most opportunity for practice-based research. However, unlike experimental research, observational research can suffer from many threats to validity, requiring careful design and interpretation. The choice

of research design will depend on a number of factors including the research question, data availability, setting, time, and resources available. The most basic type of observational research is descriptive research. Descriptive studies are typically the first scientific inquiry into new areas of interest or concern and aim to provide clues about causation for further analytic research. Examples of descriptive study designs are case reports, series, and surveillance reports. The goal of analytic research is to test hypotheses about the association between two or more variables in order to further arguments about causation.

Cohort studies. Cohort studies are the most intuitive and straightforward of all observational research.[2,3,16] The hallmark of all cohort studies is the following of groups, or cohorts, of subjects through time (virtual or real) with ultimate ascertainment of their development of a disease or outcome. Group assignment is typically defined by exposure (e.g., patients taking HRT) or magnitude of exposure (e.g., drug dosing). Cohort studies can be either prospective (real time) or retrospective (virtual time). In prospective cohort studies, cohorts are enrolled in the study by the investigator at the inception of the study and followed through time to ascertain the development of outcomes or diseases. Retrospective cohort studies involve using a previously existing data set, such as an administrative claims data set or medical record, to virtually assemble the exposure cohorts and ascertain and analyze what occurred following cohort assignment. Both have differing advantages and disadvantages.

Observational cohort studies are powerful study designs because they ensure, by definition, that exposure precedes outcome. Additionally, unlike other observational designs, cohort studies enable direct estimation of disease or outcome incidence rates or risks. Incidence rates between exposed and unexposed groups are then compared to estimate rate ratios. Because the cause always precedes the outcome in a prospective cohort study, these studies are less susceptible to selection bias. An investigator performing a prospective cohort study would not know which subjects would eventually develop an outcome because the outcomes have not occurred. Cohort studies are also useful in exploring multiple outcomes associated with an exposure. This is especially relevant for observational studies of drug effectiveness because multiple outcomes, as well as adverse reactions, can be evaluated. There are, however, several limitations inherent in the conduct of cohort studies. Prospective cohort studies can be very time-consuming and expensive to complete. If appropriate data exist, retrospective cohort studies can be completed with considerably fewer resources. Cohort studies are also inefficient for evaluating rare outcomes that have long latent periods (e.g., cancer prevention). Unless appropriate safeguards are established, such as blinding, cohort studies are prone to information biases. Another major challenge of conducting cohort studies is patient follow-up. Loss to follow-up can result in a diminution of the number of outcomes observed and subsequent statistical power. Even more critical, loss of follow-up within cohort studies can be a major source of selection bias. Like other types of study designs, subjects who drop out of cohort studies do so for a reason that is unlikely to be random. Unlike prospective studies, little or no information is available as to why the subject dropped out of the study and nothing can be done to follow up with patients who have dropped out. Finally, channeling bias can also be an important problem since clinicians do not prescribe drugs to patients at random. Clinical, functional, and social characteristics likely influence which patients are prescribed specific drugs and which are not. If these factors are independent risk factors for the outcome of interest, then a channeling selection bias will exist. Channeling bias is also sometimes termed "confounding by indication."

Case–control studies. Case–control studies, unlike cohort studies, work backward analytically from outcome to exposure.[2,3,17] Because exposure and outcome are not temporally sequenced, case–control studies are sometimes poorly understood and misinterpreted. In contrast to a cohort study in which study groups are defined by exposure, study groups within a case–control study are defined by the outcome. Subjects who have an outcome of interest (cases) are typically compared to subjects who have not had an outcome (controls). Once cases and controls have been identified, their exposure history is ascertained and compared. Because study groups are not defined by exposure status, incidence rates and associated measures of association (e.g., incidence-rate ratios) cannot be directly calculated. Case–control research will produce an odds ratio (OR), reflecting the ratio of odds of exposure among cases to the odds of exposure among controls. When the prevalence of disease in the population is rare, ORs are good estimates of the underlying relative risk ratio of the outcome given exposure. Under most sampling conditions, it can be shown that ORs only marginally overestimate or underestimate relative risk when the cumulative incidence of an outcome is less than 10%.[18]

For questions involving adverse drug events that are relatively rare, it may be advantageous to use a case–control design because of the efficient use of identified cases that experienced the adverse event. In contrast, if one wished to understand both the benefits and risks of a drug in a real-world environment, a case–control study would not be ideal

because a case–control study can evaluate only one outcome.

Case–control studies are relatively inexpensive to conduct and are the most efficient observational design when outcomes are relatively rare. However, case–control studies are also the most susceptible to both selection and information biases. Selection bias can present itself as a result of selection of both cases and controls. When selecting cases, newly diagnosed subjects (incident cases) or subjects with a previously existing disease (prevalent cases) can be selected. In most situations, it is advisable to select incident cases in order to avoid confusing any causal association from an association with disease survival. For example, if sampling includes both incident and prevalent cases of a disease that has significant early mortality, then the sample population would be overrepresented by individuals who had survived long enough to be studied and would be unrepresentative of the true target population. Selection of controls is more complicated and critical for producing unbiased estimates of association. Ideally, control substances should be from the same source population that gave rise to the cases. Any restrictions applied to the selection of cases also need to be applied to controls.[19] Several commonly cited control sources include hospital-based controls, general-population controls, and case-relationship controls. Hospital-based controls are, for some, a readily available source of information about previous exposures and can also be similar to the case group with respect to site of identification. However, hospital-based controls are inherently ill and may not reflect a healthy nondiseased population. This may lead to an overestimate or underestimate of risk if controls are less or more likely to be exposed to the proposed risk factor by virtue of their hospital status. Another commonly used control source is from the general population enrolled through random-digit phone number dialing or other outreach activities. While this approach may produce a control group that originates from the same source population as cases, it is also more expensive and time-consuming. Additionally, a general-population control sample does not assure nonbiased selection because controls who are available by phone or interview may be systematically different than cases in important ways (e.g., they are home during the workday). Another commonly used control group is made up of individuals who have a friendly or proximal (e.g., neighbor) relationship with a case. These controls may be more willing to participate in the study than subjects selected from the general population because of their relationship to the case. They are potentially similar to cases on factors that may be related to the exposure and could therefore help control for confounding. The main disadvantage of using relationship controls is that the exposure of interest may also be related (e.g., diet, exercise habits) and may not be truly representative of the source population.

Case–control studies are especially prone to information biases because the temporal sequence between outcome and exposure is reversed. Cases and controls may remember past exposures and experiences in different ways. Because cases have the disease or outcome of interest, they may be more probing or thoughtful when examining past exposures or activities compared to controls. This differential assessment is termed "recall bias." If one is examining the association between a disease and past prescription-drug exposure among patients enrolled in a health care plan, he or she should use administrative prescription claims data or medical records to ascertain past exposure most reliably and eliminate the possibility of recall bias. In this situation, control selection is still critical, but it is unlikely that cases and controls would differ systematically in their recall of past exposures because exposures are determined using an existing data set that was populated presumably before they were designated as study participants. In the absence of these types of data, other methods are available, such as blinding the subject or observer to case–control status and using multiple potential risk factors to screen and blind to the true hypothesis. Diagnostic bias is a similar type of information bias in which the diagnosis of a disease is influenced by past exposures. For example, a case–control study exploring the association between oral contraceptive use and venous thrombosis could potentially be biased by virtue of the fact that clinicians may be especially vigilant of this condition among women who use oral contraceptives.

Specialized case–control studies. Several specialized types of case–control trials exist and mix elements of other observational designs. Nested case–control studies are case–control studies that are embedded into existing prospective cohort studies. They are particularly useful when additional collection of data (beyond what is defined at cohort inception) is needed. Nested case–control studies are only assayed on a sample of subjects with and without disease, which dramatically reduces the costs of running tests on the entire cohort. Additionally, because all data are collected before outcome occurrence, a nested case–control study is unlikely to be impacted by recall bias. Another variant of the case–control study is the case-crossover. Case-crossover studies borrow methodology from the experimental crossover design in which study subjects randomized to a treatment for a period of time crossover into the placebo group to serve as their own control.[20] For each case, a period of time before disease development, or hazard window, is selected to ascertain exposure. A second time period preceding the case-

hazard window is selected for ascertainment of control-period exposure. Case-crossover studies are best suited for diseases that are believed to be rapidly occurring upon exposure (e.g., acute triggers of myocardial infarction), and the exposure should be something that varies over time.[21] Because each case serves as his or her own control, case-crossover studies are adept at controlling for many potential confounders as well as selection bias. However, confounding variables may still distort outcome-exposure associations if they change over time.

Cross-sectional studies. Cohort and case–control studies can broadly be described as longitudinal because the association between an exposure and outcome is observed in at least two discrete time points. Both designs incorporate a temporal sequence between exposure and outcome that is intrinsically required for establishing a causal effect. The cross-sectional design assesses both exposure and outcome simultaneously and is sometimes characterized as a snapshot in time. In contrast to cohort and most case–control studies, cross-sectional studies deal with prevalent cases. A cross-sectional study is a reasonable option for quickly exploring a series of exposures and one or more prevalent diseases. However, because the timing of exposure and disease is unknown, causality is difficult to establish. A cross-sectional design would be a reasonable approach if one were interested in assessing the prevalence of aspirin use among a population of patients with known cardiovascular disease for quality-assessment purposes.

Experimental research

The RCT is one of the great advances in medicine and is recognized as the gold standard for medical evidence. When executed properly, the RCT can effectively eliminate issues of confounding and minimize many types of bias which plague most observational research. Among the hierarchy of medical research, the RCT is typically considered the best type of research design for determining causality and, therefore, is the foundation for determining the effectiveness of health technology. With this research design, study subjects are randomly assigned to treatment or control groups. The purpose of the random assignment is to eliminate selection bias and any confounding that may occur caused by an imbalance between study groups. Random assignment of patients to a particular treatment usually results in study groups that are well matched when comparing important characteristics, such as sex, age, disease severity, and comorbidities. However, random assignment of patients into study groups does not guarantee matched groups because this process is subject to chance. Therefore, it is still important to collect, analyze, and report patient demographics and other important potential confounding characteristics. A well-designed study may still try to adjust for any confounders that may be present, especially if there are a large number of known potential confounding variables. While minimizing some types of bias, random assignment of patients does not affect bias that can occur in a study after patients have been assigned to their treatment groups. However, these types of bias can often be minimized through blinding and the use of standardized objective outcome variables.

Randomization. Subjects can be randomized using a number of methods, some of which are considered better than others. A truly random process will usually result in more closely matched groups. The use of a random-number generator to assign patients to a study group is considered a strong randomization method. Methods such as assigning subjects to study groups based on the day of the week or alternating enrollment order are not truly random and should be avoided. However, even a randomization process that is carefully executed can lead to unequal distribution of subject characteristics by chance, especially if the study sample is small. To protect against this, a system called "permuted block randomization," or block randomization, was developed.[22] Block randomization assures that there will not be a large imbalance between study groups at any point during enrollment. With this method, random assignments are made in blocks, each containing the same number of treatment and control patients. For example, if an RCT has a randomization algorithm to two groups (treatment and placebo) with a block size of four, two subjects would be enrolled in every group for every four subjects enrolled, and the order in which they would be allocated would be randomized again at every block (TTPP, PPTT, TPTP, TPPT, PTTP, or PTPT). If subject characteristics are systematically different at one period of enrollment compared to another period, a blocked randomization would guard against the inadvertent randomization of one characteristic to one study group. Smaller blocks may be used to have more control over the influx of patients into a particular arm of the study and to ensure that group sizes are very similar. However, small block sizes may make it easier to guess to which group the next patient may be enrolled, potentially leading to a bias. For example, if it is believed that a particular study subject may benefit more from being in one of the study groups, efforts may be made (intentionally or not) to enroll the patient in that study group. To reduce the likelihood of an investigator being unblinded to the blocking scheme, the block size can also be randomly changed.

Blinding. Blinding, or masking, is a method of eliminating or reducing two types of information bias in a study that occur after a subject's enrollment. The first bias

occurs when subjects know that they are being studied and may act and react differently than they otherwise would. This patient bias is termed the Hawthorne effect.[7] For example, subjects randomized to a treatment might respond better to the treatment if they know what it is and might not respond if they were assigned to a control group. The first method of blinding, called a single blind, attempts to keep the study subjects from knowing which group they were assigned to. Blinding is often accomplished by using a placebo treatment, which minimizes the impact of the Hawthorne effect on the outcomes being studied. Note that blinding does not eliminate the Hawthorne effect. The placebo effect, a specific type of Hawthorne effect in which patients respond to receiving an inactive placebo, has been well documented in numerous studies.[23] However, blinding the patients to their group assignment often minimizes differential Hawthorne effects so that all study groups experience a similar degree of change.

The second type of information bias reduced by blinding is investigator detection bias.[7] When an investigator is aware of the treatment status, outcomes may be (subconsciously or consciously) interpreted and ascertained in biased fashion based on preconceived notions of efficacy. Specifically, an investigator who prefers a particular therapy may interpret the outcomes of patients receiving that therapy as being better. The investigator may also influence patient responses to the outcome. Preventing the investigator from knowing the study subject's group assignment prevents these investigator biases from occurring. Clinical trials that blind the investigator as well as the patient are termed double blind. Therapies that have pronounced and rapid adverse effects on the body, such as severe nausea, can often lead patients or investigators to correctly guess which group they were assigned to. When a therapy with notable adverse effects is being studied, the investigators should have an exit interview assessing whether patients or investigators were successfully able to identify group assignments.[22]

Intention-to-treat (ITT) analysis

As with other types of studies, RCTs are also susceptible to attrition bias. Attrition bias is a type of selection bias in which subjects are lost to follow-up or are noncompliant in a way that is not random with respect to study group assignment. For example, if less healthy subjects randomly assigned to a new drug experienced an adverse reaction that caused them to drop out at a higher rate than similar subjects randomized to placebo, a bias may be introduced. Unlike retrospective studies, reasons for attrition may be collected and explored, potentially reducing the effect of the bias. One method of accounting for this is analyzing data according to the ITT principle.[24] The ITT principle stipulates that all randomized subjects need to be accounted for in the analysis. Other analytic approaches include per-protocol, in which only subjects who comply with the protocol are analyzed. However, only ITT preserves initial randomization and is the preferred method for standard RCTs. Analysis according to the ITT principle will produce the most conservative estimates of treatment effect and produce results most likely to be consistent with the null hypothesis of no difference between groups. If missing data and lost patient follow-up are large enough, bias can persist despite application of ITT. Several data imputation methods exist, ranging from the last observation carried forward to statistical modeling techniques; however, all available methods operate on assumptions that are sometimes questionable.[25]

Noninferiority trials

Specialized noninferiority or equivalence RCTs are conceptually different because the null and alternative hypotheses are reversed. In contrast to a standard superiority RCT, the null hypothesis in a noninferiority trial is that a difference exists and the alternative hypothesis is that no difference exists (or, more accurately, the difference is within some small margin). Noninferiority trials aim to demonstrate whether a new treatment is therapeutically equivalent to an existing treatment. Because ITT effectively dilutes any treatment effect, it will produce results more likely to reject the null hypothesis of a difference and accept the conclusion of noninferiority or equivalence. As a result, a per-protocol analysis is the recommended approach for these types of RCTs.[26,27]

External validity of RCT

RCTs are typically designed to estimate the efficacy of a drug, device, or procedure under experimental conditions designed to maximize internal validity, which essentially is the ability of a study to eliminate confounding and determine the impact of a therapy on the outcome. In an effort to minimize confounding and bias, RCTs often employ procedures that are not common in clinical practice and threaten the study's external validity (applicability and generalizability of the study's findings). For example, studies may limit the study subjects to healthier volunteers using strict inclusion criteria. By keeping the studied population more homogeneous, a smaller number of patients are needed to show a difference between treatment and control groups. However, people enrolled in clinical trials may be very different from those in clinical practices where a greater disease burden, more extremes of age, a greater ethnic disparity, and other patient factors may dramatically influence outcomes. In addition, procedures conducted in clinical trials, such as greater time spent with a clinician, more frequent laboratory assessments, surveys, and

more frequent assessment and reinforcement of adherence to the study medications, are not always the same in clinical practice. As a result, study subjects display higher adherence rates, are more likely to have minor adverse drug events identified, may have serious adverse drug events caught sooner, and are more likely to have positive outcomes. Different types of study designs have been developed to address some of these limitations of RCTs and extend the findings to these other populations of interest.

Clinical trials designed to answer questions important to health care decision-makers are termed practical (or pragmatic) clinical trials (PCTs).[28] Unlike RCTs, PCTs compare relevant treatments instead of comparing to a placebo, include patients who more closely resemble the general population, and include a broad range of relevant (usually terminal) health outcomes instead of focusing on intermediate health outcomes. With PCTs, patients are still randomized to a therapy. Outcomes data are often collected for a large number of patients using a registry. An example of a frequently cited PCT is the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial.[29]

Quasi-experimental research. Quasi-experimental methods encompass a spectrum of research designs that have elements of both experimental and observational designs. Quasi-experimental research is particularly useful for policy and programmatic changes in health care systems because changes in exposure are adopted for entire populations and randomization is logistically impossible and not politically feasible.[30] Similar to observational cohort studies, subject assignment to a study group is not random and is either self-selected or selected through administrative policy selection. However, the investigator or policy originator may have considerable discretion as to who receives the intervention and when.

Derived from the social sciences, major quasi-experimental designs can generally be classified into three groups in ascending order of methodological hierarchy: (1) pre–post designs with no control group (i.e., each subject serves as his or her own control), (2) pre–post designs with control groups, and (3) interrupted time-series designs with or without control groups. There are many other permutations of these basic designs; however, their use is uncommon and they are discussed elsewhere in detail.[31]

Quasi-experiments suffer from many of the same methodological drawbacks as other nonrandomized observational designs. The most commonly cited threats to internal validity within quasi-experimental research include controlling for important confounders (e.g., secular trends) and regression to the mean. A secular-trend confounder is a pattern of behavior or practice in the study population occurring at the same time as the phenomena of interest that could be responsible for the study observations. For example, a study evaluating the implementation of a prescription-drug, cost-sharing policy in an insured population might be confounded by other changes in prescribing behavior unrelated to the policy, such as changes in the standard of care or a separate policy. Regression to the mean is a type of selection bias to which quasiexperimental studies are especially prone. This bias occurs when study subjects are chosen based on an extreme observation. Retesting the observation or outcome will likely move the sample mean toward the population mean because it is more likely that individuals in the group will do better (since they were already chosen for having an extreme value on the first observation). For example, evaluating a disease-management program focusing on patients with heart failure identified through a hospital after an exacerbation would likely have a sample of patients with high initial care needs and costs. However, over time, their disease would likely stabilize independent of the program.

Uncontrolled pre–post designs are commonly encountered quasi-experimental designs and involve a single pretest observation (O_1) and the intervention (X) followed by a single posttest observation (O_2). Under this design, O_1 serves as the control for O_2. However, if the sample population were selected based on a threshold level of the initial observation (O_1), it would be difficult to discern if observations made during O_2 were attributable to X or regression to the mean. Selecting a control group from a similar population for this analysis better controls for regression to the mean and secular confounding, as well as other confounding factors. For example, an investigator exploring the impact of a previous authorization policy for brand-name nonsteroidal antiinflammatory drugs (NSAIDs) on total NSAID use in a state Medicaid population could select some other nonrelated drug class (e.g., statins) within the same population or compare NSAID use in another state Medicaid program if possible. Controlled pre–post designs are commonly described using the following notation:

$$\begin{array}{ccc} O_{1a} & X & O_{2a} \\ O_{1b} & & O_{2b} \end{array}$$

Finally, the interrupted time-series design with or without a control series is the strongest quasi-experimental design. It is especially well suited for interventions that occur abruptly, such as policy changes.[30,32] Time-series analyses are similar to pre–post designs except multiple observations are made before and after the intervention:

$$\begin{array}{l} O_1\ O_2\ O_3\ O_4\ O_5\ O_6\ X\ O_7\ O_8\ O_9\ O_{10}\ O_{11} \\ \quad O_{12} \end{array}$$

By collecting a series of observations, both before and after the intervention, time-series analyses are better able to control for confounding secular trends and regression to the mean. Additionally, time-series studies are able to explore the impact of the intervention immediately and in a longer time period. Figure 2 shows a common analytic approach to time-series analysis called a segmented regression model. This model allows an investigator to examine immediate changes in use (β_2), as well as changes in the overall trend (β_3) after a populationwide intervention, such as implementation of a prior-authorization process for drugs. A control series from the same population, but with unrelated outcomes, or a different population can be incorporated to further strengthen the design. This may be necessary if another intervention also affecting the outcome measure was implemented concurrently. The reader is referred to the work of Ray[30] and Wagner et al.[32] for practical reviews of the methodology.

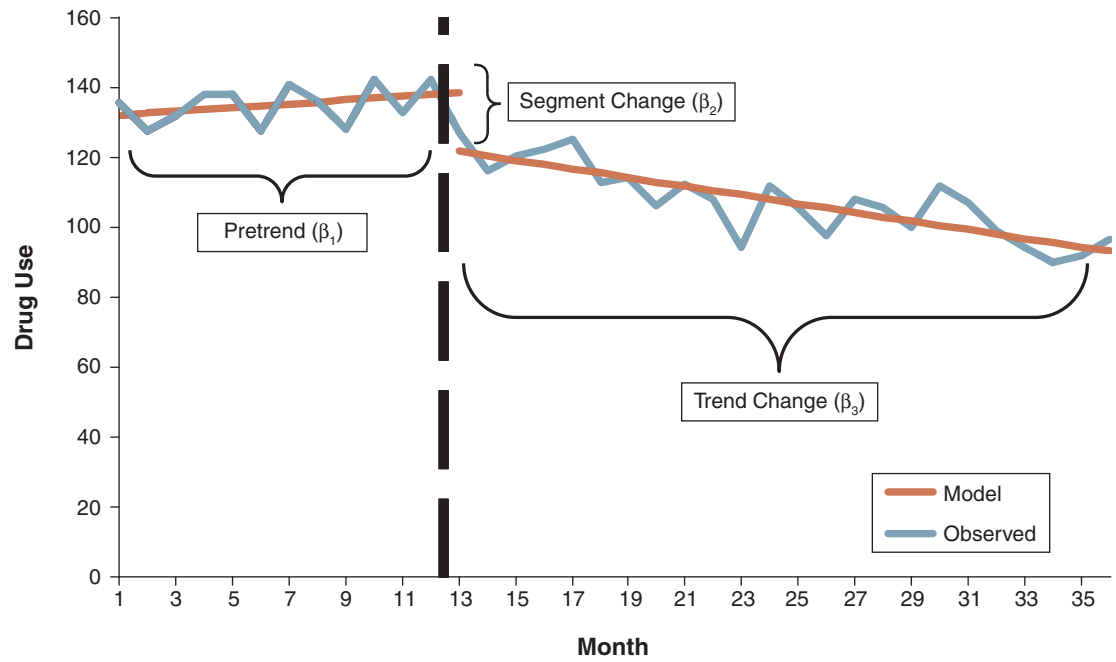
Conclusion

While experimental research is the strongest method for establishing causality, it can be difficult to accomplish under many scenarios. Observational clinical research offers many design alternatives that may be appropriate if planned and executed carefully.

References

1. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet.* 2002; 359:57-61.
2. Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown; 1987.
3. Hulley SB, Cummings SR, Browner WS et al. Designing clinical research: an epidemiologic approach. Philadelphia: Lippincot Williams & Wilkins; 2001.
4. Hill AB. The environment and disease: association or causation? *Proc R Soc Med.* 1965; 58:295-300.
5. Rothman KJ. Epidemiology: an introduction. New York: Oxford; 2002.
6. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet.* 2002; 359:248-52.
7. Sackett DL. Bias in analytic research. *J Chronic Dis.* 1979; 32:51-63.
8. Hulley S, Grady D, Bush T et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA.* 1998; 280:605-13.
9. Rossouw JE, Anderson GL, Prentice RL et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA.* 2002; 288:321-33.
10. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol.* 2003; 158: 915-20.
11. Grodstein F, Clarkson TB, Manson JE. Understanding the divergent data on postmenopausal hormone therapy. *N Engl J Med.* 2003; 348:645-50.
12. Hartung DM, Touchette DR, Bultemeier NC et al. Risk of hospitalization for heart failure associated with thiazolidinedione therapy: a Medicaid claims-based case-control study. *Pharmacotherapy.* 2005; 25:1329-36.
13. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-

Figure 2. Graphic depiction of segmented linear regression of drug use over time. β_1 is the regression model estimate of preperiod trend (slope), β_2 is the model estimate of the change in use postperiod, and β_3 is the model estimate of the change in trend (slope) in the postperiod.



randomized control group. *Stat Med.* 1998; 17:2265-81.
14. Glynn RJ, Schneeweiss S, Stürmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol.* 2006; 98:253-9.
15. Seeger JD, Williams PL, Walker AM. An application of propensity score matching using claims data. *Pharmacoepidemiol Drug Saf.* 2005; 14:465-76.
16. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. *Lancet.* 2002; 359:341-5.
17. Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet.* 2002; 359:431-4.
18. Zhang J, Yu KF. What's the relative risk?: a method of correcting the odds ratio in cohort studies of common outcomes. *JAMA.* 1998; 280:1690-1.
19. Grimes DA, Schulz KF. Compared to what? Finding controls for case-control studies. *Lancet.* 2005; 365:1429-33.
20. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol.* 1991; 133:144-53.
21. Schneeweiss S, Stürmer T, Maclure M. Case-crossover and case-time-control designs as alternatives in pharmacoepidemiologic research. *Pharmacoepidemiol Drug Saf.* 1997; 6(suppl 3):S51-9.
22. Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials. 3rd ed. New York: Springer-Verlag; 1998.
23. Miller FG, Rosenstein DL. The nature and power of the placebo effect. *J Clin Epidemiol.* 2006; 59:331-5.
24. Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ.* 2001; 165:1339-41.
25. Streiner DL. The case of the missing data: methods of dealing with dropouts and other research vagaries. *Can J Psychiatry.* 2002; 47:68-75.
26. Jones B, Jarvis P, Lewis JA et al. Trials to assess equivalence: the importance of rigorous methods. *BMJ.* 1996; 313:36-9. [Erratum, *BMJ.* 1996: 313:550.]
27. Piaggio G, Elbourne DR, Altman DG et al. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA.* 2006; 295:1152-60. [Erratum, *JAMA.* 2006; 296:1842.]
28. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA.* 2003; 290:1624-32.
29. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA.* 2002; 288:2981-97. [Erratum, *JAMA.* 2003; 289:178.] [Erratum, *JAMA.* 2004; 291:2196.]
30. Ray WA. Policy and program analysis using administrative databases. *Ann Intern Med.* 1997; 127:712-8.
31. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2002.
32. Wagner AK, Soumerai SB, Zhang F et al. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther.* 2002; 27:299-309.