SECTION I

Basic Ingredients

CHAPTER 1

Getting Started: The Anatomy and Physiology of Clinical Research

Stephen B. Hulley, Thomas B. Newman, and Steven R. Cummings

This chapter introduces clinical research from two viewpoints, setting up themes that run together throughout the book. One is the **anatomy** of research—what it's made of. This includes the tangible elements of the study plan: research question, design, subjects, measurements, sample size calculation, and so forth. An investigator's goal is to design these components in a fashion that will make the project feasible and efficient.

The other theme is the **physiology** of research—how it works. Studies are useful to the extent that they yield valid inferences, first about what happened in the study sample and then about how these findings generalize to people outside the study. The goal is to minimize the errors, random and systematic, that threaten conclusions based on these inferences.

Separating the two themes is artificial in the same way that the anatomy of the human body doesn't make much sense without some understanding of its physiology. But the separation has the same advantage: It clarifies our thinking about a complex topic.

ANATOMY OF RESEARCH: WHAT IT'S MADE OF

The structure of a research project is set out in its **protocol**, the written plan of the study. Protocols are well known as devices for seeking grant funds and Institutional Review Board (IRB) approval, but they also have a vital scientific function: helping the investigator organize her research in a logical, focused, and efficient way. Table 1.1 outlines the components of a protocol. We introduce the whole set here, expand on each component in the ensuing chapters of the book, and return to put the completed pieces together in Chapter 19.

Research Question

The **research question** is the objective of the study, the uncertainty the investigator wants to resolve. Research questions often begin with a general concern that must be narrowed down to a concrete, researchable issue. Consider, for example, the general question:

- Should people eat more fish?

This is a good place to start, but the question must be focused before planning efforts can begin. Often this involves breaking the question into more specific components, and singling out one or two of these to build the protocol around:

- How often do Americans eat fish?
- Does eating more fish lower the risk of cardiovascular disease?
- Is there a risk of mercury toxicity from increasing fish intake in older adults?
- Do fish oil supplements have the same effects on cardiovascular disease as dietary fish?
- Which fish oil supplements don't make your breath smell like fish?

TABLE 1.1 ANATOMY OF RESEARCH: THE STUDY PLAN

DESIGN COMPONENTS	PURPOSE
Research questions	What questions will the study address?
Background and significance	Why are these questions important?
Design	How is the study structured?
Time frame	
Epidemiologic design	
Subjects	Who are the subjects and how will they be selected?
Selection criteria	
Sampling design	
Variables	What measurements will be made?
Predictor variables	
Confounding variables	
Outcome variables	
Statistical issues	How large is the study and how will it be analyzed?
Hypotheses	
Sample size	
Analytic approach	

A good research question should pass the "So what?" test. Getting the answer should contribute usefully to our state of knowledge. The acronym **FINER** denotes five essential characteristics of a good research question: It should be **feasible**, **interesting**, **novel**, **ethical**, and **relevant** (Chapter 2).

Background and Significance

A brief **background** and **significance** section in a protocol sets the proposed study in context and gives its rationale: What is known about the topic at hand? Why is the research question important? What kind of answers will the study provide? This section cites relevant previous research (including the investigator's own work) and indicates the problems with the prior research and what uncertainties remain. It specifies how the findings of the proposed study will help resolve these uncertainties, lead to new scientific knowledge, or influence practice guidelines or public health policy. Often, the literature review and synthesis done for the significance section will lead the investigator to modify the research question.

Design

The **design** of a study is a complex issue. A fundamental decision is whether to take a passive role in making measurements on the study subjects in an **observational study** or to apply an intervention and examine its effects in a **clinical trial** (Table 1.2). Among observational studies, two common designs are **cohort studies**, in which observations are made in a group of subjects that is followed over time, and **cross-sectional studies**, in which observations are made on a single occasion. Cohort studies can be further divided into **prospective** studies that begin in the present and follow subjects into the future, and **retrospective** studies that examine information collected over a period of time in the past. A third common option is the **case–control** design, in which the investigator compares a group of people who have a disease or other outcome with another group who do not. Among clinical trial options, the **randomized blinded trial** is

TABLE 1.2 EXAMPLES OF CLINICAL RESEARCH DESIGNS TO FIND OUT WHETHER FISH INTAKE REDUCES CORONARY HEART DISEASE RISK

EPIDEMIOLOGIC DESIGN	KEY FEATURE	EXAMPLE
	Observational Designs	
Cohort study	A group of subjects identified at the beginning and followed over time	The investigator measures fish intake in a group of subjects at baseline and periodically examines them at follow-up visits to see if those who eat more fish have fewer coronary heart disease (CHD) events.
Cross-sectional study	A group examined at one point in time	She interviews a group of subjects about current and past history of fish intake and correlates results with history of CHD and current coronary calcium score.
Case–control study	Two groups selected based on the presence or absence of an outcome	She examines a group of patients with CHD (the "cases") and compares them with a group who do not have CHD (the "controls"), asking about past fish intake.
	Clinical Trial Design	
Randomized blinded trial	Two groups created by a random process, and a blinded intervention	She randomly assigns subjects to receive fish oil supplements or a placebo that is identical in appearance, then follows both treatment groups for several years to observe the incidence of CHD.

usually the best design but nonrandomized or unblinded designs may be all that are feasible for some research questions.

No one approach is always better than the others, and each research question requires a judgment about which design is the most efficient way to get a satisfactory answer. The randomized blinded trial is often held up as the best design for establishing causality and the effectiveness of interventions, but there are many situations for which an observational study is a better choice or the only feasible option. The relatively low cost of case–control studies and their suitability for rare outcomes makes them attractive for some questions. Special considerations apply to choosing designs for studying diagnostic tests. These issues are discussed in Chapters 7 through 12, each dealing with a particular set of designs.

A typical sequence for studying a topic begins with observational studies of a type that is often called **descriptive**. These studies explore the lay of the land—for example, describing distributions of health-related characteristics and diseases in the population:

- What is the average number of servings of fish per week in the diet of Americans with a history of coronary heart disease (CHD)?

Descriptive studies are usually followed or accompanied by **analytic** studies that evaluate associations to permit inferences about cause-and-effect relationships:

- Do people with a CHD who eat a lot of fish have a lower risk of recurrent myocardial infarction than people with a history of CHD who rarely eat fish?

The final step is often a **clinical trial** to establish the effects of an intervention:

- Does treatment with fish oil capsules reduce total mortality in people with CHD?

Clinical trials usually occur relatively late in a series of research studies about a given question, because they tend to be more difficult and expensive, and to answer more definitively the narrowly focused questions that arise from the findings of observational studies.

It is useful to characterize a study in a *single sentence that summarizes the design and research question*. If the study has two major phases, the design for each should be mentioned.

- This is a cross-sectional study of dietary habits in 50- to 69-year-old people with a history of CHD, followed by a prospective cohort study of whether fish intake is associated with lower risk of subsequent coronary events.

This sentence is the research analog to the opening sentence of a medical resident's report on a new hospital admission: "This 62-year-old white policewoman was well until 2 hours before admission, when she developed crushing chest pain radiating to the left shoulder."

Some designs do not easily fit into the categories listed above, and classifying them with a single sentence can be surprisingly difficult. It is worth the effort—a concise description of the design and research question clarifies the investigator's thoughts and is useful for orienting colleagues and consultants.

Study Subjects

Two major decisions must be made in choosing the study subjects (Chapter 3). The first is to specify **inclusion** and **exclusion criteria** that define the target population: the *kinds* of people best suited to the research question. The second decision concerns how to **recruit** an appropriate *number* of people from an accessible subset of this population to be the subjects of the study. For example, the study of fish intake in people with CHD might identify subjects seen in the clinic with diagnostic codes for myocardial infarction, angioplasty, or coronary artery bypass grafting in their electronic medical record. Decisions about which patients to study often represent trade-offs; studying a random sample of people with CHD from the entire country (or at least several different states and medical care settings) would enhance **generalizability** but be much more difficult and costly.

Variables

Another major set of decisions in designing any study concerns the choice of which variables to measure (Chapter 4). A study of fish intake in the diet, for example, might ask about different types of fish that contain different levels of omega-3 fatty acids, and include questions about portion size, whether the fish was fried or baked, and use of fish oil supplements.

In an analytic study the investigator studies the associations among variables to predict outcomes and to draw inferences about cause and effect. In considering the association between two variables, the one that occurs first or is more likely on biologic grounds to be causal is called the **predictor variable**; the other is called the **outcome variable**.[1] Most observational studies have many predictor variables (age, race, sex, smoking history, fish and fish oil supplement intake) and several outcome variables (heart attacks, strokes, quality of life, unpleasant odor).

Clinical trials examine the effects of an **intervention**—a special kind of predictor variable that the investigator manipulates, such as treatment with fish oil capsules. This design allows her to observe the effects on the outcome variable using **randomization** to minimize the influence of **confounding variables**—other predictors of the outcome such as smoking or income level that could be associated with dietary fish and confuse the interpretation of the findings.

[1]Predictors are sometimes termed **independent variables** and outcomes **dependent variables**, but the meaning of these terms is less self-evident and we prefer to avoid their use.

Statistical Issues

The investigator must develop plans for estimating sample size and for managing and analyzing the study data. This generally involves specifying a **hypothesis** (Chapter 5).

> **Hypothesis:** 50- to 69-year-old women with CHD who take fish oil supplements will have a lower risk of recurrent myocardial infarction than those who do not.

This is a version of the research question that provides the basis for testing the **statistical significance** of the findings. The hypothesis also allows the investigator to calculate the **sample size**—the number of subjects needed to observe the expected difference in outcome between study groups with reasonable probability (an attribute known as **power**) (Chapter 6). Purely descriptive studies (what proportion of people with CHD use fish oil supplements?) do not involve tests of statistical significance, and thus do not require a hypothesis; instead, the number of subjects needed to produce acceptably narrow **confidence intervals** for means, proportions, or other descriptive statistics can be calculated.

PHYSIOLOGY OF RESEARCH: HOW IT WORKS

The goal of clinical research is to draw **inferences** from findings in the study about the nature of the universe around it. Two major sets of inferences are involved in interpreting a study (illustrated from right to left in Figure 1.1). Inference #1 concerns **internal validity**, the degree to which the investigator draws the correct conclusions about what actually happened in the study. Inference #2 concerns **external validity** (also called **generalizability**), the degree to which these conclusions can be appropriately applied to people and events outside the study.

When an investigator plans a study, she reverses the process, working from left to right in the lower half of Figure 1.1 with the goal of maximizing the validity of these inferences at the end of the study. She **designs a study plan** in which the choice of research question, subjects, and measurements enhances the external validity of the study and is conducive to **implementation** with a high degree of internal validity. In the next sections we address design and then implementation before turning to the errors that threaten the validity of these inferences.

Designing the Study

Consider this simple descriptive question:

> What is the prevalence of daily ingestion of fish oil supplements among people with CHD?

This question cannot be answered with perfect accuracy because it would be impossible to study all patients with CHD and our approaches to discovering whether a person has CHD

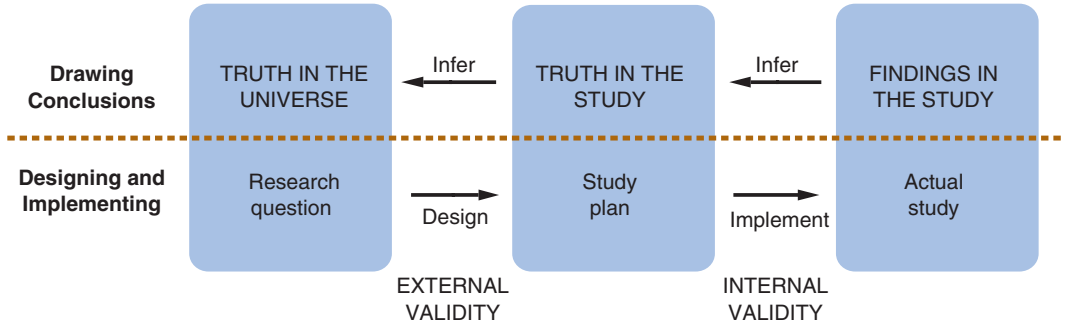


FIGURE 1.1 The process of designing and implementing a research project sets the stage for drawing conclusions based on inferences from the findings.

and is taking fish oil are imperfect. So the investigator settles for a related question that *can* be answered by the study:

> Among a sample of patients seen in the investigator's clinic who have a previous CHD diagnosis and respond to a mailed questionnaire, what proportion report taking daily fish oil supplements?

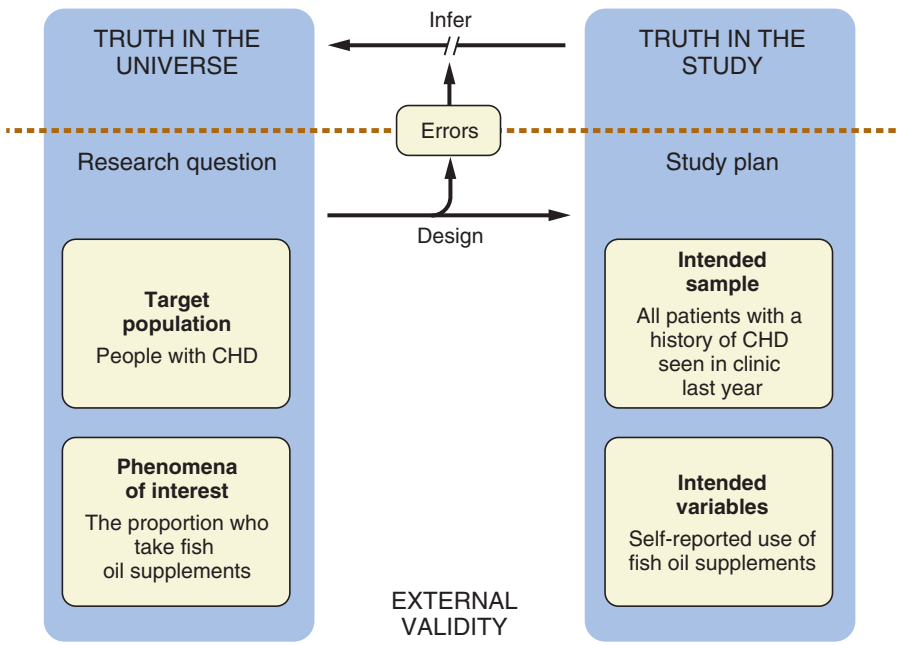
The transformation from research question to study plan is illustrated in Figure 1.2. One major component of this transformation is the choice of a **sample** of subjects that will represent the **population**. The group of subjects specified in the protocol can only be a sample of the population of interest because there are practical barriers to studying the entire population. The decision to study patients in the investigator's clinic identified through the electronic medical record system is a compromise. This is a sample that is feasible to study but has the disadvantage that it may produce a different prevalence of fish oil use than that found in all people with CHD.

The other major component of the transformation is the choice of **variables** that will represent the **phenomena of interest**. The variables specified in the study plan are usually proxies for these phenomena. The decision to use a self-report questionnaire to assess fish oil use is a fast and inexpensive way to collect information, but unlikely to be perfectly accurate because people usually do not accurately remember or record how much they take in a typical week.

In short, each of the differences in Figure 1.2 between the research question and the study plan has the purpose of making the study more practical. The cost of this increase in practicality, however, is the risk that design choices may cause the study to produce a wrong or misleading conclusion because it is designed to answer a somewhat different question from the research question of interest.

Implementing the Study

Returning to Figure 1.1, the right-hand side is concerned with **implementation** and the degree to which the actual study matches the study plan. At issue here is the problem of a wrong answer



■ **FIGURE 1.2** Design errors and external validity: If the intended sample and variables do not sufficiently represent the target population and phenomena of interest, these errors may distort inferences about what actually happens in the population.

to the research question because the way the sample was actually drawn, or the measurements made, differed in important ways from the way they were designed (Figure 1.3).

The actual sample of study subjects is almost always different from the intended sample. The plans to study all eligible clinic patients with CHD, for example, could be disrupted by incomplete diagnoses in the electronic medical record, wrong addresses for the mailed questionnaire, and refusal to participate. Those subjects who are reached and agree to participate may have a different prevalence of fish oil use than those not reached or not interested. In addition to these problems with the subjects, the actual measurements can differ from the intended measurements. If the format of the questionnaire is unclear subjects may get confused and check the wrong box, or they may simply omit the question by mistake.

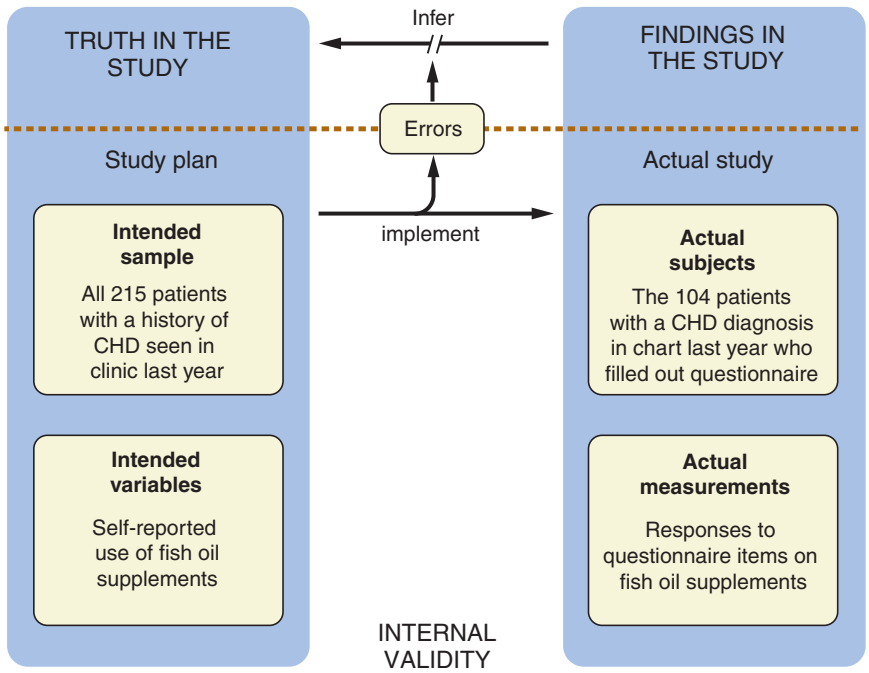
These differences between the study plan and the actual study can alter the answer to the research question. Figure 1.3 illustrates that errors in implementing the study join errors of design in leading to a misleading or wrong answer to the research question.

Causal Inference

A special kind of validity problem arises in studies that examine the **association** between a predictor and an outcome variable in order to draw causal inference. If a cohort study finds an association between fish intake and CHD events, does this represent cause and effect, or is fish intake just an innocent bystander in a web of causation that involves other variables? Reducing the likelihood of **confounding** and other rival explanations is one of the major challenges in designing an observational study (Chapter 9).

The Errors of Research

Recognizing that no study is entirely free of errors, the goal is to maximize the validity of inferences from what was observed in the study sample to what is happening in the population.



■ **FIGURE 1.3** Implementation errors and internal validity: If the actual subjects and measurements do not sufficiently represent the intended sample and variables, these errors may distort inferences about what happened in the study.

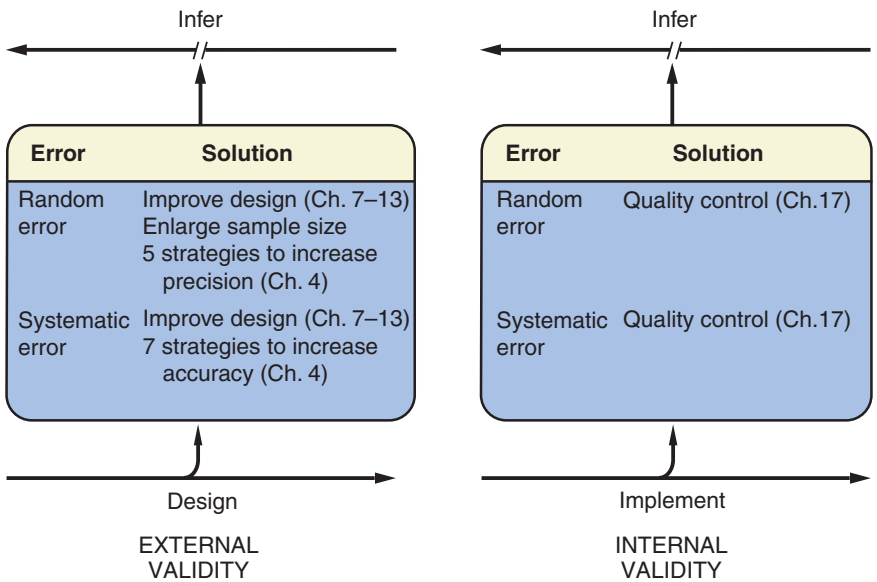
Erroneous inferences can be addressed in the analysis phase of research, but a better strategy is to focus on design and implementation (Figure 1.4), preventing errors from occurring in the first place to the extent that this is practical.

The two main kinds of errors that interfere with research inferences are random error and systematic error. The distinction is important because the strategies for minimizing them are quite different.

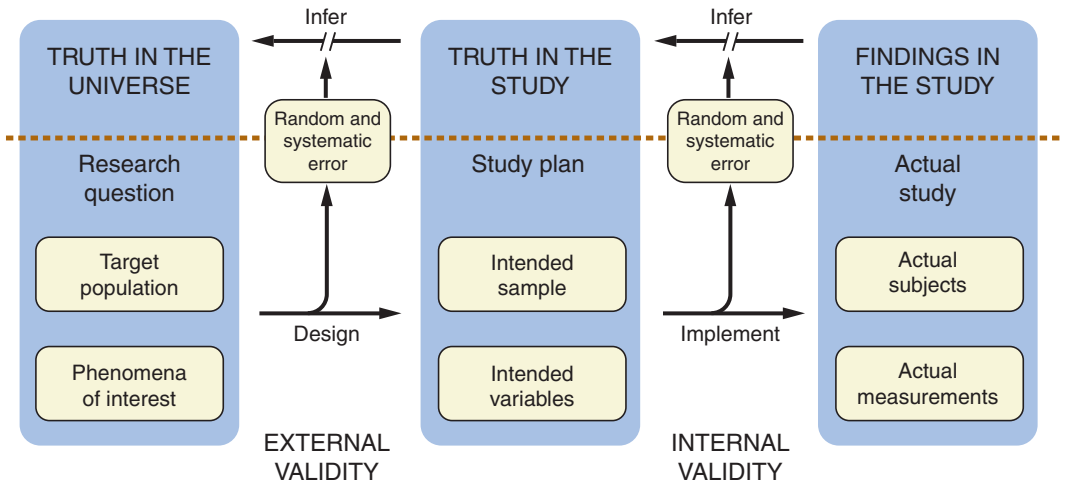
Random error is a wrong result due to **chance**—sources of variation that are equally likely to distort measurements from the study in either direction. If the true prevalence of daily fish oil supplement use in the several hundred 50- to 69-year-old patients with CHD in the investigator's clinic is 20%, a well-designed sample of 100 patients from that population might contain exactly 20 patients who use these supplements. More likely, however, the sample would contain a nearby number such as 18, 19, 21, or 22. Occasionally, chance would produce a substantially different number, such as 12 or 28. Among several techniques for reducing the influence of random error (Chapter 4), the simplest is to increase the sample size. The use of a larger sample diminishes the likelihood of a substantially wrong result by increasing the **precision** of the estimate—the degree to which the observed prevalence approximates 20% each time a sample is drawn.

Systematic error is a wrong result due to **bias**—sources of variation that distort the study findings in one direction. An illustration is the decision in Figure 1.2 to study patients in the investigator's clinic, where the local treatment patterns have responded to her interest in the topic and her fellow doctors are more likely than the average doctor to recommend fish oil. Increasing the sample size has no effect on systematic error. The best way to improve the **accuracy** of the estimate (the degree to which it approximates the true value) is to design the study in a way that reduces the size of the various biases. Alternatively, the investigator can seek additional information to assess the importance of possible biases. An example would be to compare results with those from a second sample of patients with CHD drawn from another setting, for example, examining whether the findings of such patients seen in a cardiology clinic are different from those seen in a primary care clinic.

The examples of random and systematic error in the preceding two paragraphs are components of **sampling error**, which threatens inferences from the study subjects to the population.



■ **FIGURE 1.4** Research errors. This blown-up detail of the error boxes in Figures 1.2 and 1.3 reveals strategies for controlling random and systematic error in the design and implementation phases of the study.



■ **FIGURE 1.5** Physiology of research—how it works.

Both random and systematic errors can also contribute to **measurement error**, threatening the inferences from the study measurements to the phenomena of interest. An illustration of random measurement error is the variation in the response when the diet questionnaire is administered to the patient on several occasions. An example of systematic measurement error is underestimation of the prevalence of fish oil use due to lack of clarity in how the question is phrased. Additional strategies for controlling all these sources of error are presented in Chapters 3 and 4.

The concepts presented in the last several pages are summarized in Figure 1.5 Getting the right answer to the research question is a matter of designing and implementing the study in a fashion that minimizes the magnitude of inferential errors.

■ DESIGNING THE STUDY

Study Plan

The process of developing the **study plan** begins with the one-sentence **research question** that specifies the main predictor and outcome variables and the population. Three versions of the study plan are then produced in sequence, each larger and more detailed than the preceding one.

- **Study outline** (Table 1.1 and Appendix 1). This one-page summary of the design serves as a standardized checklist to remind the investigator to address all the components. As important, the sequence has an orderly logic that helps clarify the investigator's thinking on the topic.
- **Study protocol.** This expansion on the study outline usually ranges from 5 to 15 pages, and is used to plan the study and to apply for IRB approval and grant support. The protocol parts are discussed throughout this book and summarized in Chapter 19.
- **Operations manual.** This collection of specific procedural instructions, questionnaires, and other materials is designed to ensure a uniform and standardized approach to carrying out the study with good quality control (Chapters 4 and 17).

The research question and study outline should be written out at an early stage. Putting thoughts down on paper leads the way from vague ideas to specific plans and provides a concrete basis for getting advice from colleagues and consultants. It is a challenge to do it (ideas are easier to talk about than to write down), but the rewards are a faster start and a better project.

Appendix 1 is an example of a study outline. This one-page outline deals more with the anatomy of research (Table 1.1) than with its physiology (Figure 1.5), so the investigator must remind herself to worry about the errors that may result when it is time to draw inferences

from measurements in the study sample to phenomena of interest in the population. A study's virtues and problems can be revealed by explicitly considering how the question the study is likely to answer differs from the research question, given the plans for acquiring subjects and making measurements, and given the likely problems of implementation.

With the study outline in hand and the intended inferences in mind, the investigator can proceed with the details of her protocol. This includes getting advice from colleagues, drafting specific recruitment and measurement methods, considering scientific and ethical appropriateness, modifying the study question and outline as needed, pretesting specific recruitment and measurement methods, making more changes, getting more advice, and so forth. This iterative process is the nature of research design and the topic of the rest of this book.

Trade-offs

Regretably, errors are an inherent part of all studies. The main issue is whether the errors will be large enough to change the conclusions in important ways. When designing a study, the investigator is in much the same position as a labor union official bargaining for a new contract. The union official begins with a wish list—shorter hours, more money, health care benefits, and so forth. She must then make concessions, holding on to the things that are most important and relinquishing those that are not essential or realistic. At the end of the negotiations is a vital step: She looks at the best contract she could negotiate and decides if it has become so bad that it is no longer worth having.

The same sort of concessions must be made by an investigator when she transforms the research question to the study plan and considers potential problems in implementation. On one side are the issues of internal and external validity; on the other, feasibility. The vital last step of the union negotiator is sometimes omitted. Once the study plan has been formulated, the investigator must decide whether it adequately addresses the research question and whether it can be implemented with acceptable levels of error. Often the answer is no, and there is a need to begin the process anew. But take heart! Good scientists distinguish themselves not so much by their uniformly good research ideas as by their alacrity in turning over those that won't work and moving on to better ones.

SUMMARY

1. The **anatomy** of research is the set of tangible elements that make up the study plan: the **research question** and its **significance**, and the **design**, **study subjects**, and **measurement approaches**. The challenge is to design elements that are relatively **inexpensive** and **easy** to implement.
2. The **physiology** of research is how the study works. The study findings are used to draw **inferences** about what happened in the study sample (**internal validity**), and about events in the world outside (**external validity**). The challenge here is to **design** and **implement** a study plan with adequate control over two major threats to these inferences: **random error** (chance) and **systematic error** (bias).
3. In designing a study the investigator may find it helpful to consider Figure 1.5, the relationships between the **research question** (what she wants to answer), the **study plan** (what the study is designed to answer), and the **actual study** (what the study will actually answer, given the errors of implementation that can be anticipated).
4. A good way to develop the **study plan** is to begin with a one-sentence version of the **research question** that specifies the main variables and population, and expand this into a one-page **outline** that sets out the study elements in a standardized sequence. Later on the study plan will be expanded into the **protocol** and the **operations manual**.
5. Good **judgment** by the investigator and **advice** from colleagues are needed for the many **trade-offs** involved, and for determining the overall viability of the project.

APPENDIX 1
Outline of a Study

This is the one-page study plan of a project carried out by Valerie Flaherman, MD, MPH, begun while she was a general pediatrics fellow at UCSF. Most beginning investigators find observational studies easier to pull off, but in this case a randomized clinical trial of modest size and scope was feasible, the only design that could adequately address the research question, and ultimately successful—see publication by Flaherman et al (1) for the findings, which, if confirmed, could alter policy on how best to initiate breast feeding.

TITLE: EFFECT OF EARLY LIMITED FORMULA USE ON BREASTFEEDING

Research question:

Among term newborns who have lost ≥ 5% of their birth weight before 36 hours of age, does feeding 10 cc of formula by syringe after each breastfeeding before the onset of mature milk production increase the likelihood of subsequent successful breastfeeding?

Significance:

1. Breast milk volume is low until mature milk production begins 2–5 days after birth.
2. Some mothers become worried if the onset of mature milk production is late and their baby loses a lot of weight, leading them to abandon breastfeeding within the first week. A strategy that increased the proportion of mothers who succeed in breastfeeding would have many health and psycho-social benefits to mother and child.
3. Observational studies have found that formula feeding in the first few days after birth is associated with decreased breastfeeding duration. Although this could be due to confounding by indication (see Chapter 9), the finding has led to WHO and CDC guidelines aimed at reducing the use of formula during the birth hospitalization.
4. However, a small amount of formula combined with breastfeeding and counseling might make the early breastfeeding experience more positive and increase the likelihood of success. A clinical trial is needed to assess possible benefits and harms of this strategy.

Study design:

Unblinded randomized control trial with blinded outcome ascertainment

Subjects:

- **Entry criteria:** Healthy term newborns 24–48 hours old who have lost ≥ 5% of their birth weight in the first 36 hours after birth
- **Sampling design:** Consecutive sample of consenting patients in two Northern California academic medical centers

Predictor variable, randomly assigned but not blinded:

- **Control:** Parents are taught infant soothing techniques.
- **Intervention:** Parents are taught to syringe-feed 10 cc of formula after each breastfeeding until the onset of mature milk production.

Outcome variables, blindly ascertained:

1. Any formula feeding at 1 week and 1, 2, and 3 months
2. Any breastfeeding at 1 week and 1, 2, and 3 months
3. Weight nadir

Primary null hypothesis:

Early limited formula does not affect the proportion of women who are breastfeeding their baby at 3 months.

REFERENCE

1. Flaherman VJ, Aby J, Burgos AE, et al. Effect of early limited formula on duration and exclusivity of breastfeeding in at-risk infants: an RCT. *Pediatrics*, in press.