Chapter 1

METHODOLOGY: WHAT IT IS AND WHY IT IS SO IMPORTANT

Alan E. Kazdin

Scientific knowledge is very special. This knowledge is based on the accumulation of empirical evidence and obtained through systematic and careful observation of the phenomenon of interest. At a very general level, the ways in which the observations are obtained constitute the methods of science. Yet, these methods can be considered at multiple levels, including the principles and tenets they are designed to reflect, a way of thinking and problem solving, and concrete practices that scientists use when actually conducting an investigation. This book draws on each of these levels because they work together and make for good science and scientific research.

The purpose of this introductory chapter is to convey what methodology is, why it is needed, and the key tenets that guide what we do as scientists. These foci may seem obvious—after all, everyone knows what methodology is and why it is needed. Perhaps so, but the answers are not all so obvious. It is useful to give the rationale for what we do and why because it provides the common base we as psychologists and social scientists share with all of the sciences. Also, that base underpins all of the chapters that follow. Let us begin.

SCIENTIFIC METHODOLOGY AND ITS COMPONENTS

Methodology in science refers to the diverse principles, procedures, and practices that govern empirical research. It is useful to distinguish five major components to convey the scope of the topics and to organize the subject matter.

1. *Research design*: This component refers to the experimental arrangement or plan used to examine the question or hypotheses of interest. It includes fundamental issues related to who the participants will be, how they will be assigned (e.g., randomly), and the comparisons (various groups) included in the study. Many different arrangements exist, including those in which some experimental manipulation is made (true experiments) or groups are formed (observational study), by which to evaluate differences in characteristics of interest.
2. *Assessment*: This component pertains to the measurement strategies (e.g., self-report, neuroimaging) and the measures that will be used to provide the data. There are many different types of measures and multiple measures within each type. Key issues related to assessment, such as reliability and validity of the measures, are pivotal to research.
3. *Data evaluation and interpretation*: This component encompasses all of the methods that will be used to handle the data—to characterize the sample, to describe performance on the measures, and to draw inferences related to the hypotheses. Statistical significance testing is dominant and the most familiar method used to develop and evaluate data but, as later chapters show, other methods are also used.

http://dx.doi.org/10.1037/14805-001
Methodological Issues and Strategies in Clinical Research, Fourth Edition, A. E. Kazdin (Editor)

4. *Ethical issues and scientific integrity*: This multifaceted component includes a variety of responsibilities that the investigator has in the conduct of the study and can encompass all of the other components (e.g., design, data analyses, publication of findings). Ethical issues include multiple responsibilities to participants (e.g., their rights and protections) and adherence to the professional standards of one's discipline (e.g., ethical codes). Scientific integrity includes responsibilities to the scientific community and the public at large (e.g., transparency, accurately reporting findings) and is also part of professional standards and ethical codes. Before a study begins, proposals are usually required (e.g., by universities, agencies) that discuss not only specifics of the project (e.g., research design, assessment) but also ethical issues and assurances that participant rights are protected (e.g., scrutiny of the procedures for any untoward effects, informed consent, protection of privacy).
5. *Communication of research findings*: Communication of our work is key to building the knowledge base, stimulating responses to our work, and promoting and fostering new theory and findings as we ourselves or others follow up on the study we have described. Findings can be communicated to other professionals through many different venues (e.g., journal articles on empirical studies, review articles, conference symposium presentations, poster sessions). Communication also includes the media (dissemination of information to the public via TV, radio, and the web). Communication of findings has its own responsibilities and challenges, as discussed later.

I have divided methodology into these components in part to convey the breadth and depth of the topic. There are books, courses, and journals devoted specifically to each of these components. As one example, psychological assessment is an enormous topic encompassing models of scale development, validation, the vast range of assessment modalities, and sources of artifact and bias that can greatly affect data obtained from a measure. Similarly, data analyses and the vast array of statistical models and analyses have their own courses and journals. This book covers all five components and does so in a way that underscores their integration and interrelation. There are always more topics and components of methodology one could add. For example, the historical roots of science and science and social policy are legitimate topics that could be covered as well. Yet, in developing an appreciation for methodology and the skills involved in many of the key facets of actually conducting research, the five will suffice.

WHY DO WE NEED SCIENCE AND ITS METHODS AT ALL?

Rationale

I have already mentioned the components of scientific methods, but now let us step back a bit. Why do we even need methodology in general and its components? Four reasons can make the case for why we need science and the methodology of science. First, we need consistent methods for acquiring knowledge. There are many sciences, and it would be valuable, if not essential, to have principles and practices that are consistent across them all. We would not want the criteria for what counts as knowledge to vary as a function of quite different ways of going about obtaining that knowledge. This consistency is more important than ever today, because much of research on a given topic involves the collaboration of scientists from many different fields and many different countries to address a set of questions for a given project. Scientists from many different areas must speak the same methodology language, share the same underlying values about how to obtain knowledge, and agree on procedures and practices (e.g., statistical evaluation, reporting data that do and do not support a particular hypothesis). Consistency is also critical within any given scientific discipline. For a given science (e.g., psychology), we would want consistency throughout the world in the standards for obtaining scientific knowledge—the accumulation of knowledge from all individuals in a given field requires this level of consistency. Science says, essentially, these are our goals (e.g., describe; understand; explain; intervene when needed, possible, and desirable) and these are our means (use of theory, methodology, guiding concepts, replication of results). Science is hardly a game because so many of its tasks and topics are so serious—indeed, a matter of life and death (e.g., suicide, risky behavior, cigarette smoking). Yet there are rules and there are enormous benefits to be gained by all sciences and scientists. Think of the chaos if methods varied across countries or professions; we simply could not accumulate an agreed-on body of knowledge.

Second, methodology is needed to identify, detect, isolate, and reveal many of the extremely complex relations that exist in the world. Science uses special controlled arrangements and special methods (e.g., equipment, measures) to isolate influences that are otherwise difficult, if not impossible, to detect from casual observation in everyday life. Consider a brief sample of findings from the natural and social sciences conveying the complexities of our world that the methods of science were needed to reveal. Consider the guiding question in the examples and the answers that scientific method provided:

- What is near the boundary of our universe? Well, for starters, a galaxy (a system of millions of stars held by gravitational attraction) has been identified that is more than 13 billion light-years away (e.g., Maartens, 2013).
- How did dinosaurs become extinct? Approximately 66 million years ago (give or take 300,000 years), a huge asteroid (15 kilometers, or more than 16,400 yards, wide) crashed into the earth (near Yucatan, Mexico) and led to the extinction of more than half of all species on the planet, including the dinosaurs. The material blasted into the atmosphere led to a chain of events that resulted in a global winter (e.g., Brusatte et al., 2014).
- Are male and female interactions and behaviors influenced by a woman's menstrual cycle? Where a woman is in her menstrual cycle apparently has an effect on her behavior (e.g., selection of clothing, gait when walking, and the type of man that seems attractive) and how men respond to it. All of this occurs outside of consciousness but conveys dynamically changing interactions influenced in part by ovulation cycles (e.g., Haselton & Gildersleeve, 2011).
- When prisoners come before a parole board, are there any unexpected influences on the decision of whether they can be released before their prison sentence is complete? Surprisingly, the point during the day at which a given prisoner sees the parole board is relevant to the outcome. An evaluation of multiple parole decisions revealed that the likelihood of being granted parole is much higher in the morning and immediately after a lunch break than at other times (Danziger, Levav, & Avnaim-Pesso, 2011). Indeed, as hunger (or fatigue) increases and as lunch time approaches, the chances of being paroled decrease, but they bounce up again right after the lunch break. The same raters were involved, and the result cannot be explained by severity of the crimes or types of prisoners.
- Do early harsh environments for children (e.g., exposure to violence, enduring stress, corporal punishment) have any long-term effects? Yes, they can lead to many untoward outcomes, including poor academic performance (e.g., poor grades, dropping out of school) and mental illness (e.g., posttraumatic stress disorder, depression, anxiety). Also, the outcomes can include enduring impairment of the immune system (ability to ward off infection and inflammation) and are likely the reason why many such children have premature deaths from serious disease much later in adulthood (e.g., Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002).

The findings in these examples required very special observation procedures under special arrangements, measures, assessments, and methods of data evaluation. The conclusions I list are not discernible by everyday observation. If you said, "I knew all along based on my casual observations that there was a galaxy at the boundaries of our universe; what's the big deal?" or "Of course prisoners who are seen after the parole board's lunch break are more likely to be granted parole," you are among a very elite group. The rest of us needed careful research and scientific methods to grasp these phenomena!

Third, whether the relations are complex or not, for many questions of interest extensive information (a lot of data) are needed to draw conclusions. How to obtain that information (assessment, sampling) requires very special procedures to yield trustworthy results. For example, how many individuals experience some form of psychiatric disorder? To answer this question, one needs a large sample, a representative sample, and special procedures (e.g., use of measures known to provide consistent information and to reflect the phenomenon of interest). As it turns out, approximately 25% of the U.S. population at a given point in time meet criteria for one or more psychiatric disorders (Kessler & Wang, 2008). Approximately 50% experience a disorder at some point in their lifetime. This kind of information cannot be obtained from casual observation or individual experience. Large data sets and systematically collected data are needed to address many questions, and science is needed to provide the information in a trustworthy, consistent, transparent, and replicable way.

Finally, we need science to help surmount the limitations of our usual ways of perceiving the environment and reaching conclusions. Along with these limitations in our perceptions, there are many sources of subjectivity and bias that interfere with obtaining more objective knowledge—that is, information that is as free as possible from subjectivity and bias. How we perceive and think is wonderfully adaptive for handling everyday life and the enormous challenges presented to us (e.g., staying out of danger, finding mates and partners, rearing children, adapting to harsh and changing environments, meeting the biological needs of ourselves and our family—it is endless). Evolution spanning millions of years has sculpted, carved, sanded, and refined these skills. Yet those very adaptive features can actually interfere, limit, and distort the information presented to us and do so by omission (our perception omits many facets of experience that we do not detect well) and by commission (we actively distort information on a routine basis). Scientific methodology has emerged in part to surmount the limitations of more casual observation.

That said, a few limitations are worth noting. Science does not get rid of these limitations. Rather, methodological practices are designed to help manage and overcome them.

Brief Illustrations of Our Limitations in Accruing Knowledge

Senses and their limits. The limitations of our senses—including vision, hearing, and smell—serve as a familiar example to convey how very selective we are in the facets of reality that we can detect. We consider what we see, hear, and smell to represent reality, that is, how things are. But this reality is very selective. For example, we see only a small portion of the electromagnetic spectrum and refer to that as the *visible spectrum.* Probably a better term would be the *human visual spectrum.* We cannot see infrared, or ultraviolet, for example. Other animals (e.g., birds, bees and many other insects) see part of the spectrum we do not see, which helps with their adaptation (e.g., identifying sex-dependent markings of potential mates that are only visible in ultraviolet light). The same is true for sounds and smells; many nonhuman animals have senses that evaluate different parts of the world from those we can experience. Many animals can hear sounds that we do not hear (e.g., dogs, elephants, pigeons) and have a sensitivity to smell that vastly exceeds our own (e.g., bears, sharks, moths, bees). More generally, many nonhuman animals trump our vision, hearing, and smell or have differences that are not better (more sensitive) or worse but just different (e.g., seeing different parts of the electromagnetic spectrum).

These examples are intended to make one point: As humans, we see one part of the world, and that picture is quite selective. The picture we have of what is omits piles of things that are. So one reason for science is to overcome some of the physical limitations of our normal processing of information. Much of what we want to know about and see cannot be discerned with our ordinary capacities (our senses). In fact, much of what we have learned about the universe and also about interpersonal interaction and attraction comes from what is not obvious, detected, or detectable by means of usual sensory perception.

Cognitive heuristics. Leaving aside physical limitations in seeing, smelling, and hearing the world, more persuasive arguments for the need for science come from many areas of cognitive psychology. These arguments are more persuasive in the sense that when we look at experience well within the capacities of our senses, we may still have enormous limitations in how we process that information. You already know the everyday expression "seeing is believing"; psychological research has provided considerable support for the additional claim "believing is seeing." We process the world in special ways, and various cognitive processes have been well studied. These processes can and often do systematically distort and lead us to make claims and inferences that do not reflect reality, as revealed by less biased or unbiased means.

Several characteristics of normal human functioning, referred to as *cognitive heuristics*, reflect how we organize and process information. These processes are out of our awareness and serve as mental shortcuts or guides to help us negotiate many aspects of everyday experience (Kahneman, 2011; Pohl, 2012). These guides help us categorize, make decisions, and solve problems. The heuristics emerge as bias when we attempt to draw accurate relations based only on our own thoughts, impressions, and experience. There are several cognitive heuristics, but let me convey a sample to make concrete what I am talking about.

The *confirmatory bias* reflects the role of our preconceptions or beliefs and how they influence the facets of reality we see, grasp, and identify. Specifically, we select, seek out, and remember evidence in the world that is consistent with and supports our view. That is, we do not consider and weigh all experience or the extent to which some things are or are not true on the basis of the realities we encounter. Rather, we unwittingly pluck out features of reality that support (confirm) our view. This is particularly pernicious in stereotypes, as one case in point. For example, experimental manipulation of ethnic characteristics (e.g., skin tone among African Americans, ethnicity of victims in a crime) leads to different evaluations of crime and sentencing practices (e.g., Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). Objective facts about the material presented can be carefully controlled in research to allow demonstration of ethnic biases in how participants react to stereotypes and biases they would not otherwise express. More generally, if we believe that one ethnic group behaves in this or that way or that people from one country or region have a particular characteristic, we will see evidence that supports it—the supportive evidence is more salient in our mind and memory and is constructed rather than recording the incoming data objectively. Counterevidence does not register as salient or, if and when it does, is dismissed as an exception.

Cognitive heuristics are not the only set of influences that guide our perception. Our motivation and mood states can directly influence how and what we perceive of reality (Dunning & Balcetis, 2013). Both biological states (e.g., hunger, thirst) and psychological states (e.g., mood) can directly guide how reality is perceived. This is sometimes referred to as *motivated perception* or *wishful perceiving*. For example, when a person feels threatened or angry, he or she is likely to see another as holding a weapon rather than a neutral object (Baumann & DeSteno, 2010). That is, the reality we perceive is influenced by us as a filter, and our changing biological and psychological states have an impact on what we see, hear, and recall. Obviously, motivated perception can have life-and-death consequences because the person perceiving (e.g., civilian, police officer) feels threatened and acts accordingly. We are not likely to be empathic when we hear a person shot someone else when in fact there was no danger. The "in fact" may not have been so relevant because the perception of the individual who fired was guided by perceived threat. My comments are not about blame or justification; rather, they are intended to convey that reality is filtered and that filter can be biased and influenced in ways quite different from the actual facts or events.

Memory. Other examples illustrate how our normal processing of information influences and distorts and, again, why we need assistance from methodology to help surmount these influences. *Memory* refers to the ability to recall information and events, although there are different kinds of memory and different ways of studying them. As humans we believe (and are often confident) that our memory records reality, but research has very clearly shown that we recode reality (Roediger & McDermott, 2000). That is, more often than not we do not recall things as they happened. And this has come up in many contexts.

First, as we consider stories of our past (e.g., childhood, high school years), little details and sometimes larger ones get filled in and become part of our remembered story. Our memories draw on information for experience of the external world, but these memories are filled in by internal processes (e.g., imagination, thought). As we recount a story, we cannot distinguish between what in the story actually happened and what did not.

Reality monitoring is the name for the memory function that differentiates memories that are based on external (the world) versus internal (one's own thoughts, perceptions; Johnson, 2006). Thus, I can separate my imagined phone call from the Nobel committee (last night's dream) from reality (the phone call I actually received yesterday was from my dry cleaner—I had to pick up my shirts, or they would be thrown out). Errors occur when that distinction is not made, which is a function of several things, including how vivid the imagined events are and how consistent they are with the external stimuli. We develop a story or scheme of an event or occurrence and fill in details where needed, and when we recall the event, we cannot always distinguish the source. Sometimes our own mind fills in details, and sometimes this process is aided by the stories others have told us that become our stories and are planted as part of our experience.

Second and related, the notion of false memories has appeared in the public as well as the scientific literature. Interest in false memories emerged from the experiences of many clients in therapy who, over the course of treatment, newly recalled childhood experience of abuse. In several cases, it in fact appears as though the memories were actually induced by the very process of therapy. This does not mean, of course, that all, most, or any given recollection of abuse is false, but we know that some are, and that is just enough to establish that it can happen. Researchers have studied false memories—can we induce them in stories, memory tasks, and laboratory studies (e.g., Brainerd & Reyna, 2005)? Yes, experiments have shown that they can even be implanted. In these experiments, when people were asked to recall material, false memories (things that did not occur at all) were often recalled and mixed with those that had occurred. The key is that people do not see them as false memories, nor do they flag some memories as accurate or true and others as implanted. When someone says he or she remembers something perfectly or well, it may be useful to regard that as a statement of confidence in a memory rather than in accuracy of the account.

Finally, consider recall, used heavily by the courts in legal proceedings. In jury trials, the most persuasive type of evidence is eyewitness testimony. Juries are persuaded by a witness on the stand saying he or she saw the defendant do this or that and perhaps even identified the defendant in a lineup as the perpetrator. The reliance on eyewitness testimony makes forensic psychologists want to jump out of their basement windows because rather extensive research has shown that this type of testimony is the least reliable form of evidence and is responsible for sending more innocent victims to prison than any other form of evidence (Wells & Loftus, 2013). Clearly, memory, perception, and confident accounts must give one pause or caution.

General comments. Several facets of perception, thoughts, and emotions influence how we characterize the world, although I have mentioned only a small sample (e.g., only one cognitive heuristic, although there are several; only a few areas of memory research, including reality monitoring, false memories, and eyewitness testimony, while omitting others). The point was just to convey that, as humans, we have limitations that can readily influence the conclusions we reach. These limitations can have little impact (e.g., details regarding who was at a social event last month and who drank and ate what) or enormous impact (e.g., who goes to jail or receives the death penalty). Also, humans negotiate life rather well. As a rule, we do not bump into buildings or each other when walking down the street, put on our clothing correctly most days, and say "hi" rather than "goodbye" when we first encounter a friend or colleague during the day. So we should not distrust our senses, cognition, and affect or be dismissive in any way about their utility.

Accumulating scientific knowledge is another story. The limitations I have illustrated convey how essential it is to develop the means to counter normal experience, perception, memory, and the like in developing a knowledge base of our world. The challenge is as follows: We know we have limitations in our perception and hence in our ability to acquire unbiased knowledge without some systematic set of aids. We need more reliable tools to codify current and past experience and surmount some of our normal recall and other limitations. The paradox is this: We ourselves, with these imperfections, have the responsibility for developing the tools (methods) to surmount those limitations.

Think of science as a way of knowing filled with checks and balances. For example, one check, arguably the most important, is repetition of findings by other investigators. This repetition of findings is referred to as *replication*. For example, if I find an amazing result and no other investigator can reproduce (replicate) that result after many excellent tries, my finding is suspect. I am not necessarily suspected of anything odd, but the finding is not reliable. Perhaps the finding depended on something no one knows about or occurred by chance, was a fluke, or happened because of a bias I did not detect or control. At this moment in our discussion, the reason does not matter (although all of this will be discussed later), but we have to say that my finding is not to be taken as a reliable finding, and we go on. Perhaps some people replicate my finding, but others do not. This suggests that some other condition or circumstance (e.g., some characteristic of the participants, how the experimental manipulation was conducted) may influence whether the finding is obtained. These possibilities can be readily studied. If my study cannot be replicated, that is annoying at the moment, but we are committed to the process, and the last thing we want is false knowledge—that is, findings that do not hold up across investigators, laboratories, and time.

Methodology does not eliminate bias and problems, and so a great dose of humility about the process is just wise. Also, science is a human enterprise, so the full range of human characteristics (e.g., commitment, integrity, and creativity and also deception, fraud, and falsely claiming credit) is present. Methodology is the best we have now in the way of developing the knowledge base. Methods are constantly evolving to improve what we know and how we know it and to correct sources of bias or influences that can interfere with obtaining knowledge.

KEY TENETS AND STRATEGIES

A few overarching tenets or principles guide science and the methods we use to obtain knowledge. These are useful at a general level to understand science. In addition, they are extremely useful at a very concrete and specific level, namely, in interpreting the results of a given study and in communicating the results of a study to one's colleagues. In this section, I describe translating ideas into testable hypotheses, parsimony, plausible rival hypotheses, replication, and caution and precision of thinking as core elements.

First, ideas for scientific research must be translated into testable hypotheses. Scientific research depends on putting ideas to a test, which means making predictions, using systematic measures, and evaluating whether the data do or do not support a hypothesis. The concept of falsifiability has been used as part of the notion of testing ideas. The idea must be one that can be put to a test and in principle shown to be false.

In everyday life, one can see that this requirement is rarely invoked (or needed). For example, we might say a person is passive–aggressive. Usually that means we are interpreting their behavior as being nasty even though it does not appear that way. In everyday life, the concept conveys a point, and we usually do not challenge the person making it. To translate the concept into a scientific hypothesis (not too difficult to do), we would need a measure (systematic, objective, reliable, valid) of passive-aggressive style or behavior and to specify what evidence from our study would support or refute the view that passive–aggressiveness could explain the behaviors of interest. Perhaps the person is busy, slow, not wildly competent, or forgetful. We need a concrete way to test (and possibly support or refute) that passive–aggressiveness, rather than these other constructs, explains the behavior.

Occasionally, within psychology, theories are advanced that include components that are not easily testable. For example, psychoanalysis, developed by Sigmund Freud (1856–1939), was a comprehensive theory proposing that psychological processes and specific early childhood experiences play critical roles in our everyday behavior (e.g., comments we make to others, relationships), psychological defenses to ward off threatening material (e.g., denial, projection), and psychiatric disorders or social, emotional, and behavioral problems (e.g., anxiety, depression, aggression) and in how to change maladaptive functioning in the context of psychotherapy. Among the criticisms is that the theory is not very testable. Many key components could not be shown to be false because the theory seemed to be slippery enough to explain results no matter what happened. A key facet of science is devising testable hypotheses. Within the tradition of contemporary science, the concept of falsifiability, that is, being able to show that an idea or hypothesis is in fact not true, is critical. In short, a key component of science and the methodology on which it draws is to translate ideas into testable hypotheses and hypotheses that can be demonstrated to be false or that can be supported.

Second, scientific knowledge is based on parsimony. *Parsimony* refers to the practice of providing the simplest version or account of the data among the alternatives that are available. This does not in any way mean that the explanations themselves are simple. Rather, it refers to the practice of not adding all sorts of complex constructs, views, relationships among variables, and explanations if an equally plausible account can be provided that is simpler. We add complexity to our explanations as needed. If two or more competing views explain why individuals behave in a particular way, we adopt the simpler of the two until the more complex one is shown to be superior in some way.

A well-known illustration of competing interpretations comes from cosmology and pertains to the orbiting of planets in our solar system. Nicolas Copernicus, a Polish scientist and astronomer (1473–1543), advocated the view that the planets orbited around the sun (heliocentric view) rather than around the earth (geocentric view). This latter geocentric view had been advanced by Claudius Ptolemy (ca. 85–165), a Greek astronomer and mathematician. Ptolemy's view had dominated for hundreds of years. The superiority of Copernicus's view was not determined by public opinion surveys or the fact that Ptolemy was no longer alive to defend his position. Rather, his account could better explain the orbits of the planets as well as other phenomena and do so more simply, that is, parsimoniously. The heliocentric view could explain more with one key construct.

Parsimony relates to methodology in concrete ways. When an investigation is completed, we ask how to explain the findings or lack of findings. The investigator may have all sorts of explanations for why the results came out the way they did. Methodology has a whole set of explanations that may be as or more parsimonious than the one the investigator promotes. For example, say I develop a new psychotherapy (Kazdin's mindlessness therapy) and show that it is better than no treatment. I now explain how engaging in mindless behaviors (e.g., wandering the streets, grocery shopping, counting backward from 100 in Sanskrit) leads to reduced depression. I might be right. Yet, my view is not parsimonious and ought not to be adopted yet. There is a large literature showing that doing anything (e.g., meeting with a therapist, attending sessions) and expecting improvement in treatment leads to therapeutic change. These latter influences are referred to as *common factors of therapy* because they are present in many techniques. Common factors are more parsimonious than my mindless interpretation because common factors already explain the findings from hundreds of treatment studies. We do not need another set of constructs to explain the findings from my study. Additional research is needed to show that we need my explanation, but on the basis of my study and its design (just a no-treatment control group), there is no need for that explanation now.

Third, plausible rival hypothesis is another key concept of science (Cook & Campbell, 1979). A *plausible rival hypothesis* refers to an interpretation of the results of an investigation that is based on some other influence than the one the investigator has studied or wants to discuss. The question to ask at the completion of a study is whether other plausible interpretations exist to explain the findings. This sounds so much like parsimony that the distinction is worth making explicit. *Parsimony* refers to adopting the simpler of two or more explanations that account equally well for the data. The concept is quite useful in reducing the number and complexity of concepts that are added to explain a particular finding. Plausible rival hypothesis as a concept is related to parsimony but has a slightly different thrust: At the end of the investigation, can other plausible interpretations be made of the finding than that advanced by the investigator? Simplicity of the interpretation (parsimony) may or may not be relevant. At the end of the study, there could be two or 10 equally complex interpretations of the results, so parsimony is not the issue.

For example, let us say we have an amazing hypothesis that people who take cough drops during a cold get better faster than those who do not. We do a massive survey of students and identify two groups—those who take cough drops when ill and those who do not. We keep track of the participants and closely monitor all those who get colds in the next few months. Then we call them every day and find out when their cold is over. Lo and behold, we find that those individuals who take cough drops get better in one half of the time it takes those who did not take cough drops. We are all ready to conclude that, as predicted, taking cough drops is helpful in limiting the duration of colds.

Plausible rival hypotheses come in here by raising the question "Are there other plausible explanations for the results?" The answer is a resounding yes. It may be that people in the cough-drop group are just healthier in general. Perhaps those who take cough drops tend to take better care of themselves in general (better eating habits, more exercise), take care of themselves during a cold more (a little more bed rest and chicken soup), or have a history that indicates they are healthier to begin with. Perhaps they would have colds of shorter duration for these reasons, and cough drops are not needed at all. We could provide many more explanations for these findings. They do not have to be more or less parsimonious than the cough drop interpretation. They are all plausible, and one cannot decide from the study whether cough drops make any difference.

Methodology is all about plausible rival hypotheses and designing studies so that some hypotheses or interpretations are made more plausible than others. We engage in various methodological practices (e.g., random assignment, use of comparison and control groups, keeping experimenters naïve with regard to the hypotheses) to make some interpretations implausible so that the interpretation we are evaluating can be better evaluated. In the example, we might control for some of the interpretations I added to give a better test of the cough drop hypothesis.

Fourth, as mentioned previously, replication is central to science. *Replication* refers to repetition or repeatability and is important for two reasons. To begin, the procedures used in research must be repeatable. If one were to ask another scientist, "How did you get that finding?" the unacceptable and inappropriate answers (drawn from childhood) are "I'm not telling!" or "That's for me to know and you to find out." Science operates so that what investigators do in a study, how they do it, and all of the circumstances are described. Others must be able to repeat the study. One might refer to this as *replicability of the procedures*. The second way in which replication is central regards whether the findings can be reproduced. A question about the results of any study is whether the findings can be repeated or obtained again by someone using identical or similar procedures.

Replication relates to parsimony and plausible rival hypotheses. Were the results evident in this one study because of a chance effect or some odd circumstance in the situation about which the investigator may be unaware? Could there be a simple (parsimonious) explanation or one that is equally plausible (rival hypothesis)? Replication is needed to establish the credibility and genuineness of the finding. It would be unthinkable for a researcher to say he or she demonstrated a cure for a type of cancer but that other scientists could not replicate the results. One study might excite the news media and the public who reads the story. Yet, the scientific community might be very skeptical until this finding is replicated by other investigators. Only through repeated demonstration does one gain confidence in a finding. The initial investigation that obtained the result is tantalizing, promising, and maybe true, but more is needed. Skepticism of a finding is not suspicion directed toward the original investigator; rather, it comes from the realization that many influences might explain the results. For example, a finding might emerge because of chance, a statistical artifact that is embedded in the procedures. That is, sometimes the data will show a statistically significant effect even if there is no real effect in the world.

Finally, science encourages caution and precision in thinking. This does not mean scientists (humans) are invariably cautious and precise. Rather, the scientific community to which we belong demands of us as investigators that we are careful in not going beyond the data, or at least not too far. This means we can describe only what the study demonstrated, and any other part remains to be determined by further research. For example, in research we distinguish key concepts such as correlation, risk factor, and cause. A correlation indicates that two variables are related to each other at a particular point in time. A risk factor indicates that two variables are correlated with each other but that one clearly comes before the other. An example of a risk factor is cigarette smoking, which is a risk factor for later lung and heart disease. That is, those who smoke are more likely to contract later disease. We know the ordering of these. A *cause* means that one variable leads to, produces, and is responsible for a particular outcome. For example, nonhuman animal research has shown that various experiences early in development (e.g., physical contact with a parent, ingestion of toxins such as lead) influence brain development. The studies can show a causal relation because presenting or withholding experiences dictates the outcome. When a study shows only a correlation, the scientific community and the investigators themselves are cautioned not to go beyond the data and assume or state something more.

The caution and precision of scientific statements and inferences stand in contrast to inferences drawn in everyday life. The public at large and the media freely use such terms as *because, due to*, and *cause* in connecting concepts; for example, are children aggressive because they were abused, will I appear younger due to this cosmetic cream, and will that special new diet really cause weight loss? The questions are simplistic because they suggest that one variable leads to a particular outcome and that the relation is causal. Information in daily life moves seamlessly from casual to causal inferences. Science encourages greater care.

For example, we know that having a cigarette smoker in the home increases the likelihood that an infant will die from sudden infant death syndrome (SIDS Network, 2015; *sudden infant death* refers to the death of an infant, usually between the ages of 2 weeks and 2 years, that cannot be traced to disease, physical abuse, or other disorders. Many more children die of SIDS in a year than die of cancer, heart disease, pneumonia, child abuse, AIDS, cystic fibrosis, and muscular dystrophy combined). Hearing the relationship I mentioned (smoking as a risk factor) in casual thinking almost naturally leads one to assume smoking is the culprit (cause). Moreover, it is even more tempting to move to interventions—we ought to stop smoking to decrease SIDS. Scientific thinking is a bit more cautious. Smoking as a cause has not been established. Could it be one of many causes? Could it perhaps not be a cause at all but be correlated with something that is more likely to play a causal role? These questions and of course the very process of questioning is science.

The tenets and practices I have highlighted provide the underpinnings of methodology and are points to which we as scientists return in our thinking about a particular study and its findings and conclusions. Healthy skepticism about a given finding or set of findings is also part of methodological thinking. This skepticism is reflected in thoughts about a study and includes what might be called "yes–but" questions. Scientists read a study or learn of a finding and say, "Yes, but . . ."

- are there simpler interpretations to explain the findings than what the author has concluded (parsimony)?
- is there is a more reasonable interpretation than what the author concluded (plausible rival hypothesis)?
- does the relation (correlation, risk factor) really explain the connection of these variables?

- has that finding been replicated?
- do the design, assessments, and data analyses allow for the conclusions to be reached, whatever their interpretation (caution and precision in thinking)?

Skepticism does not mean rejection of findings or criticism for its own sake. We build the knowledge base, often slowly, by clarifying, elaborating, repeating, and better establishing what we think we know. Skepticism in asking the "yes–but" questions operates at its best when investigators adopt this stance about a finding rather than their peers.

DRAWING VALID INFERENCES

The overall purpose of methodology is to permit one to draw sound or valid inferences, that is, to reach conclusions that are as free as possible from competing interpretations (plausible rival hypotheses). We engage in the practices and procedures of methodology not for their own sake but to help with interpretation and to bring clarity or relative clarity to our findings. In this process, methodology has two major roles.

Codifying Sources of Problems

Methodology codifies the sources of problems that emerge in drawing inferences. Essentially, methodology provides a list of most of the problems to be wary of, what the investigator ought to think about before running the first participant, and what might go wrong during a study that would interfere with drawing clear conclusions. In everyday life, we develop fears and worries based on experience; some biological predisposition; information from relatives, peers, and the media; and no doubt many other sources. We have some friends who worry about exams, elevators, catching a disease, being struck by lightning, and so on—it seems as though it would be so much better if there were a master list of worries somewhere from which we could select! Well, in the context of science, methodology provides a master list of sources of worry or concern. *What to worry about* is not a very sophisticated term. Understandably, other terms used in methodology include *sources of artifact and bias* and *threats to experimental validity* (e.g., Kazdin, 2016; Shadish, Cook, & Campbell, 2002). Table 1.1 lists many of the main sources of bias and artifacts that can interfere with drawing valid inferences. There are others referred to in earlier citations, but this set is a great place to begin.

Some of the problems in research are so pervasive that one can almost always point them out. For example, we have learned from years of reviews from the 1960s to the present that most studies in psychology do not provide very good tests of their hypotheses (e.g., Bakker, van Dijk, & Wicherts, 2012; Cohen, 1962). The key reason is that they have low statistical power—too low to find a difference even if there really is a difference. Power depends on many factors in a study, but sample size is the first place to look. Studies typically use too few participants to find real differences.

But if this problem is not new, why is it still with us? That is a separate topic, but perhaps the best answer is that psychology as a discipline does not enforce or require well-powered studies. Also, we have learned that repeatedly pointing out the problem and pleading with researchers to increase power does not have very much impact (see Sedlmeier & Gigerenzer, 1989). In general, it is wise to know on the basis of such issues as statistical power whether one could obtain the desired effect well in advance of running the first participant.

I mentioned that methodology codifies many of the problems that can emerge in conducting and interpreting research, and low power as a result of small sample size is just one of them. (Later chapters will discuss power and other methodological problems as well.) The overall point, for our purposes in this introductory chapter, is that methodology provides guidelines—fairly clear guidelines, actually—about where to look for potential problems that could undermine a study. Most of these problems can be identified and addressed at the design stage before a study is begun.

Codifying Solutions

Methodology also codifies many of the solutions to the problems and practices that can help scientists draw valid inferences. Here, too, it would be useful to have a list of solutions to the problems we worry about in everyday life and outside of the context of methodology. How great would it be to, on the back

TABLE 1.1

Examples of the Problems in Drawing Valid Inferences and Some Practices to Address Them

Name of the problem	Definition	Possible solutions
History and maturation	Changes that occur over time as a result of events (history) or processes within the individual (maturation). The results of the study may perhaps be due to such changes rather than to the manipulation, intervention, or independent variable.	Include a control group that does not receive the experimental manipulation or intervention. If the groups are composed by random assignment, the historical and maturational influences will be controlled. It is likely that such influences will apply to both (all) groups, and group differences are likely to be due to the experimental manipulation.
Testing	Taking any test on more than one occasion often leads to changes in performance (e.g., improvement). Studies that assess participants on multiple occasions (e.g., pretest and posttest) might show change just because of repeated testing.	A control group that receives the same assessments but does not receive the intervention or experimental manipulation makes repeated testing implausible as an explanation for any group differences.
Selection biases	The groups (e.g., experimental, control) are different to begin with because of how they were selected or formed. Any differences between groups at the end of the study may be due to these differences rather than to anything the investigator does.	Assign participants randomly to conditions. With reasonably large samples, this random assignment is likely to produce groups that are equivalent before the manipulation is provided. Alternatively, participants can be matched on variables (e.g., level of anxiety) and randomly assigned to groups in matched sets so groups are equivalent on key variables.
Attrition (dropping out)	Loss of participants over the course of the study can make groups different. The random composition of the groups has changed because participants selectively excluded themselves. The groups may be different from each other, leading to a selection bias, as noted previously.	Try to minimize the loss of participants. Also, evaluate characteristics of participants who did drop out versus those who did not. Complete statistical tests designed to control for the impact of loss of participants (e.g., intention-to-treat analyses or other methods of imputation [data estimation]).
Cues of the experimental situation	Incidental cues of the experiment (what participants believe, what they are told, the expectations unwittingly conveyed about how they ought to perform) may explain the group differences rather than the experimental manipulation. Cues may foster a way of responding that account for the results.	Use a control group that receives all but the special part of the intervention so that almost all cues are identical across groups. Interview participants in the study or in pilot work to ask how they are likely to behave to determine whether cues would lead to a systematic way of responding. Use measures that are not so transparent that participants can readily discern what is being measured.
Sample characteristics	The findings may be restricted to the special sample that was used; the finding is genuine but may not apply to others (e.g., of different ages, sex, ethnicity, culture).	Include different types of participants (e.g., males, females; more than one ethnic group). College student samples, used in most psychological research, have special characteristics that may make them unique. Think carefully about the sample that provides the best test of the hypotheses. Analyze the results in a way that permits one to see whether a characteristic of the sample in fact relates to the finding; replicate results with other samples.
Sample size (low power)	No differences were obtained in the study because the power (ability to detect a true difference when there is one) was too low.	Use a larger sample. Use a within-subjects design (pre and post measures on the same participants). Also, make directional statistical predictions and use directional (one-tailed) statistical tests. Use statistics (such as effect size) other than those that focus on statistical significance.

TABLE 1.1 (*cont.*)

Name of the problem	Definition	Possible solutions
Questionable reliability and validity of the measure	No statistically significant differences were obtained because the measure has considerable error variability (unreliability) or it is not clear that this measure is a very good measure of the construct of interest. If a researcher invented a measure and the results did not come out as expected, it could easily be the case that the items (face validity) do not measure what you believe or do not measure it very well (face validity).	Use measures that have validity and reliability data pertinent to the focus of the investigation. Use multiple measures and combine them statistically (e.g., by factor analysis) for a better index of the construct of interest. Use multiple measures that rely on different methods (e.g., self-report, direct observation). Conduct analyses on the measures to assess directly within the study whether measurement issues might explain the findings (e.g., internal consistency, test–retest reliability). If you invent a measure, include some data to suggest that it is valid as an index of the concept you wanted to measure.

Note. The purpose of this table is to illustrate how methodology codifies problems to which researchers must be sensitive in their demonstrations and solutions to rule out or address these problems. The list of problems or solutions presented here is not exhaustive (see Cook & Campbell, 1979; Kazdin, 2016; Shadish, Cook, & Campbell, 2002).

of the sheet that says worry about lightning, Ebola virus, mad cow disease, and germs from friends, unprotected sex, and unprotected celibacy, have a long list of solutions? In the context of research, methodology provides a long list of solutions and strategies. In Table 1.1, I have listed in the third column some of the solutions that methodology encourages. These are only a sample of some of the major solutions to some of the major problems. One can see that we engage in research methodological practices (solutions) to address specific problems.

If methodology were merely a list of problems and solutions, however, then one might not need a course or a book. One could just master the lists and go on to other topics. The challenge is that designing a study involves many decision points, and each of these points has some implications for drawing conclusions. Consider that I want to study new treatment for depression. Okay, who will be the participants, and how will I select them? The answer to this question can determine whether treatment works and, if it does work, whether this effect can be shown statistically. For example, if I recruit individuals of any age who are depressed (e.g., 15–60 years) and individuals who are depressed for whatever reason (e.g., clinical depression, bereavement), I may be less likely to find an effect because of the variability or individual differences in participants and types of depression. Also, if I include only college students who seem depressed on a self-report measure, this may limit the extent to which the results might apply to patients with clinical depression.

Also, methodological challenges and solutions are evolving. Ethical issues, for example, constantly require new considerations and guidelines. Will using social media as part of my study create new risks for violating confidentiality? Is the database protected (encrypted) so that no one can invade the computer storage and uncover confidential information and the identity of the participants associated with that information? When my study is completed, I may make the raw data and analyses freely available to other scientists (as many journals now require). Are there any risks with that, for example, might the data be used in a way that unwittingly portrays a particular population (e.g., ethnic group, workers in a particular industry) or geographical area (e.g., region of United States, country) in a negative light? Research methods evolve (e.g., combining large data sets on the basis of how individuals use the Web or their electronic health records) and, as they evolve, how individuals are protected also needs to change.

Research situations require consideration of what problems will emerge and what solutions are possible and the trade-offs of one solution versus another.

As in many other aspects of life, sometimes choosing one solution limits other courses of action. That is, maximizing one option may lead to sacrifice of some other benefit. Methodology requires a deeper understanding of practices so that these trade-offs can be thoughtfully considered in relation to the specific hypothesis the investigator wants to test.

General Comments

How methodology codifies various problems and solutions that relate to drawing valid (well-based) inferences might sound too abstract and cerebral to be of broad relevance beyond the confines of scientists conducting studies. Actually, issues can affect decisions in everyday life and indeed even decisions about life and death. For example, we are all keenly interested in the development of treatments for life-threatening illnesses. How could methodology figure into this? A review of medical research of a variety of diseases and conditions revealed that more than 25% of the studies surveyed (published from 1975 to 1990) revealed no differences among the treatments that were studied (Moher, Dulberg, & Wells, 1994). In the majority of these studies, statistical power was weak. Of course, I cannot speak for you, but I personally am not pleased to learn that viable treatments might be available but one cannot tell because the studies were not well designed. That is, the sample size was too small to demonstrate a real effect if there was one. Let us try to ignore for the moment the tax dollars likely to be wasted (from federally and state-funded research), the enormous inconvenience and perhaps pain (physical and psychological) of many of the participants run in such trials, and the ethical issues all of this raises by exposing participants to any condition when the research might not be able to obtain an answer. We want all investigators who design studies, who review proposals for research, or who are involved in the research process in some way to understand methodology to minimize delays in accumulating knowledge that can affect people in everyday life.

In many cases, the life-and-death features come from the topics under study. In psychology, we want to discover the etiologies and means of prevention of suicide, psychiatric disorders, addictions, dementia, and aging, to mention huge areas of research. Each area has a need for basic research on processes, mechanisms of action, and so on and eventually, one hopes, intervention trials (treatment, prevention). Progress is slow even when all the methodological stars are aligned (i.e., strong empirical tests based on great methods). Among the challenges that make the progress slow is replication of effects and then, if relevant to the public at large (e.g., clinical practice, child rearing, day care), dissemination of the results. This is not a quick process under the best of circumstances. Methodologically weak studies along the way merely delay or even mislead us as we move forward.

METHODOLOGY IN CLINICAL PSYCHOLOGY

Although methodology is central to all science in general, there are special reasons to focus on issues related to research in clinical psychology. Clinical psychology embraces all of the usual features of scientific research, such as defining the research idea, generating hypotheses, designing investigations, collecting and analyzing data, and so on. Yet, in clinical psychology research is conducted in laboratory and clinical settings and addresses theoretical and applied issues. Other areas of psychology such as counseling, educational, and school psychology and other disciplines such as psychiatry, nursing, and social work also engage in research that spans quite diverse settings, participants, and goals. Research in these areas often presents novel challenges to the investigator. Consider the diversity of topics, samples, and settings in which clinical psychological research is conducted.

The scope of research in clinical psychology is enormous. Among the topics addressed are the assessment, diagnosis, course, treatment, and prevention of social, emotional, and behavioral problems; affective and cognitive processes, personality, family processes, peer relationships, the interface of mental and physical health, and cross-cultural differences. The populations studied include children, adolescents, adults, elderly people, and people with special experiences (e.g., homelessness, divorce, prisoners of war), medical impairment and disease (e.g., cancer, AIDS, spinal cord injury, diabetes), or psychological disorder or dysfunction (e.g., depression, anxiety, posttraumatic stress disorder, autism). People in contact with special populations, that is, those who are exposed to someone with a special condition, are themselves often studied (e.g., children of parents who abuse alcohol, spouses of depressed patients, siblings of children with physical disabilities). Yet, the populations also include nonhuman animal subjects in an effort to test basic theory, processes, or experiences in the lab and to evaluate animal models of some state or clinical dysfunction. Research in clinical psychology is conducted in diverse settings (e.g., laboratory, clinics, hospitals, prisons, schools, industry) and in the absence of structured settings (e.g., runaway children, homeless families). Finally, research in clinical psychology is also conducted in conjunction with many other areas of research and different disciplines (e.g., criminology, health psychology, neurology, pediatrics, psychiatry, public health).

Understandably, diverse methods of study are required to meet the varied conditions in which clinical psychologists work and the special challenges in drawing valid scientific inferences from these situations. The methodological diversity of clinical research, as with the substantive diversity, can be illustrated in many different ways. Studies vary in the extent to which the investigator can exert control over the assignment of participants to conditions or administration of the intervention and the selection of preexisting groups (with a type of condition or experience) and how they are followed and evaluated.

Occasionally, clinical psychologists conduct research with college students recruited from introductory psychology classes at a university. Participants are seen for a session or two and complete a laboratory task. If, however, one looks at the premier journals in clinical psychology and allied disciplines (e.g., *Journal of Abnormal Psychology*, *Journal of Consulting and Clinical Psychology*, *Clinical Psychological Science*), it becomes clear that this is not the usual paradigm for research. And yet, many basic questions about phenomena of interest (e.g., emotion regulation, psychopathy, and brain activation in relation to varied social cues that require carefully controlled conditions in the lab) might well be studied with nonclinic samples and answer questions that are theoretically important or lead to understanding clinical phenomena and populations.

Often, clinical samples are studied and evaluated over a period of time, even if only a month or two. The challenges of recruiting participants; retaining them if the study lasts weeks, months, or years; ensuring that their care is fine if they are in a clinical sample; and obtaining enough participants to test the hypotheses are a few of the salient challenges. These challenges can have implications for the conclusions and whether the conclusions can be generalized to other investigators, samples, and settings. The challenges also mean that quite different methodological approaches are often used, including diverse designs (e.g., group and single case) and methods of data evaluation (e.g., statistical and nonstatistical data evaluation; Kazdin, 2016).

My purpose in highlighting the diversity of clinical psychology is to underscore the importance of facility with the methods of research. Special demands or constraints are frequently placed on the clinical researcher. Ideal methodological practices (e.g., random assignment) are not always available, but they are not always necessary. Also, restrictions may limit the researcher's options (e.g., a control group might not be feasible, only small sample sizes may be available). If one is conducting research that requires more than one occasion (e.g., repeated assessments on two or more occasions, intervention studies, longitudinal studies on the course of clinical dysfunction), retaining participants in the project and loss of participants (attrition) present new challenges (e.g., for data analyses).

The task of the scientist is to draw valid inferences from the situation and to use methodology, design, and statistics toward that end. In clinical psychology and related areas of research, the options in methodology, design, and statistics must be greater than in more basic research areas to permit the investigator to select and identify creative solutions. Clinical research is not in any way soft science; indeed, the processes involved in clinical research reflect science at its best precisely because of the thinking and methodological ingenuity required to force nature to reveal its secrets.

Deploying strategies to accomplish this requires an appreciation of the purposes of research and the underpinnings of research strategies in addition to the concrete practices on which methodology draws.

GOALS AND FOCUS OF THIS BOOK

The goal of this book is to help the reader design, conduct, recognize, and appreciate high-quality research. High-quality research begins with the idea, theory, or prediction that underlies the study and its contribution to knowledge. The quality of the research continues to be defined by how that idea is translated into an investigation. Addressing potential problems and solutions, as highlighted earlier, contributes to how well that idea has been translated. Can the investigator draw the conclusions he or she wants on the basis of how the study was planned, executed, and evaluated? High-quality research occurs when scientists are skilled in each of these phases of research. This book addresses the full process of research, moving from ideas, design, assessment, data evaluation, interpretation, and presentation to communication of findings. Pervasive in all facets of research are ethical issues and responsibilities to the individual participants but also to the scientific enterprise more generally. The chapters in this book address multiple facets and components of research and the flow from planning, to executing, and to evaluating a study.

Part I of this book, *Introduction: Overview and Background*, conveys what methodology is and the roles it plays in scientific knowledge. Perhaps the most critical point is to conceive of methodology not only as a set of practices but as a way of approaching the subject matter of interest. The goal of research is to draw valid inferences, and methodology provides the means to accomplish this.

Part II, *Beginning the Research Process*, focuses on sources of ideas for research and different areas in which the process of developing a study might begin. Key concepts that often guide research (e.g., correlates, risk factors, mediators) and various ways to study a phenomenon (e.g., basic laboratory to applied research) are discussed. The material is designed to help the researcher identify the problem or topic of interest and how it might be studied.

Part III, *Sampling and Assigning Participants to Conditions*, includes chapters that address procedures, practices, and decisions as one moves from research ideas to concrete details of the study. Sampling and whom to include as participants can make a huge difference, and current issues (related to the use of college students as participants and online participant pools) are covered. Also, the importance of and limitations of random assignment, a core feature of experiments, are addressed. Finally, different design options are detailed, from experiments in which conditions are manipulated to observational studies in which groups are selected because of some particularly important characteristic.

Part IV, *Assessment*, presents scale evaluation and development. Although most research work does not involve development of scales, selecting measures for research is a critical step. Articles on scale development and measurement reliability and validity convey key issues that influence or ought to influence the measures one selects and the interpretation of research results. Issues raised in the assessment of underrepresented groups are also included in this part. Ethnic, minority, and cultural issues are of great interest in their own right because science and psychological science are designed to understand everyone and how differences and similarities emerge. In addition, the study of diverse groups raises important methodological issues related to sampling, measurement validation, and generality of results.

Part V, *Data Analysis, Evaluation, and Presentation*, includes several articles that relate data evaluation to other facets of research. Major attention is given to statistical significance testing, including its origins, strengths, and limitations. There is deep concern regarding the uses and misuses of statistical significance and null-hypothesis testing. The chapters in this part encompass critical concepts and statistics, including statistical power, effect size, and alternatives to null-hypothesis testing. Data presentation is also covered in a chapter on novel and standard ways of graphing one's data.

Part VI, *Special Topics: Evaluation in Clinical Practice and Research*, presents assessment in ways that unify clinical practice and research. The importance, contribution, and examples of evaluation in clinical work and qualitative measures to enhance the understanding of therapeutic change from the clients' perspective are discussed. Also, assessment in treatment outcome research extends the discussion with similar concerns, namely how to evaluate therapeutic change in ways that are meaningful to the client. Ethnic and cultural variation can greatly influence evaluation and interpretation of results. Methodological challenges in designing and evaluating interventions with ethnically and culturally diverse groups are covered in this part.

Part VII, *Multiple Methodologies*, presents various methodological approaches to research. The dominant methodological approach in research is null-hypothesis statistical testing. Yet, other approaches are equally as rigorous and scientific but receive much less attention. Three methodological approaches are highlighted in this part, including math and statistical modeling (a very user-friendly chapter), qualitative research, and single-case experimental designs. As the book notes in a few places, how one studies a phenomenon can greatly influence as well as restrict what one learns. Diverse methodologies can elaborate that same phenomenon in novel ways not available through another approach.

Part VIII, *Ethical Issues and Scientific Integrity*, focuses on investigators' responsibilities to participants, science, and the public at large. This part begins with the ethical principles and code of conduct developed by the American Psychological Association as well as a brief guide from the National Institutes of Health. Research ethics for ethnic and minority samples can raise special issues that are addressed in their own chapter. Scientific integrity issues are raised and include questionable research practices and decision making that can bias how research findings are presented, fraud and fabrication of data, and allocation of credit of authorship. Ethics and scientific integrity reflect a range of responsibilities that go beyond the sets of issues illustrated in the chapters in this section, which is why the overarching ethical principles and code of conduct are so important to have as a reference and broader guide.

Part IX, *Reproducibility of Findings: Replication and Improving Research Practices*, focuses on renewed interest in improving our scientific research by increasing replications, that is, verification of findings by seeing whether they can be reproduced in additional studies. The renewed interest stems from many influences, including concerns over the bias of journal publication, by which statistically significant findings are favored, questionable research practices that may further bias and foster more chance findings in the research literature, and fraud, among other influences. Replication is not merely just doing the study over again. Several relevant methodological issues are discussed, along with broader issues about how to foster replication and what changes might be needed in incentives for researchers—in journal publication and the review of manuscripts, for example—to increase the likelihood of replication attempts.

Part X, *Publication and Communication of Research*, addresses the preparation of manuscripts designed to communicate research. Communication is a logical conclusion to completion of research. There are reporting standards and guidelines for preparing journal articles that involve empirical studies or literature reviews, and these are covered in the chapters. Concrete guidelines are provided on preparing articles for publication that underscore how to convey the rationale for the study as well as many decisions along the way as part of the design, assessment, and data evaluation procedures. The chapters in this part are designed to convey the thought processes prompted by methodology that deserve attention in preparing reports of one's own research, reviewing the literature in an area, conducting meta-analyses, and preparing grant applications.

Finally, in Part XI, *Perspectives on Methodology*, I provide closing comments to convey some of the broader lessons methodology teaches. Among the recommendations is to encourage the use of novel and diverse methods of investigation. The purpose is not diversity and novelty for their own sake. Rather, methods of studying a phenomenon often influence what can be learned and what specifically is learned. Complementary methods of study can elaborate the phenomenon in new ways. Researchers are encouraged to study phenomena in diverse ways and to develop collaborations in their careers that facilitate this. Novel findings can come from novel methods.

Such findings also come from novel ideas, but even novel ideas studied with the usual participants, usual measures, and usual data analyses can be constrained.

The chapters in each section raise points central to what might be considered phases of a study or some temporal sequence, beginning with the research idea, of course, and then moving to details about how the study will test the idea in concrete ways, who the participants will be, what measures will be used, and so on. Of course, there is an obvious sequence of steps. At the same time, many chapters connect the research process by spanning different phases of research and showing their interconnections. For example, ethical issues and data evaluation and analyses emerge at the proposal stage before the first participant is even run. So not all aspects of methodology follow a stepwise progression. Even the study itself might be conceptualized as not having a fixed starting and ending point. For example, the write-up of a study is not the end of a sequence of tasks in research. The well-described and well-presented write-up ought to point rather clearly to the next studies and hence constitutes a new beginning. I have divided the book into parts, but the research process is continuous, and considerations (assessment, design, data analyses) emerge at multiple places (before the study is begun, when the write-up is underway).

We begin now with the often most daunting research challenge, namely, developing the idea that is to serve as the basis for a study. There are many guidelines and aids, and these are provided in Part II.

References

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*, 543–554. http://dx.doi.org/10.1177/1745691612459060

Baumann, J., & DeSteno, D. (2010). Emotion guided threat detection: Expecting guns where there are none. *Journal of Personality and Social Psychology*, *99*, 595–610. http://dx.doi.org/10.1037/a0020665

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195154054.001.0001

Brusatte, S. L., Butler, R. J., Barrett, P. M., Carrano, M. T., Evans, D. C., Lloyd, G. T., . . . Williamson, T. E. (2014). The extinction of the dinosaurs. *Biological Reviews*, *90*, 628–642. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/brv.12128/full

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, *65*, 145–153. http://dx.doi.org/10.1037/h0045186

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 6889–6892. http://dx.doi.org/10.1073/pnas.1018033108

Dunning, D., & Balcetis, E. (2013). Wishful seeing how preferences shape visual perception. *Current Directions in Psychological Science*, *22*, 33–37. http://dx.doi.org/10.1177/0963721412463693

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*, 383–386. http://dx.doi.org/10.1111/j.1467-9280.2006.01716.x

Haselton, M. G., & Gildersleeve, K. (2011). Can men detect ovulation? *Current Directions in Psychological Science*, *20*, 87–92. http://dx.doi.org/10.1177/0963721411402668

Johnson, M. K. (2006). Memory and reality. *American Psychologist*, *61*, 760–771. http://dx.doi.org/10.1037/0003-066X.61.8.760

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus & Giroux.

Kazdin, A. E. (2016). *Research design in clinical psychology* (5th ed.). Needham Heights, MA: Allyn & Bacon.

Kessler, R. C., & Wang, P. S. (2008). The descriptive epidemiology of commonly occurring mental disorders in the United States. *Annual Review of Public Health*, *29*, 115–129. http://dx.doi.org/10.1146/annurev.publhealth.29.020907.090847

Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). *World report on violence and health*. Geneva, Switzerland: World Health Organization.

Maartens, R. (2013). The accelerating universe. *Quest*, *9*, 14–16.

Moher, D., Dulberg, C. S., & Wells, G. A. (1994). Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*, *272*, 122–124. http://dx.doi.org/10.1001/jama.1994.03520020048013

Pohl, R. F. (Ed.). (2012). *Cognitive illusions: A handbook on fallacies and biases in thinking, judgment and memory*. New York, NY: Psychology Press.

Roediger, H. L., III, & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science*, *9*, 123–127. http://dx.doi.org/10.1111/1467-8721.00075

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*, 309–316. http://dx.doi.org/10.1037/0033-2909.105.2.309

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

SIDS Network. (2015). *Sudden infant death syndrome and other infant death information web site*. Retrieved from http://sids-network.org

Wells, G. L., & Loftus, E. F. (2013). Eyewitness testimony memory for people and events. In I. B. Weiner (Ed.), *Handbook of research: Forensic psychology* (Vol. 11, pp. 617–629). New York, NY: Wiley.