

EVALUATING NUMERICAL INFORMATION

1.1 INTRODUCTION

In everyday use, **statistics** can refer to specific pieces of numerical information, such as average income for all employed persons in the United States. In science and technical fields, the term *statistics* more often describes techniques for analyzing and interpreting numerical information. Readers should not assume that all published numerical information is correct. **Numeracy** skills are needed to understand and evaluate how numerical information is collected, analyzed, and presented.

1.2 GUIDELINES FOR NUMERACY

A report published by the American Statistical Association's Committee on Guidelines for Assessment and Instruction in Statistics Education (GAISE College Report ASA Revision Committee, 2016) described numeracy skills as follows:

Students should become critical consumers of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted. To be a critical consumer of statistically-based results, it is necessary to understand the components that produced them: the *design* of the investigation, the *data*, its *analysis*, and its *interpretation*. Identifying the *variables* in a study, which includes consideration of the measurement units, is a necessary step to inform judgments or comparisons. Identifying the *subjects* (*cases*, *observational units*) of a study and the *population* to which the results of an analysis can be *generalized* helps the consumer to recognize whether the reported results can reasonably support the conclusions claimed for an analysis. Being able to interpret displays of data (tables, graphs, and visualizations) and statistical analyses also informs the consumer about the reasonableness of the claims being presented. (Italics added)

Italicized terms in the preceding quotation identify components of the research and data analysis process; these are discussed further in Chapter 2 and research methods courses. This chapter briefly considers other fundamental issues in the communication of numerical information: (a) sources (or communicators), (b) types of evidence, (c) questions about generalizability and causal inference, (d) quality control mechanisms, (e) ethical responsibilities, and (f) degrees of belief.

1.3 SOURCE CREDIBILITY

1.3.1 Self-Interest or Bias

Communicators can be motivated by self-interest or bias. Self-interest is often clear in mass media; messages are often intended to influence audience beliefs or behaviors (such as voting or product purchases). Science communicators can also be motivated by self-interest; for instance, some researchers receive funding from alcohol or pharmaceutical companies, and their future funding may depend on research results. Many science journals require authors to declare potential conflicts of interest.

Self-interest of information providers is not always obvious. Many webpages offer “sponsored content”: paid messages from advertisers that look like news articles but in fact promote the interests of advertisers. For instance, a new diet pill might be presented as “news” when in fact the article is an advertisement. *Communicator self-interest raises concerns about credibility of messages.*

1.3.2 Bias and “Cherry-Picking”

Communicators generally cannot (or do not) present all available information. Selection of information by communicators can be influenced by **confirmation bias**, a preference for information that confirms preexisting beliefs or ideas. Biased selection of evidence can be informally called **cherry-picking**. Information and ideas that are excluded may be as important as information that is included.

As an example of cherry-picking, suppose 20 studies show no association between consuming meat and cancer risk, and 3 studies do show an association. A journalist might report only the 3 studies that showed an association or might report only the single most recent study. Whether the bias was intentional or not, the article will not provide an accurate summary of research results.

When scientists write **literature reviews** (reviews of past research), they are expected to discuss all past relevant research.¹ Literature reviews are included in the introductions to most primary source research reports; literature reviews can also be stand-alone papers or books.

1.3.3 Primary, Secondary, and Third-Party Sources

An old game called “telephone” illustrates the problem of distance from a source. People form a line; the first person whispers a message to the second person, the second person whispers it to the third, and so forth. When the final message is compared with the original message, there are changes and distortions. Transmission of information can introduce errors because of each person’s biases or misunderstandings.

In science, a **primary source** is a research report written by a researcher who has firsthand knowledge of behaviors and events in a study. Primary source reports (sometimes called articles or papers) are published in **science journals**.² Primary source data may also appear in books written for science audiences.

A **secondary source** is a description or summary of past research, created by someone who did not experience the reported data collection or observations firsthand. In many disciplines, secondary sources are scholarly books. Some journal articles are also secondary sources because they only review past research and do not present new data about which their authors have firsthand knowledge. Literature reviews in the introductions to science journal articles are secondhand discussions of past studies. (In the sciences, *literature* refers to past published research.)

Unfortunately, primary source reports are usually long and difficult to read (particularly for readers unfamiliar with statistics and technical terms). Language in research reports is

sometimes unnecessarily obscure. Some full-length science research reports are published on the web as open-access materials; anyone can view these. However, many publishers require fees or subscriptions for access. The consequence is that many people can't easily understand most primary source information in science and sometimes cannot even gain access to it.

Much content on websites for news organizations is **third-party content**. This is content written by someone who may have examined only secondary sources or other thirdhand content, such as news reports or press releases. Often, third-party content is authored by someone who has no technical knowledge of the research field and statistical methods. Examples include articles published by news organizations. These articles usually don't provide complete or accurate information about research results.

In the past, editors of prestigious newspapers required reporters to fact-check claims carefully. Increasingly, news reports on the web are paraphrases of, or uncritical reposting of, third-party content from other news sources. Some mass media news sources specifically disclaim responsibility for accuracy. Here is an example; many other news organizations post similar disclaimers:

CNN is a distributor (and not a publisher or creator) of content supplied by third parties and users. . . . Neither CNN nor any third-party provider of information guarantees the accuracy, completeness, or usefulness of any content. . . . (CNN, 2018)

Communicators can provide better quality information when they are closer to original sources of information, and they are likely to provide better quality information when they assume responsibility for accuracy.

In everyday life, most of us rely on thirdhand information most of the time. Because so much of what we think we know is based on thirdhand information, we should not be overly confident about things we think we know.

1.3.4 Communicator Credentials and Skills

Communicators are more believable when they have training and background related to information in the message. Researchers generally have credentials that provide evidence of this training and background, including advanced degrees such as a PhD or MD, affiliations with respected organizations such as universities, and publications in high-quality science journals. Some journalists have strong credentials in science, but many do not. People who do not have training in statistics can easily misunderstand studies that use statistical terms such as *logistic regression* and *odds ratios*.

Celebrity status is not a meaningful credential. Famous media personalities, such as Dr. Oz³ and other self-appointed lifestyle or health experts, may base recommendations on incomplete or incorrect information.

Scientific research reports include source information (authors, university affiliations, and so forth). News reports and websites sometimes do not include source information; they provide no basis to evaluate self-interest, distance from information source, and credentials. Guidelines for evaluation of websites are provided by Kiely and Robertson (2016) and Montecino (1998).

1.3.5 Track Record for Truth-Telling

There are independent, nonpartisan organizations that evaluate communicator track records for truth-telling in journalism, for example, the Pulitzer Prize-winning site www.politifact.com. PolitiFact rates statements as true, mostly true, half true, mostly false, false, and “pants on fire” (extremely false). Other respected fact-checking sites are www.snopes.com

and www.factcheck.org. These fact-checkers do the work that information consumers usually don't have the time to do.

Information published in scientific journals can be incorrect because of fraud; fraud in science is rare, but it has occurred. A notorious example was a claim by Andrew Wakefield that vaccines cause autism (discussed by Godlee, Smith, & Marcovitch, 2011). There are severe penalties for fraud or plagiarism in science, including forced retraction of publications, withdrawal of research funds, loss of reputation, and job dismissal. Rare instances of fraud in science can be identified by a web search for the researcher name and terms such as *fraud*. *Information consumers should be skeptical of information from sources with poor records for truth-telling.*

1.4 MESSAGE CONTENT

1.4.1 Anecdotal Versus Numerical Information

Anecdote means “story,” often about an individual person or situation. First-person accounts are often called **testimonials**. Audiences may find narrative stories or anecdotes more persuasive and memorable than numerical information. There are many potential problems with **anecdotes (anecdotal evidence)**. Sometimes individual situations are not reported accurately (for example, advertisements for weight loss products often include falsified before and after photos). Even when anecdotal evidence is accurate, it is difficult to know whether the experience shown is generalizable: Has this experience happened to many other people, or was this a unique situation? Diet product advertisers are required to acknowledge this and typically do so in a tiny footnote: “Individual results may vary.”

In science, a detailed report of an individual person or situation is called a **case study**. The study of unique cases, such as the brain damage suffered by railway worker Phineas Gage (Kihlstrom, 2010; Twomey, 2010) can be valuable. However, generalizability concerns are still relevant.

Anecdotal evidence can dramatize genuine problems. However, anecdotal evidence can also dramatize and promote incorrect beliefs. It is obviously easy to cherry-pick anecdotes. Supporting evidence in the form of systematic numerical information can provide a more accurate overview of evidence than anecdotal reports.

1.4.2 Citation of Supporting Evidence

In science, identification of outside sources of evidence is done by **citation**. Author names and years of publication are included in the text (to identify sources of ideas and evidence), and complete information to locate each source is included in a reference list. Citation has two purposes. First, it gives credit to others for their ideas and evidence; this avoids **plagiarism**, which occurs if authors present ideas or contributions of other people as if they were the authors' own new contributions. Second, it shows how the present study builds upon an existing body of evidence.

A message is more believable when it includes or refers to specific supporting evidence. In science, the most complete and detailed supporting evidence appears in primary source research reports in science journals. Documentation of information sources is typically less detailed and systematic in journalism and mass media. (The best science journalists provide references or links to primary source research reports.)

It is possible for a writer or an advertiser to claim a spurious air of authority by citing numerous sources. However, a long list of references does not guarantee accuracy. On closer examination, readers may find that communicators have cherry-picked, misinterpreted, or

misrepresented evidence; cited sources that are not relevant to the topic; or referred only to opinion pieces that do not actually contain evidence.

To evaluate the quality of evidence, we need to know how it was collected. Collection of evidence in science is systematic; that is, there are rules and procedures that specify what researchers should do to gather evidence and limit the kinds of interpretations they are permitted to make. Rules for statistical analysis are an important part of this.

1.5 EVALUATING GENERALIZABILITY

Researchers and journalists usually want to generalize about their findings. In other words, instead of just saying: “45% of the *respondents I talked to* said they plan to vote for candidate X,” they want to say something like “45% of *all registered voters* plan to vote for candidate X.” **Generalizability of results** is the degree to which a researcher can claim that results obtained in a specific sample would be the same for a population of interest. Results from a sample can be generalized to an actual population of interest if the sample is representative of the population; representativeness can often be obtained using random or systematic methods to select the sample. Results from an accidental or a convenience sample may be generalizable to a hypothetical population if the sample resembles that hypothetical population. Results from a biased sample are not generalizable. In experiments, generalizability also depends on similarity of type and dosages of experimental treatment to real-world experiences with the treatment variable, setting, and other factors.

Polling organizations, such as Gallup, collect public opinion information in ways that provide a good basis for generalization. They use large samples (usually at least 1,000 individuals) and obtain these samples using combinations of random and systematic selection so that the people who responded to the survey resemble the larger population (such as all registered voters) in terms of age, income, and so forth (Gallup, n.d.).

When journalists report information from polls and demographic studies, they are (once again) in a position to cherry-pick. Because of differences in procedures and types of people contacted, various polling organizations may report different predictions about presidential candidate preference. A journalist who wants to make a case to support Candidate X may report only the poll in which Candidate X had the highest approval ratings.

In behavioral and social science, the problem of generalizability can have a different form. A researcher may want to know whether cognitive behavioral therapy (CBT) reduces depression. Typically, studies examine small to moderate numbers of cases, for instance, 35 patients who receive CBT and 35 who do not. To generalize results about effects of CBT to a large hypothetical population of “all depressed persons,” ideally, we would want a random sample drawn from that population. However, participants are often convenience samples, that is, people who were easy to recruit.

It is important to know what kinds of people were (and were not) included in a study. For example, if a drug study finds evidence that a new medication is effective and safe for healthy young men, that does not necessarily mean that the drug is also effective and safe for women, elders, children, and other kinds of people not included in the study.

Be careful not to overgeneralize results, particularly when there is little information about the types and numbers of people (or cases) included. *It makes sense to generalize information from a small group to some larger population only when people in the group resemble the population of interest.* This is discussed further in Chapter 2 in sections about samples and populations.

In science communication, authors are expected to discuss limitations that must be considered before drawing any conclusions. Limitations include the number and kinds of people (or cases) included in a study. *Science writing should make limitations of evidence clear; media reporting often does not.*

1.6 MAKING CAUSAL CLAIMS

In everyday life, and in science, we often want to know about causal connections. Consider a question raised by Wootson (2017). Do diet (artificially sweetened) soft drinks cause weight gain? If you are concerned about weight gain, and if artificially sweetened soft drinks cause weight gain, then you might consider avoiding diet soft drinks to avoid weight gain. However, it is possible that the association reported in some studies did not arise because of any direct causal impact of diet soft drinks on weight. Perhaps when people drink diet soft drinks, they feel free to indulge in other high-calorie foods, and perhaps it is those other high-calorie foods, not the soft drinks in and of themselves, that cause weight gain. If that is the correct explanation, then what you need to do to avoid weight gain is to avoid consuming high-calorie foods (rather than reduce diet soda consumption).

Causal explanations are attractive because they tie events together in meaningful ways. This is useful in science as well as everyday life. Sometimes when a cause–effect relationship is known, it suggests what we can do to change outcomes.

Demonstrating that two events are causally connected can be difficult, because there are often rival possible explanations. Well-controlled experiments can rule out many rival explanations. In everyday life, people sometimes jump to conclusions about causality on the basis of insufficient evidence.

1.6.1 The “Post Hoc, Ergo Propter Hoc” Fallacy

News commentators frequently offer causal explanations for events (e.g., the stock market went down because of a blizzard the previous day). This explanation is often just an opinion of the news commentator. The stock market might have gone down for other reasons (including random variations). This is an example of a common logical fallacy called “**post hoc, ergo propter hoc**.” This Latin phrase means “after this, therefore, because of this.” This (incorrect) reasoning goes like this: If Event A happens, and then Event B happens, then A must have caused B. Before we can conclude that Event A caused Event B, additional conditions are required. Here is another example. If you have a cold, take a large dose of vitamin C, and then the cold goes away, you might conclude that vitamin C cured the cold. However, the cold might have gone away on its own, whether you took vitamin C or not. Post hoc, ergo propter hoc reasoning uses one instance of co-occurrence (vitamin C, end of cold) to draw a causal conclusion. That is poor-quality reasoning that often leads to mistaken beliefs in causality. To conclude that vitamin C cures colds, you would need an experiment to evaluate whether the duration of colds was less in a group that took vitamin C than in a group that did not (controlling for other factors, such as placebo effects).

1.6.2 Correlation (by Itself) Does Not Imply Causation

You may have frequently heard the warning that correlation does not imply causation. This warning should be stated more precisely. It is more accurate to say, *Existence of a statistical relationship, such as a correlation, between variables X and Y, is needed to make claims that X causes Y. However, the mere existence of a statistical relationship does not prove that X causes Y. Alternative explanations for the statistical relationship between X and Y must be ruled out before we can believe that X causes Y.*

Let’s examine this idea carefully.

The word **correlation** has two meanings. First, sometimes people use the term *correlation* to refer to a specific statistic: the **Pearson product–moment correlation**, also called **Pearson’s r** . Second, the term *correlation* can be used in a broader sense; we can say that variables are correlated if they are statistically related using some statistical analysis. The statistical analysis can be something other than Pearson’s r . For example, if we compare average height for male and female groups and find that men are taller than women, we can say that sex (X) is statistically related to height (Y) or that sex is correlated with height.

We cannot claim that an X variable “causes” a Y variable if there is no statistical relationship of any kind between X and Y . In other words, the existence of a statistical relationship between X and Y is a *necessary* condition before we can consider causal inference.

However, existence of a statistical association is not enough evidence by itself to prove causality. Sometimes variables are statistically related (correlated) just by chance, or because the X and Y variables are related to some third variable Z , and Z may be the real “cause.”

Consider this example: If we measure ice cream sales (X) and number of homicides (Y) once a month for a year, there is a correlation between them. Months that have the most ice cream sales also have the largest number of homicides (Peters, 2013). Does eating ice cream cause people to commit homicide? That idea is obviously silly. A more plausible explanation is that temperature is related to both ice cream consumption and homicide. During hotter months, people may buy more ice cream; homicide rates are higher in hotter months (perhaps because people hang around outside more, or perhaps heat makes people more irritable).

Correlation (statistical association) is a **necessary but not sufficient** condition for making causal inference. Statistical association is necessary because we can’t conclude that X causes Y unless X and Y go together or co-occur. Statistical association is not sufficient by itself to prove causation because, even if X and Y covary, this co-occurrence may be due to the influence of one or more other variables; one of those other variables might be the real cause of X , or of Y , or both. In this example, heat or temperature might cause (or at least predict) ice cream purchase and homicide.

The effects of rival explanatory variables can be reduced or eliminated in well-controlled experiments and reduced by statistical controls. Mere co-occurrence is not enough evidence to make a causal inference.

Sometimes the need to look for a different explanation is obvious (as in the ice cream/homicide example). It would be absurd to argue that ice cream causes homicide. However, the need to consider rival explanations also arises in situations that are not so obviously silly. In the diet soft drink/weight gain example, it is conceivable that artificial sweeteners have causal effects on appetite or metabolism that really do lead to weight gain, even though the artificial sweeteners contain zero (or negligible) calories. However, the other explanation (that drinking diet beverages leads people to indulge in other high-calorie foods) is also plausible. (It is also conceivable that both these explanations are partly correct.) Both experimental and nonexperimental studies, with humans and nonhuman animals, would be helpful in sorting out the relations among variables and whether any of the associations are causal.

1.6.3 Perfect Correlation Versus Imperfect Correlation

Perfect co-occurrence (perfect correlation or statistical association) is rare. Consider the genetic mutation for hemophilia (Table 1.1). If a male child inherits this genetic mutation, he will have hemophilia. Most other heritable diseases do not show this perfect association. (For female children, effects of the hemophilia gene are ruled out by information on the other X chromosome.)

Table 1.1 Example of Perfect Co-occurrence or Perfect Correlation (Between Gene and Disease)

	Male Child Has Hemophilia	Male Child Does Not Have Hemophilia
Hemophilia gene is present	100%	0%
Hemophilia gene is absent	0%	100%

	Person Does Not Get Sick	Person Gets Sick
Person washes hands regularly	75%	25%
Person does not wash hands regularly	33%	67%

If a male child does not inherit the gene for hemophilia, he will not have hemophilia. In logical terms, the mutated gene is both **necessary and sufficient** for the disease. The mutated gene is necessary for hemophilia because a person can't get hemophilia without it. The mutated gene is sufficient for hemophilia, because if a person has it, he always has hemophilia. In other words, hemophilia always occurs when the mutated gene is present and never occurs when the mutated gene is absent.

Most associations in behavioral and social sciences and medicine are *not* perfect. Consider this hypothetical example for a behavior (washing or not washing hands) and a disease outcome (getting sick).

Table 1.2 shows an imperfect association. Only 25% of regular hand washers got sick, while 67% of the those who don't regularly wash their hands got sick. While most people who washed their hands did not get sick, hand washing did not guarantee that they could avoid getting sick.

The association between lung cancer and smoking is also not perfect. The risk for getting lung cancer is much higher for smokers than for nonsmokers. However, a few nonsmokers do get lung cancer, and many smokers do not get lung cancer.

In situations where associations are **not** perfect, it is likely that other variables are involved. Behaviors or conditions that sometimes (but not always) precede disease are often usually called "**risk factors**" rather than causes. Smoking is a risk factor for lung cancer. Some diseases have numerous risk factors (for example, risk for heart disease is related to smoking, body weight, sex, age, high blood pressure, and other factors).

We call behaviors that reduce risk for a negative outcome "**protective factors**." For example, hand washing is a protective factor against getting sick.

1.6.4 "Individual Results Vary"

Unless there is a perfect correlation (as in the hemophilia example), statistical associations or correlations between variables do not predict exact outcomes for all individuals. Consider the results of a study by Judge and Cable (2004), informally reported in Dittman (July/ August 2014). They reported that taller persons tend to earn more money (that is, height is correlated with salary). This is not a perfect correlation. If you are short, that does not necessarily mean that you will earn very little. Mark Zuckerberg (the founder of Facebook) is reported to be 5'7", but that did not prevent him from becoming one of the wealthiest men in the world. If you think about the implications correlations might have for your own outcomes, realize that individual outcomes differ when correlations are not perfect.

1.6.5 Requirements for Evidence of Causal Inference

Training in research methods and statistics provides the skills scientists need to think carefully about the evidence needed to support causal claims. Mass media journalists often rely on secondary sources or third-party content. By the time information filters through multiple communication links, details about the nature of the evidence and concerns about

limitations that affect the ability to generalize and make causal inferences are often lost. Third-party content often does not provide accurate information about generalizability and potential causality.

1.7 QUALITY CONTROL MECHANISMS IN SCIENCE

1.7.1 Peer Review

The science research process has mechanisms for information quality control. The most important mechanism is **peer review**. Researchers submit research reports to science journals (also called academic journals) for consideration (see note 2). The editor sends papers to peer reviewers (peers are expert researchers in the same field). Reviewers provide detailed criticism of studies, including evaluation of their research methods. On the basis of reviews, editors decide whether to reject a paper as inadequate, ask authors to revise the paper to correct errors or deficiencies, or (very rarely) accept the paper with only minor corrections. Papers are rarely accepted in their initially submitted form. Rejection rates for some journals are 80% or higher.

Peer review is fallible. Reviewers can also be subject to confirmation bias (they are more likely to favor conclusions consistent with their own beliefs). Reviewers may not notice all of the problems in a research report. However, peer review weeds out much poorly conducted research and improves the quality of published papers. The community of scientists in effect systematically polices the work of all individual scientists.

1.7.2 Replication and Accumulation of Evidence

A second important mechanism for data quality control in academic research is **replication**. Replication means repeating or redoing a study. This can be an **exact replication** (keeping all methods the same) or a **conceptual replication** (changing elements of the study, such as location, measures, or type of participants, to evaluate whether the same results occur in different situations). We should not treat findings from any one study as a conclusive answer to a research question. Any single study may have unique problems or flaws. In an ideal world, before we accept a research claim, we should have a substantial body of good-quality and consistent evidence to back up that claim; this can be obtained from replications.

Peer review and replication in science are fallible. However, they provide the best ongoing quality control checks we have. In contrast to science, there are few quality control mechanisms for most mass media communication.

1.7.3 Open Science and Study Preregistration

There are recent initiatives to improve the reproducibility and quality of research results in biomedicine, psychology, and other fields (Begley & Ioannidis, 2015; Open Science Collaboration, 2015). The Open Science model includes components such as preregistration of research plans and sharing details of data and methods. For further discussion, see Cumming and Calin-Jageman (2016).

1.8 BIASES OF INFORMATION CONSUMERS

1.8.1 Confirmation Bias (Again)

Information consumers or receivers also tend to select evidence consistent with their preexisting beliefs. Media consumers need to be aware that they can systematically miss kinds of information (which may be of high or low quality) when they select news sources they

like. Ratings of many web news sources on a continuum from left/liberal to right/conservative, along with assessment of accuracy, are provided at <https://mediabiasfactcheck.com/politifact/>. News sources that are extremely far left or far right tend to be less accurate.

Because of confirmation bias, people can get stuck: They continue to believe “facts” that aren’t true, and ideas that are wrong, because they never expose themselves to information that might prompt them to consider different possibilities. Consumers of mass media usually avoid evidence that challenges their beliefs. Philosopher of science Karl Popper argued that scientists also need to examine evidence that might falsify their beliefs. Scientists and people in general should consider evidence that challenges their beliefs.

1.8.2 Social Influence and Consensus

Should we believe something simply because many people, particularly those whom we know and respect, believe it? Not necessarily. Some incorrect beliefs are widely reported in mass media and held by millions of people. My personal favorite conspiracy theory is that alien reptiles control U.S. politics. Bump (2013) reported that more than 12 million people, or 4%, of the U.S. population said that they believed this theory in 2012–2013. To be clear, I strongly disbelieve that we are ruled by alien reptiles. (I am also not sure whether to believe Bump’s report that 12 million people really believe this; surveys are not always accurate.)

Consensus among science researchers can enhance the believability of a claim. However, even in science, consensus does not always guarantee accuracy. Experts can turn out to be wrong. For example, there was a consensus among nutrition researchers that eggs are bad for health because of their cholesterol content. Some recent research suggests that this widely held belief may be incorrect⁴ (Gray & Griffin, 2009), but the issue continues to be controversial.

A belief shared by millions of people is not necessarily wrong. However, consensus is neither necessary nor sufficient evidence that information is correct.

1.9 ETHICAL ISSUES IN DATA COLLECTION AND ANALYSIS

1.9.1 Ethical Guidelines for Researchers: Data Collection

Ethical issues arise when collecting data about people and nonhuman animals. For psychologists, the American Psychological Association has codes of ethics that protect the well-being of subjects (Campbell, Vasquez, Behnke, & Kinscherff, 2009). Research that involves human participants is evaluated by an **institutional review board**; research that involves nonhuman animals is evaluated by an **institutional animal care and use committee**. Ethical codes govern research in other areas such as biomedicine. Data collection cannot begin until ethics board approval of procedures has been obtained. Adherence to those rules is an ethical obligation for researchers. We should not harm the people or entities we study.

As an example of potential harm to a research participant, suppose that a study reveals that a person has a history of addiction. If that information gets into the hands of potential landlords or employers, it could have an impact on that person’s search for housing and jobs. Researchers must keep such records confidential.

Researchers also have an ethical responsibility to think about the potential impact of their research (both positive and negative) on public policy and the behavior of organizations and individuals.

1.9.2 Ethical Guidelines for Statisticians: Data Analysis and Reporting

The GAISE report states, “Students should demonstrate an awareness of ethical issues associated with sound statistical practice” (GAISE College Report ASA Revision

Committee, 2016). A separate document (American Statistical Association, 2015) discusses ethical issues in detail. Here is a list of ethical practices for data analysts, paraphrased from the American Statistical Association's ethics document. You will be reminded about these issues as you continue through the book.

1. Ensure that numbers are accurate. Fully disclose data handling procedures (such as deletion of cases or replacement of missing values) that could alter conclusions.
2. Make the limitations of the type of statistical analysis clear. (As each new analysis is introduced, you will learn about its limitations.)
3. Avoid behaviors that can lead to errors (including, but not limited to, cherry-picking a few results).
4. Avoid misleading presentations (such as “lying graphs”; see Section 1.10).
5. Avoid language that obscures results.
6. Do not overgeneralize. Do not make strong claims about characteristics of a population when your sample does not resemble that population.

Real-world problems in applications of data analysis are often not clear in introductory courses; students learn to do one analysis at a time using one small set of numbers. In actual practice, data analysts often work with large sets of messy data. Data analysts need to make many choices that involve difficult judgment calls. This book points out differences between the *ideal* use of statistics in artificially simplified situations and the *actual* application of statistics to real-world data. Sometimes decisions about “best practice” are difficult.

As Harris (2001) said, “Statistics is a form of social control over the professional behavior of researchers. The ultimate justification for any statistical procedure lies in the kinds of research behavior it encourages or discourages.” Science has rules and standards about good practice in collection, analysis, and presentation of evidence. These are discussed throughout this book.

Researchers should be aware that press releases from universities sometimes overhype research findings (Resnick, 2019).

This book discusses good practices in applied statistics that can potentially improve the clarity and honesty of research reports. When communicators present information in misleading, unclear, or dishonest ways, they risk loss of credibility, trust, and respect, not just for themselves but for the professions of statistics and science. When information consumers rely on incorrect information, they may make poor decisions.

1.10 LYING WITH GRAPHS AND STATISTICS

The most extreme form of lying with statistics is fabrication or falsification of data; this is rare. However, some common research practices slant information presentation in ways that can be called “lying with statistics.” The classic book *How to Lie With Statistics* (Huff, 1954) presented numerous examples.

Deceptive bar graphs are among the most common ways information communicators mislead information consumers. If you will be an information producer, you need to know how to set up “honest” bar graphs. When you are an information consumer, you need to know how to examine graphs to make sure that they are not misleading. Chapter 5 provides examples of clear versus misleading graphs and guidelines for evaluation of graphs.

1.11 DEGREES OF BELIEF

People rarely have time to collect all necessary information. Even for questions in science, we often do not have enough information to be confident about conclusions. Uncertainty is more common than people realize, even in areas such as medicine. There are many questions in medicine (such as what causes autoimmune disorders) for which medical research does not have good answers (Fox, 2003).

It is useful to think about scientific knowledge in terms of **degree of belief** instead of certainty. The philosopher David Hume said that “a wise [person] . . . proportions his [or her] belief to the evidence” (Schmidt, 2004). Degree of belief should be based on the *quantity* of *consistent* and *good-quality, systematically collected* supporting evidence. When there is little evidence (for example, results from only one study), people should not have strong belief in a claim. As additional good-quality evidence accumulates, degree of belief can increase. People should revise degree of belief upward or downward as new (good-quality) evidence becomes available.

This rating scale illustrates the concept of degree of belief. The use of a five-point scale and the exact verbal descriptions for each numerical rating are arbitrary.

1	2	3	4	5
Probably untrue	May be untrue	Not sure; insufficient evidence	May be true	Probably true

Fairly often, the best answer to research or public policy questions is that we do not have enough high-quality evidence to be confident that we know the correct answer. We should never assume that numerical results of one single study or mass media report are conclusive.

1.12 SUMMARY

Here are some questions to keep in mind when evaluating numerical (and other) information.

1. Is there evidence of communicator bias or self-interest?
2. Is evidence cherry-picked to fit the communicator’s argument?
3. Is the communicator far from the information source or not well qualified to evaluate the information?
4. Does the communicator have a good record for truth-telling?
5. What types of evidence are included. Anecdotes? Citations of specific, credible sources?
6. Have you considered your own possible biases as an information consumer? Do you accept information uncritically because it confirms when you already believe? Are you influenced by what other people believe?
7. Do data come from people (or cases) who resemble the population of interest? Are results generalizable?
8. Are causal inferences drawn when there is not enough information to prove a causal association? Remember that imperfect correlation or co-occurrence does not indicate causation.
9. Has information been subjected to quality control? (In science, this includes peer review and replication.)

10. Is the presentation of information deceptive (e.g., lying graphs)?
11. What ethical issues are at stake in the conduct and application of the research?
12. Is your degree of belief proportional to the quantity of good quality and consistent evidence? (You should never believe a claim on the basis of just one scientific study or one journalism report.)

Sometimes the best answer to questions such as “Are eggs harmful to cardiovascular health?” is that we don’t have enough evidence yet to answer the question. Unfortunately, lack of evidence does not prevent some communicators from making premature claims. When claims are made on the basis of limited evidence, contradiction and confusion often arise. It is better to reserve judgment until a large quantity of good-quality evidence is available. One single media report, or one single science report, is not “proof.”

Even if you do not plan to be a researcher, you can benefit from thinking like a scientist and statistician about numerical evidence you encounter in everyday life. Some decisions have high stakes. For example, you may need to decide whether to undertake a risky but potentially beneficial medical treatment. Ideally, you should have accurate information about potential outcomes. The higher the stakes, the more you need to know how to obtain trustworthy information.

The take-home message from this chapter is: *We all know a lot less than we think we do*, because most of us rely heavily on third-party content that has little or no information quality control. All of us (scientists, journalists, and information consumers) should be cautious about degree of belief. Sometimes the best answer to a question is: We don’t have enough good quality evidence. Courses in statistics and research methods teach you good practice in evaluation and presentation of evidence.

COMPREHENSION QUESTIONS

1. What is cherry-picking of evidence, and why is it deceptive? (Can you think of a book or media report that seems to present cherry-picked evidence?)
2. Give examples of self-interest that might make a communicator less believable.
3. Why is distance to original source of information an important factor when you evaluate message credibility?
4. What does it mean to say that a correlation (or association) between variables is imperfect?
5. Give an example of a risk factor, and a protective factor, not discussed in the chapter.
6. Why is the existence of a correlation (existence of co-occurrence or association) between X and Y not enough evidence for us to say that X causes Y ?
7. What is the post hoc, ergo propter hoc fallacy? (Give an example you have seen, different from the one in this chapter.)
8. What is confirmation bias?
9. What quality control mechanisms are used in science?
10. What is peer review? How can it improve the credibility of science reporting?
11. What is research replication? How can this improve the quality of evidence in science? How do exact replication and conceptual replication differ?
12. A researcher might say “the results of this one study prove” something. Is this justified?
13. What (approximate) degree of belief should you have on the basis of only one study?

NOTES

¹ Scientists are expected to be objective when they select information to report. However, scientists tend to focus selectively on information consistent with the most widely accepted existing theories; Kuhn and Hacking (2012) called this “selection of significant fact.”

² Numerous predatory, for-profit online journal publishers have emerged in recent years. It has become more difficult to determine whether online publications are credible. Research reports published in predatory journals are not valued by professional colleagues and universities. Beall’s List of Predatory Journals and Publishers names many publishers that are almost certainly predatory (<https://bealllist.weebly.com>). Additional warning signs that a publisher may be predatory:

- The journal invites you to submit your undergraduate or graduate thesis for publication (particularly if the journal title is not in your discipline or field).
- The journal offers to publish your paper without peer review.
- The journal asks you to pay for publication. (However, many legitimate publishers charge author fees to make journal articles open access on the web; therefore, a request for payment is not always an indication that a journal is predatory.)

If you are not sure whether a journal or publisher is predatory, search <journal name> or <publisher name> along with the term *predatory*. You can also ask mentors, advisers, or colleagues.

³ About half of Dr. Oz's medical advice is not supported by medical research (Belluz, 2014). Dr. Oz was investigated in a congressional hearing and paid large settlements in lawsuits for false advertising (Cohen, 2015).

⁴ This video about an imaginary time-traveling dietician makes fun of changes in dietary recommendations across the decades: <https://www.youtube.com/watch?v=5Ua-WVg1SsA>.

DIGITAL RESOURCES

Find **free study tools** to support your learning, including **eFlashcards**, **data sets**, and **web resources**, on the accompanying website at edge.sagepub.com/warner3e.

Do not copy, post, or distribute