

Chapter 2

General Research and Writing Practices



2.1 Insufficient Knowledge of the Literature

The first step in developing a scientific document is not writing but thinking and reading. Good authors are good readers. To write a good paper, you need to develop your own critical thinking, creative thinking, and understanding. You need to have read and critically considered what others have previously reported.

This error can take several forms, such as not having read the relevant literature, not understanding and integrating the work of others into the paper, or ignoring work that threatens or contradicts your findings or beliefs. Authors need to understand what has been previously published on the topic in order to frame the research question and to highlight novel elements of their contribution. If the author lacks sufficient interest in the topic to read about it in detail, then the author is not well positioned to convince readers to be interested in that topic. Failure to demonstrate familiarity with the literature and understanding of the topic also jeopardizes an author's credibility.

Remember, experts in the field will be reviewing your paper. Your initial drafts will be reviewed first by your primary reviewer and then by your co-investigators, co-authors, and research group head. When you submit a manuscript to a journal, it will be peer-reviewed. If you don't find the most up-to-date relevant information, then a reviewer is likely to do it for you, resulting in embarrassment and/or rejection of your paper.

An author needs to understand and communicate what the state of knowledge in the field is and describe what your paper adds to what is already known. You are trying to advance the field of knowledge, not just duplicate it. You cannot do this unless you are intimately familiar with what is already known. This should transcend, "There is almost no data on this subject in Bangladesh," the implication being that anything I say will be an improvement. While prior work may be limited, you need to look at similar settings or even dissimilar settings and see what other

researchers have found. What are the principal ideas, explanations, and data that are relevant to your particular paper?

If you cannot answer the question, “What does this paper add to what is already known about this subject in the literature?,” then you are not ready to write the paper. Expect to spend many days finding relevant articles and reading them critically before you can understand and then communicate clearly what new information or idea your paper adds. Different electronic search engines can help you identify different articles: By default Google Scholar lists the number of times an article is cited and so can help you identify articles that are influential, while PubMed can be set to list the most recent articles first.

When conducting a literature review, it is, at times, acceptable to put together a concept note or a first draft of a protocol by reviewing abstracts of journal articles. However, to cite information in a paper for submission to a journal, we recommend reading the manuscript for two reasons. First, a scientific argument that is sufficiently refined to be included in a peer-reviewed scientific article requires a nuanced understanding of the work you cite, a level of specificity that is unavailable from an abstract. Second, there may be data or an argument in a cited article that directly challenges a central idea you are presenting in your paper. If you fail to note it and address the implications for your paper, you risk losing credibility in the minds of readers and reviewers.

Finally, the excuse of “I couldn’t get the paper” is not acceptable in the arena of international scholarship. It is more difficult when articles are behind pay walls, but with persistence, nearly any article can be secured. Online resources and collaboration with other institutions and even directly writing authors can secure helpful sources.

Examples of the error		Alternative, better options	
X	Key studies in the field are not quoted.	✓	Search the literature carefully.
X	The studies quoted do not represent the best or the latest studies.	✓	Update literature search, and identify “citation classics”.
X	Study findings are misrepresented.	✓	Read all cited papers fully, not only the abstracts.

2.2 Insufficient Citations

Citations in the text that point to a list of references at the end of your article provides a standardized approach to acknowledge the sources of information and ideas that you have used. It allows readers to locate and review the basis of your arguments.

Learn and use a reference management software. Options include EndNote, Mendeley, Zotero, Papers, JabRef, and many others. Reference software helps you track the source of the information and ideas that contribute to your own scientific understanding.

Keep track of your sources in a physical or electronic logbook during your research. When you identify useful ideas or information, include a citation and reference in your notes.

2.2.1 *Not Providing a Reference to Support an Observation*

Scientific arguments require specificity. All statements that are not common knowledge or do not flow directly from your data require a citation within the text that points to a reference that supports the assertion. This requirement flows from the central importance of empirical findings in constructing and defending scientific arguments. Regular citation also reinforces the social construction of scientific understanding.

For novice scientific authors, this requirement may seem odd or stilted. We don't use this in our normal conversation. Even journalists commonly make general assertions without citations and references. Peer-reviewed scientific literature is different.

Consider the statement “*It is estimated that by 2050, half of all deaths will be a result of environmental mismanagement.*” Who made such an estimate? What is the basis of this estimate? Is this one person's opinion (Error 2.3.3)? If this estimate was based on a model, what are the model assumptions? Such general statements may be common in popular discourse, but in scientific writing, the reader needs to know the basis of each of your assertions. Readers can then judge whether this assertion and so the scientific argument within your manuscript is credible or not.

Examples of the error		Alternative, better options	
X	Pneumonia is a major public health problem in India.	✓	In 2018, pneumonia was the leading cause of death among children in India (ref).
X	Handwashing is effective against diarrheal diseases.	✓	Community level interventions that promoted handwashing have been associated with reduced incidence of childhood diarrhea (ref).

2.2.2 *Plagiarism*

Plagiarism is presenting other people's work as your own. This is a particularly serious error. It has destroyed the reputation and careers of many scientists. Web search tools make it increasingly easy to detect plagiarism.

A particularly egregious form of plagiarism is copying text word for word from another source and not attributing the source. Anytime an author quotes >3 words from a source, you should use quotation marks as well as a citation. More commonly in scientific writing, authors paraphrase the ideas and results from other articles and add a citation.

Authors can commit plagiarism unintentionally when they are pulling ideas together for a scientific manuscript or proposal. Authors might copy text from various articles and paste this text into a working document to help assemble relevant observations and ideas. Authors may subsequently insert this text into a draft manuscript losing track that the specific words originated from someone else. To avoid this error, whenever you copy text from another article, use quotation marks when you paste it into your own notes, and include a citation that points to the original author's work.

Example of the Error

*Built environment has direct and indirect effects on mental health** and *poor quality housing increases psychological distress* and *insufficient daylight is associated with increased depressive symptoms* (Evans 2003).

Cited reference:

Evans, G. W. The built environment and mental health *J Urban Health* 2003 Dec;80(4):536–55.

Abstract of Cited Reference

The built environment has direct and indirect effects on mental health. High-rise housing is inimical to the psychological well-being of women with young children. *Poor-quality housing* appears to *increase psychological distress*, but methodological issues make it difficult to draw clear conclusions. Mental health of psychiatric patients has been linked to design elements that affect their ability to regulate social interaction (e.g., furniture configuration, privacy). Alzheimer's patients adjust better to small-scale, homier facilities that also have lower levels of stimulation. They are also better adjusted in buildings that accommodate physical wandering. Residential crowding (number of people per room) and loud exterior noise sources (e.g., airports) elevate psychological distress but do not produce serious mental illness. Malodorous air pollutants heighten negative affect, and some toxins (e.g., lead, solvents) cause behavioral disturbances (e.g., self-regulatory ability, aggression). *Insufficient daylight is* reliably *associated with increased depressive symptoms.*

*The ***bold italic*** format reflects direct quotations from the published work.

X This is an error because the author is directly quoting from a source but using neither quotation marks nor a citation.

Alternative, Better Options

As Evans notes, “the built environment has direct and indirect effects on mental health . . . (and) poor quality housing appears to increase psychological distress. . . and insufficient daylight is associated with increased depressive symptoms” (Evans 2003).

Poor quality housing that provides little daylight worsens psychological health (Evans 2003)

Science is a social enterprise. Scientific writing requires that we give credit to others who have informed our ideas. A less egregious form of plagiarism than unattributed direct quotation involves using the ideas of others but failing to cite the source of these ideas and so presenting the ideas as your own. Because scientific discourse builds on the ideas and findings of others, scientific authors aim to situate their work within broader scientific discussion. It is important to cite the sources that led to the specific framing of the issues presented in your work.

Some journals are concerned with “self-plagiarism.” There are two related concerns here. First, most scientific manuscripts are framed as presentations of original data. Duplicate publication of the same work in more than one journal typically violates both the norms of science and the rules of individual journals. Journals want to ensure that original work is genuinely novel. When publishing multiple

articles from the same underlying study, sometimes some analysis, for example, the baseline comparison of characteristics between intervention groups, may be of interest to readers of multiple papers. If you are presenting some results that have been previously published, it is important to make this clear within the manuscript.

A second concern is that authors often sign over copyright to journals. Thus, if they are using the same language they have used before, they are actually using copyrighted material of a copyright they may no longer own. At its extreme, a concern with avoiding self-plagiarism means that an author would need to rewrite the methods section using different words even when producing the tenth paper describing various outcomes of a randomized controlled trial. This can become absurd. Best practice is to refer to a prior article that provided details and then offer a succinct summary.

(Thanks to Laura Kwong for her assistance in drafting this section on plagiarism.)

2.3 Weak Citations

Scientific reasoning is based upon what can be observed in the world. Authors support scientific arguments by pointing to various observations. An original scientific paper includes new observations and argues that they inform broader understanding. Although it is sometimes appropriate to cite specific arguments, ideas, or theoretical models, the most common citations are observations reported by other scientists. Three common forms of the weak citation error are:

2.3.1 *Citing a Secondary Source*

In this form of the error, the author cites an article that cites the original observation. Standard scientific practice is to cite the primary observation. It is a flagrant error if you cite an article that makes a similar point to the argument you want to make in your article, and the article that you are citing perhaps, in its introduction, cites the primary articles. Avoid this error by simply citing the primary article.

Sometimes, it is appropriate to cite meta-analyses or other reviews, but the best practice in most cases is to cite the relevant primary literature even if it requires multiple citations. Citing the primary literature points directly to the empirical basis of the assertion. It specifies where critical readers should look if they are interested in further exploring these data. It also signals to the reader, who may know the literature very well, that you are also familiar with the relevant literature. If you are citing work that people are not so familiar with, but it is important to your argument, this can be an important pathway to support a somewhat different interpretation than the dominant interpretation. This process encourages creative connections, critical thinking, and productive scientific argumentation.

2.3.2 *Presenting Conclusions Rather Than Data from References*

Scientific understanding advances by reasoned interpretation of observation. Indeed, an essential difference between scientific discourse and nonscientific discourse is this reliance on observation as the cornerstone of argument. Thus, if you want to make a persuasive scientific argument, you need to present the core data, not just a person's conclusion from that data.

Example: *A baseline evaluation of the quality of sexually transmitted disease case management was conducted in five areas of Chennai, in 2012, and it was found that there is an urgent need for health-care providers to adopt the syndromic approach to STD treatment.*

In this example, the cited study may well have concluded that the health-care providers' performance was so poor in detecting and treating sexually transmitted diseases that a move to a syndromic approach was the best option. But if this is being presented as evidence that sexually transmitted disease diagnosis and treatment were poor, why should a scientific thinker have to accept the judgment or opinion reached by someone else? Accepting another's judgment without personally evaluating the data upon which that judgment is based is nonscientific reasoning. Nonscientific reasoning is out of place in a scientific manuscript.

Consider the alternative, better option: *In a baseline evaluation of the quality of sexually transmitted disease case management conducted in five areas of Chennai in 2012, 74% of persons presenting with symptoms of sexually transmitted diseases were given treatment that differed from World Health Organization guidelines.*

Now, the reader is no longer being asked to accept the interpretation of the author of the original study or of the author of the present manuscript. The reader has been given the primary observation that forms the basic unit of reasoning and so can either accept it as appropriate to the idea being developed or not. With the data, the reader can follow the author's reasoning.

2.3.3 *Arguing from Authority*

An argument from authority asserts that readers should accept a statement as true because of the authority of the person who spoke it. In everyday life, we depend upon arguments from authority to help navigate the world. We believe the auto mechanic who tells us our car will not start because the battery is too weak to hold a charge. We believe the attorney we consult who suggests that adding a specific clause in a contract will prevent subsequent legal problems. Arguments from authority are commonly used in many religious traditions and among journalists.

A distinctive feature of scientific reasoning, by contrast, is that it eschews arguments from authority and instead asserts that statements are credible because of the empirical evidence that supports them. Scientists do not believe statements because

they were uttered by a prestigious university or government official. Scientific reasoning requires evidence.

Examples of the error		Alternative, better options	
X	Many experts emphasize that shared toilets are the only solution for urban slum residence.	✓	Because of severe constraints on space, shared toilets will continue to be a common option in urban slums the foreseeable future.
X	Daniel Kahneman, a Nobel prize winning economist, notes that human decision-making is frequently illogical.	✓	Numerous formal assessments find that human decision-making is frequently illogical (references).

2.4 References Not in Standard Style

There are many times that a scientist is required to exercise creativity and ingenuity. Writing endnotes is not one of those times. Endnotes for manuscripts have standard formats well detailed in the “Uniform Requirements for Manuscripts submitted to Biomedical Journals” (www.icmje.org).

Various reference management software programs are available that assist in tracking and reporting references including EndNote, Zotero, Mendeley, Papers, JabRef, and many others. They allow an author to quickly insert bibliographical information. They automate renumbering references when text is resequenced after copying and pasting. They can quickly convert from one reference format to another if a journal requires a different reference format.

2.4.1 Varying Citation Format

Different journals use different formats for citations and references. There are two general approaches. Most journals sequentially enumerate the references in the order that they appear in the narrative. Different journals that use sequential numbering require that the citations within the text be displayed differently. Some prescribe that numbers be displayed within square back brackets. Others want numbers in parentheses. Others request superscripts. Some journals want reference numbers to precede periods or commas. Others want them to follow.

The other general approach is to list references at the end of the article alphabetically based on the first author’s last name. The in-text citations include one or more of the authors’ name and the year of publication.

When drafting a manuscript, look up your target journal’s reference format and use it. If you are writing a proposal or other piece of work that does not have a set format, then use a format that is easy for readers to understand. If you are space constrained, choose an enumerated format.

Do not mix formats that is sometimes using author's last names in parentheses and other times using numbers. Sometimes, copying and pasting from different documents create this problem. It risks confusing readers and making it difficult for them to connect to your references.

2.4.2 Not Proofreading References Prior to Submission

None of these reference management programs work flawlessly. All have their strengths, weaknesses, and idiosyncrasies. Prior to submission, the first author needs to carefully review each reference, ensure that it is complete, that capitalization is appropriate, and that there are no spelling or other obvious errors. When circulating a submission-ready manuscript to co-authors for their sign off, the references should be proofread. Submitting sloppy references communicates a lack of attention to detail. Journal editors prefer to engage authors who attend to details.

If response to further review requires any changes in the references, this often requires redeploying the reference management software that will likely replicate many of the earlier errors. These can be minimized by making corrections to the source references within the management software, but because of imperfections in reference management software, this is insufficient. Prior to resubmission, the references need to be proofread again.

2.5 Not Using Standard Draft Manuscript Form

Most journals have specific instructions for manuscripts submitted to them, usually detailed in their website under "Instructions to Authors." However, as a good starting point, the following generic style would be appropriate for a first draft manuscript sent to co-authors for review.

1. Format a title page to include:
 - The title of the article
 - First name, middle initial, and last name of each author (check the journal to see if they have a limit on the number of authors)
 - Each author's institutional affiliation as a superscripted note
 - Targeted journal(s)
 - Main text total word count
 - Abstract total word count
 - Key words
2. Include an abstract in the format and within the word length of the targeted journal. If the journal choice is uncertain, then include a structured abstract (text separated into sections labeled background, methods, results, and conclusion) of no more than 250 words.

3. The main text of the article should be in the traditional format of introduction, methods, results, and discussion (IMRAD). Different disciplines and different journals have different norms regarding the appropriate length of an article. The main text should not exceed the word limit for your target journal. Shorter articles are particularly attractive to most journal editors. If the journal does not suggest a limit, look at the length of articles that they generally publish. A manuscript that is too long risks discouraging reviewers, editors, and readers. By contrast, if a paper is too short, editors and reviewer can request that more information be included.
4. The manuscript should be double-spaced using a common font size 12. This provides more space for comments for reviewers of both the paper and electronic version.
5. The narrative text should be in a single column. Don't try to make it look like a formatted two-columned journal article. It makes it harder to review electronically, and it is also not the form it needs to be in for a specific journal submission.
6. Indent the first word of each paragraph one tab width (0.25–0.5 inch), or skip a line between paragraphs to signal the reader that this is the start of a new set of ideas.
7. Align text to the left. (Avoid Error 4.8.)
8. Insert the acknowledgments after the discussion. Then add references up to the limit permitted by the journal.
9. Tables and/or figures should be placed after the references. Journals often limit the number of tables and/or figures.

2.6 Repeating Information

Editors of scientific manuscripts prefer succinct writing. Don't repeat ideas. Say it well and say it once. A useful strategy to reduce repetition is by carefully considering the logic of your arguments in presenting the ideas so that they build progressively. If a point is so important that you want to ensure that reader see it, then include it in both the body of the paper, and the abstract, which is a summary of the manuscript.

A subtle version of this error is including both proportions of a dichotomous outcome in a results table (see last example).

One situation where a modicum of repetition may be appropriate is in the development of some ideas in the discussion when it is appropriate to link the development of these ideas to specific study results and/or to issues of study rationale raised in the introduction.

However, in a linked discussion, the important point is not to repeat the words but rather to make a logical connection between what was raised earlier and the discussion about to take place. Thus, a short recall, without quantitative details, is sufficient. Some journals, including *The Lancet*, want the first paragraph of the discussion to summarize the main results.

Examples of the error		Alternative, better option	
X	“Disease X causes XXX deaths annually worldwide” used in the first paragraph of the introduction and in the first paragraph of the discussion.	✓	Don’t repeat an idea. Say it well and say it once. If you are unsure about where to mention it, review Error 3.2 that clarifies the respective roles of each section of a manuscript to identify the most suitable place.
X	Full repetition of results, with quantified data and statistical tests in the discussion section.		
X	Household pays for electricity Yes 3 (10%) No/don’t know (90%)	✓	Household pays for electricity 3 (10%).

2.7 Labeling a Scientific Document as “Final”

Avoid the word “final” in the title or the description of any scientific document. Scientific thinking is always open to revision. To call a document final implies either dogmatic close-mindedness or naiveté, both characteristics that are inconsistent with a genuine scientific outlook.

Examples of the error		Alternative, better options	
X	Attached is the final version of the protocol.	✓	Attached is the version of the protocol approved by the Institutional Review Board.
X	Here is the final version of the manuscript.	✓	Here is the published version of the manuscript. (Who knows, there may be letters to the editor or subsequent insight that requires further revisions?).

2.8 Characterizing an Observation as “The First”

Scientists take pride in identifying novel observations. Galileo was the first person to see moons around Jupiter. Darwin was the first to both notice the very high variation of bird species on tropical islands and to suggest that this variability was best explained by evolution of species. Watson and Crick were the first to identify the structure of deoxyribonucleic acid (DNA). Part of that task of writing a manuscript is to explain to the readers what is new about the information that is being presented and how this new information changes or refines global scientific understanding. In this context, many authors will assert that their scientific findings are “the first.” However, there are three problems with describing one’s scientific findings as “the first.”

1. These assertions can create controversy and ill feeling with some scientists writing venomous letters to the editor disputing the claim of primacy. Such ill feelings do not help scientific understanding progress. Indeed, if one of your subsequent papers or research funding proposals is then reviewed by one of these

scientists who felt slighted by not being appropriately recognized in your earlier work, you risk receiving an unnecessarily devastating review that does not fairly consider the merits of your work. Indeed, many journal editors (e.g., those at the Lancet) will not publish claims of first primarily because they prefer to avoid such nonproductive ego-driven controversy.

2. Every observation can be described as a first if there are sufficient qualifications. Thus, the assertion of “first” is not, in itself, meaningful, for example, “This is the first time that hepatitis E virus has been confirmed using advanced molecular methods in environmental water supplies in Shakira District during the dry season at night using locally trained staff.” Asserting that something is “first” does not communicate why it matters.
3. These assertions distract from useful explanations of how these observations contribute to global scientific understanding. If a health condition has been found in the other 10 countries where it has been looked for, then saying that this is the first time this has been recognized in Bangladesh tells us more about the interest of Bangladeshi scientists in this condition and the funding available to work in this area than about the health condition itself or the situation in Bangladesh. It does not tell readers why this observation is important.

Like all rules in the guide, this one is not absolute. An occasional claim of first may be defensible and help to clarify to the reader how to interpret the results, but >95% of scientific articles are best written without any claim to “first.”

Examples of the error		Alternative, better options	
X	This is the first time that an association between hepatitis C infection and carcinoma of the liver has been demonstrated in Liberia.	✓	The link noted between hepatitis C and liver carcinoma in this population in Liberia provides further evidence of the importance of hepatitis C as a leading cause of hepatocellular carcinoma globally. It suggests that for a low-income country like Liberia, preventing the transmission of hepatitis C may be the most cost-effective way to prevent liver carcinoma.
X	This is the first time that Nipah virus antibodies have been identified in dogs in Bangladesh.	✓	Nipah virus infects a wide range of mammals. Earlier studies in Malaysia identified dogs with evidence of Nipah virus infection, but similar to our findings in Bangladesh, dogs appear to be dead-end hosts rather than the reservoir of the infection.

2.9 Errors in Reasoning

Scientific reasoning is central to interpreting our scientific results and to sound, persuasive communication with our colleagues. There are many ways that scientific reasoning can go awry. Indeed, one of the main benefits we derive from co-authors and external reviewers critically reviewing our manuscripts is that they criticize our reasoning and so help us to improve it. Some criticisms of scientific reasoning reflect

different interpretations of reported observations in the published scientific literature. What follows, however, are more formal errors in the structure of argument.

2.9.1 *Casual Assertion of Causality*

Scientists take the idea of causality very seriously. Indeed, much scientific work is centered around developing causal hypotheses that explain a relationship between characteristics and exposures in the world and subsequent outcome. When a scientist concludes that a particular chemical exposure caused illness, this is an argument that is based on careful observation, a biologically plausible mechanism, systematically collected data that demonstrates a statistical association, and rejection of alternative explanations including bias and chance [1].

By contrast, when nonscientists speak they tend to be much less careful in their assertion of causality. Business journalists commonly assert that the stock market went down because, for example, the weather was cold, a large company reported disappointing quarterly results, or investors were concerned about recent political developments. Similarly, politicians will assert, for example, that the reason crime has increased is because there are too few police officers. Sport journalists and fans will assert that the reason the home team lost the soccer match is because they did not take their opponents seriously. Each of these assertions may or may not reflect a genuine causal relationship, but none of the people making the assertion is offering a rigorous scientifically persuasive argument.

Such casual assertions of causality, which might be acceptable in casual conversation political speech or daily journalism, is not acceptable in scientific writing. Thus, especially in the introduction and discussion sections of the manuscript, it is critical for your credibility as a scientist not to assert causality unless there is rigorous evidence to support this assertion.

Examples of the error		Alternative, better options	
X	Banning overnight poultry storage at live bird markets have been found to reduce influenza H9N2 circulation substantially in Hong Kong.	✓	After overnight poultry storage at live bird markets in Hong Kong was banned, influenza H9N2 circulation decreased among market poultry.
X	Due to higher temperature, the number of non-cholera diarrhea cases also increased among the individuals with lower educational attainment, non-concrete roofs, and unsanitary toilets.	✓	As temperatures increased, the number of non-cholera diarrhea cases also increased among individuals with less education, non-concrete roofs and unsanitary toilets.

Examples of the error	Alternative, better options
<p>X Development project implementation also faltered, the reasons being financial constraints that produced cost overruns and procurement delays, foolhardy recruitment of under-skilled personnel and ill-planned career management, and imprecise delineation of the respective roles of development planning and supporting agencies.</p>	<p>✓ Fewer than 10% of development projects achieved their target objectives. Commentators suggest that the factors that most likely contributed to this underperformance included financial constraints that produced cost overruns and procurement delays, recruitment of under-skilled personnel and ill-planned career management, and imprecise delineation of the roles of development planning and supporting agencies.</p>

2.9.2 Assuming Association Is Causality

Much scientific work aims to identify associations between different phenomena. For example, is a particular exposure (drinking raw date palm sap) associated with a particular outcome (developing Nipah virus infection)? When we construct 2×2 tables or evaluate if there are different mean values between different groups, we are exploring whether there are associations within our data. An important element of our data analysis is to identify relevant associations within our data.

However, just because we find an association, this does not mean that the exposure caused the outcome. For example, if our analysis shows that people who have a lower income have a higher incidence of tuberculosis compared to people who have a higher income, it would be an error in scientific inference to conclude that low income causes tuberculosis infection. Consider for a moment what mechanism we would be asserting. Does the individual *Mycobacterium* have receptors that only attach to the alveolar cells of persons who have an income less than \$100 per month? Does the individual *Mycobacterium* wait to see how much money someone spends a month before deciding whether or not to infect him? In this example, low income is better considered an indicator of an environment that puts certain people at risk rather than a cause. For example, people who have low income more commonly have poor nutrition, and this poor nutrition reduces the capacity of the body to defend itself from an infection from *Mycobacterium*. Additionally, people with low income tend to live in more crowded settings where it is easier for respiratory diseases to spread from one person to another. Thus, there is an association between wealth and tuberculosis, but the causal mechanism is a deeper underlying mechanism.

There are a number of other reasons that we might find associations between exposures and outcomes in our data. Three common reasons for associations in our data are bias, chance, and confounding. There are entire books written on each of these topics, and we encourage you to read them. However, when it comes to interpreting your data, any time you see an association, you should be asking yourself

the following: What is underlying this association? Is there bias? Could this have arisen by chance? Is this a marker of confounding?

Scientific writing is most persuasive when it invokes a thoughtful, conservative interpretation of association. When discussing an association in the result section, for example, one should never use language that asserts the relationship is causal. In the results, you are only presenting the data and identifying associations.

The argument that an association is causal is an argument that should consider the potential mechanism of action; the possibility that the association is a result of bias, chance, or confounding; and results from other studies including different types of evidence that supports a causal mechanism. An assertion of a causal relationship is an argument that should be made in the discussion section; indeed, such an argument is often the major point of the discussion section.

2.9.3 Assuming Reported Behavior Reflects Actual Behavior

Research in the health sciences often considers human behavior, what people do, and what might influence what they do. Scientific study of human behavior requires deciding how to assess behavior. Usually, the easiest and least expensive approach is simply to ask study respondents how they behave. This can be appropriate and useful, but considerable literature illustrates that compared with actual practice, people generally overreport socially desirable behavior and underreport stigmatized behavior. Scientists should not take reported behavior at face value but consider the likelihood that the reported behavior is not accurately reflecting actual behavior [2]. These considerations are an important aspect of how we interpret our results and so should be considered in the discussion and the limitations.

Sometimes, we use research methods that permit us to directly observe behavior. Although the presence of an observer has been repeatedly demonstrated to alter behavior, observed behavior is often less biased compared with reported behavior. Nevertheless, even scientists who study observed behavior must keep in mind the difference between behavior when an observer is present and the behavior that occurs when people are not being observed.

For example, scientific studies comparing reported handwashing behavior to observed handwashing behavior consistently demonstrate that reported handwashing vastly exceeds observed handwashing [3–5]. Indeed, the differences are so great that reported handwashing behavior is not a valid proxy measure of handwashing practice. Similarly, the handwashing literature provides strong evidence that the presence of an observer markedly increases handwashing [6–9].

In scientific narrative when referring to behavior that has been studied by other researchers or when describing your own work, it is important to keep in mind the deep biases associated with reported behavior. Therefore, when describing behavior, it is useful to clarify whether the behavior was observed or reported.

Examples of the error		Alternative, better options	
X	After the intervention, respondents were less likely to defecate in the open.	✓	After the intervention, fewer respondents reported defecating in the open.
X	In Bangladesh, the rate of exclusive breastfeeding in the first 6 months is 64%.	✓	In the 2011 Bangladesh Demographic and Health Survey, 64% of mothers reported exclusively breastfeeding their children during the child's first 6 months.

2.9.4 *Confusing Imperfect Recall with Recall Bias*

Human memory is imperfect. If you ask a colleague what they ate for lunch 17 days ago, most would be unable to provide an accurate response. We do not remember all of our experiences. This is imperfect recall. Imperfect recall does not necessarily constitute a bias. Recall bias occurs when different groups of people within the study are likely to remember experiences differently. For example, assume you are conducting a case-control study exploring risk factors for leg fractures. If the injury occurred 2 weeks previously, and you ask people what they were doing in the minutes preceding the injury, cases, that is, people who had experienced a fracture, are much more likely to have carefully considered the events that led up to the fracture and so are likely to recall details of what type of shoes they were wearing, where they were, and what the visibility and footing was. By contrast, if you ask controls about their precise exposures at the same time of day 2 weeks previously, they are much less likely to recall rich details of their experience. Thus, there may be systematic differences in the recall of cases and controls, not because their exposures were different but because their recall of events is different. This is recall bias. All study subjects have imperfect recall. If there is no reason to believe that this recall will differentially affect reports of exposures or outcomes, it should not be labeled as recall bias.

Examples of the error		Alternative, better options	
X	Since the data on exposures to sick poultry was collected by interview, there is a risk of recall bias.	✓	Although our study subjects likely did not recall all of their exposures to sick poultry, because people in this community do not consider sick poultry to be a risk factor for human illness, we would not expect any bias.

2.9.5 *Confusing Absence of Recognition with Absence*

Authors should not blithely assume that all occurrences of a phenomenon of interest are known to science and reported in the scientific literature. Many events of scientific interest are neither recognized nor recorded in the scientific literature.

Examples of the error		Alternative, better options	
X	Mortality in ducks and geese as a result of highly pathogenic avian influenza H5N1 infection had never occurred in Bangladesh.	✓	Mortality in ducks and geese as a result of highly pathogenic avian influenza H5N1 infection had never been confirmed in Bangladesh.
X	The last of the four Nipah outbreaks from India was in 2019.	✓	The last recognized outbreak of Nipah in India was confirmed in 2019.

2.9.6 *Asserting Seasonality with a Single Year of Data*

Asserting that a phenomenon that occurs at different frequencies in different seasons of a single year is due to seasonality is an error in scientific inference. This is an error because it assumes a pattern when no repetitive pattern has been observed. With only a single year of data from South Asia, for example, only one rainy season was observed. Cases may have increased during the rainy season because a new strain of the pathogen was introduced into the community, a strain that the community did not have immunity against. The strain may have been introduced during the year of observation during the rainy season, but the following year, a new strain might be introduced at a different time of year. We are much less prone to scientific error and have much more credibility if we draw conclusions conservatively from our data. Multiple years of data that show a similar pattern provide a stronger case to assert that the variability in the observation over time is associated with seasonal patterns.

So what should we do if we have 1 year of data and see more cases in the rainy season than in the dry season? It is reasonable in the discussion section to note that the cases were more common in the rainy season, but multiple years of data would need to be observed to see if this is a seasonal pattern.

2.9.7 *Drawing Conclusions Using Confirmation Bias*

Confirmation bias refers to the human tendency to see patterns in the world that are consistent with previously held beliefs [10]. It is a particularly pernicious bias for scientists because we strive to bring forward new information and to draw sound conclusions.

Confirmation bias often affects scientists when we look at our data and see the patterns that we expect. For example, if people in the intervention group reported less illness, then the data makes sense to us, and we don't dig deeper. By contrast, when we find an association that is unexpected, for example, that disease is more common among people who received the intervention, then we carefully reevaluate the evidence. We check to see if we made a coding error in the analysis or if there was some way the question was framed that might have confused respondents. In

short, we invoke a double standard of accepting results that confirm our preconceptions and working to identify problems with evidence that runs counter to our expectations.

Another common manifestation of confirmation bias in science is interpretation of borderline p -values. If the point estimate of an association is in the direction that supports the unifying theory that the author is proposing, but the p -value is 0.10, authors commonly assert that “borderline result that supports this interpretation.” By contrast, if the association is not consistent with the author’s favored interpretation, the association is more likely to be left out of the manuscript, ignored in the narrative results, or dismissed as “not significant.”

Confirmation bias is so deeply rooted in our human capacity to see patterns in information and the incentives that scientists have to find interesting associations that it is difficult to avoid. A benefit of peer review is that reviewers may not share the authors’ preconceptions and so offer alternative interpretations of the data.

As an author, consider the risk of confirmation bias in your interpretation. Seriously consider the strengths and weaknesses of alternative interpretations. Consider the limitations in your data and available data in supporting the most likely interpretation. A conclusion that is based on evidence while also conceding weaknesses and alternative interpretations is more persuasive to a scientific audience.

	Examples of the error		Alternative, better options
X	The evidence supports that pesticides contributed to the elevated lead levels among mother.	✓	The evidence that pesticides contaminated with lead were associated with elevated blood levels is mixed. We found a strong association with reported use of a particular brand of pesticide and blood lead levels, but when we later collected samples of this pesticide, those samples did not contain lead. It is possible that lead arsenate intermittently contaminates commercial pesticides, but further study will be needed to assess this.
X	We found no association between child nutritional status and risk of infection.	✓	Both well-nourished and poorly nourished children were at risk of infection. Indeed, we found no association between child anthropometric measures and risk of infection though the number of observations were small so we had limited power for this assessment.

2.10 Constructing a Multivariate Model Using Only Statistical Criteria

Scientists are commonly interested in understanding how multiple factors interact to produce a particular outcome. Much of our research efforts are aimed at clarifying these causal pathways. When scientists explore statistical associations between exposures and outcome, they are usually striving to understand if there is an underlying causal connection.

Real-world causal pathways of health outcomes are characteristically complex. Multiple factors generally need to be present (e.g., there is a pathogen in the environment, there is a person who is exposed to the environment, the person is susceptible to the infection). In addition, causal pathways typically have sequences where one exposure must precede another in order for the effect to occur. For example, the pathogen must be present in the environment before the person enters the environment. We are much more likely to add insight to global scientific understanding of underlying causal pathways if we seriously reflect on the likely underlying causal mechanism and then construct our investigations and our data analyses to query these pathways.

All too commonly, analysts simply dump all their exposure variables into a multivariate model and use backward elimination to identify those exposures that are most strongly associated with the outcome and then offer this as a final model. This approach provides no consideration for the potential that two variables may be measuring the same underlying characteristic. It also invokes an implicit causal structure that all the exposures occur simultaneously and without interacting with each other to generate the outcome. This is a naïve and unlikely map of the way processes unfold in the world [11].

A better approach is to develop a causal model that explicates how the scientist believes the various factors are likely to co-produce the outcome and then use this conceptualization to decide which factors to test in the model. There is considerable scholarship on directed acyclic graphs that provide graphical support to help illustrate proposed causal paths and the impact of confounding and temporal sequencing [12, 13] The researcher's proposed causal model can be included as a figure in the paper. This way, readers can follow the hypothesized causal map and understand the judgments used in building a multivariate model.

This is a very different approach than large machine learning efforts that aim not to detect causal relationships but rather to find associations and then use those associations to predict subsequent activity. This type of prediction algorithm has been remarkably successful at identifying patterns in marketing data. In some settings, this widespread search for association in large data sets have been used to identify unexpected associations that may be worth further exploration. This approach remains uncommon among scientists who generally strive to elicit causal understanding. The statistical approach employed should align with the analyst's aspiration.

Examples of the error	Alternative, better options
<p>X Tobacco use and male sex are highly correlated (1/34 female respondents reported regular tobacco use as compared to 11/16 males); therefore, although both characteristics meet the specified criteria for inclusion in the final model, only male sex is included.</p>	<p>✓ Tobacco use and male sex are highly correlated (1/34 female respondents reported regular tobacco use as compared to 11/16 males); because tobacco use is known to affect taste (the primary outcome), it was included in the model and sex was dropped.</p>
<p>X We used univariate logistic regression to select predictor variables significant at the $p < 0.2$ level for inclusion in the full model. We used sequential backward elimination of variables with the weakest association to reach the final model of variables all with $p < 0.05$.</p>	<p>✓ Exposures were grouped in four blocks following the conceptual model: (1) attitude, (2) knowledge, (3) school facilities and programs, and (4) practices. We performed bivariate analysis between exposures and outcome to calculate crude association. We further considered only those exposures associated with outcomes with a $p < 0.2$. We then conducted multivariable analysis among the exposures within each block including confounders identified in the conceptual model. We retained exposure within each block associated with an outcome at the $p < 0.05$ level. We then built an overall multivariate model by using exposure variables from each block that were associated with school absence at the $p < 0.05$ level and which captured most of the measurement.</p>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

