

2.6: Statistical reasoning

Introduction

So far in the course we have talked about **statistical reasoning**. Statistical reasoning as part of the **critical thinking** tool kit. What this should mean for you in practice is that we will be developing and enhancing your critical thinking skill set. Here, we'll define our terms and place statistical reasoning in context.

Critical thinking

What do I mean by critical thinking? You'll find various definitions and discussions, but the Wikipedia entry on the subject seems as good a start as anything else I have heard or read:

“Critical thinking is a way of deciding whether a claim is always true, sometimes true, partly true, or false” (5 August 2013, http://en.Wikipedia.org/wiki/Critical_thinking).

Which raises the question, what is truth (false)? And what about the frequency qualifiers “always”, “sometimes”, or “partly”? This hints at one of the strengths of science: “Truth in science ... is never final, and what is accepted as fact today may be modified or even discarded tomorrow” (p. 2, National Academy of Science 1999). Scientists extend and even correct the work of their predecessors.

Critical thinking has many definitions, but they coalesce broadly as the search for **rational justification** — arguments and decisions based on reason not emotion — given sets of **facts**. We'll define facts as assertions about the world that are **verifiably true**. Related to facts, we define **evidence** as collection of facts used to infer an assertion or belief is **objectively true**. Objectivity holds independently of the observer. In practice, critical thinking can be improved by identifying and developing certain skill sets, from deductive reasoning to use of a statistical toolkit, although skills alone do not necessarily lead to sound critical thinking (Bailin 2002). Emphasis on the importance of statistical reasoning, for example in the **Evidence-Based Medicine** (EBM) approach, is now well established. EBM is an expansive concept, but includes the philosophy that decisions in medicine and health care should be based on evidence (results) from well designed and executed research. More generally, much has been written on the subject. My purpose here is to sell you just a bit that you will learn more than formulas and statistical tests in this course; if we do this right you will improve your critical thinking skills.

And the tool kit for critical thinking? Science has the **scientific method** to offer. In turn, by adopting statistical reasoning approaches you will be more cognizant of whether or not you are indeed engaged in critical thinking about your work.

A working checklist for critical evaluation of a project:

- “All sides” of the problem are fairly represented
- Authority is not synonymous with empirical truth: trust, but verify
- Correlation (association) may or may not infer causation
- Do conflicts of interest compromise conclusions?

In other words, identify and rigorously consider assumptions made to reach conclusions, evidence given in support of conclusions, and considerations of bias held by the researcher. The “all sides” part requires judgement — not all sides of a research question are equally valid; the point about “fair” representation requires that opposing arguments are actually held by others and not simply a strawman characterization of a position held by no one.

Rigid adherence to a “how to do science checklist” is likely to be wanting, and thus practicing and successful scientists show more flexibility. And, the list is certainly incomplete because it presumes observations and data collection are from a reliable, unbiased source. Before proceeding with how statistics informs our critical thinking in science, let's illustrate bias.

Bias

Bias is defined as any tendency that may hamper your ability to answer a question without prejudice (Pannucci and Wilkins 2010). Although we tend to think of ourselves as highly rational, psychologists have documented and named numerous, and in some cases, highly specific kinds of cognitive biases we may have about “**real world data**” that may impair our judgment about results of experiments. At its best, application of the scientific method is our best tool kit to help protect our conclusions against bias (Nuzzo 2015), by proper experimental design and possible study biases. A couple of comics from [xkcd](#) illustrate.



Figure 2.6.1: “Survivorship bias,” <https://www.xkcd.com/1827/>.

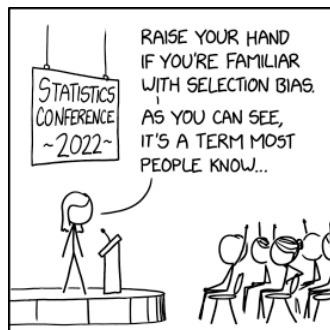


Figure 2.6.2: “Selection bias,” <https://xkcd.com/2618/>.

Wikipedia lists dozens of named biases which may influence our ability to interpret science:

- **confirmation bias**, e.g., given extraneous information about ancestry or sex, forensic investigators were more likely to assign ancestry or sex of skeletal evidence to a group that confirmed the contextual information (Nakhaeizadeh et al 2014).
- **congruence bias**, a type of confirmation bias, the tendency to test only a favorite or initial hypothesis over consideration of alternative hypotheses.
- **observer-expectancy bias**, e.g., randomized controlled trials of acupuncture often find equivalent responses to real and placebo acupuncture despite both appearing superior to no treatment (Colagiuri and Smith 2012).
- **selection bias**, the selection of individuals, groups, or data for analysis in such a way that proper randomization is not achieved, e.g., mitotic counts (Cree et al 2021) and cell differentials and Figure 2.6.2. Selection bias, where the study population is not well-defined or if is, the method by which subjects are recruited favors one kind of subject over another.
- **surveillance bias**, e.g., association between myocarditis and COVID-19 vaccines may be due to increased focus to identify myocarditis (Husby et al 2021).
- **survivorship bias**, focusing on subjects that passed a selection process as if they represent an entire group, e.g., from ecology: rare species more likely to go extinct than abundant species (Lockwood 2003) and Figure 1.

Bias in research may occur at any level, from study design to analysis and publication of results. Bias leads to **systematic errors** that may favor one outcome over others. Clearly then, bias is something one wants to avoid in science, and most scientists would probably agree strongly that bias is unacceptable. However, bias can creep into projects in many forms. Particularly pernicious causes of bias are conflicts of interest, with the notable case of a 1990's set of trials in gene therapy (Wilson 2010). Biased research may be more harmful to science than deliberate **misconduct**, at least in part because there are, or should be, mechanisms to detect fraud (e.g., Marusic et al 2007). If errors in data analysis or management are rare, we can trust the results; if errors are common, systematic, or deliberate falsification, then any conclusions drawn from such work deserves retraction at the very least (cf. discussion in Baggerly and Coombs 2009; see StatCheck project, Nuijten et al 2016).

Bias in research

Bias is not limited to obvious profit-seeking or even defense of one's ideas or reputation. Bias can enter research in subtle ways. Prior to 2000, many post-menopausal women could expect to be placed on hormone replacement therapy (HRT) to manage post-menopausal symptoms and to reduce likelihood of osteoporosis later in life. Moreover, HRT was believed based on research to show protective benefits against heart disease (e.g., Nabulsi et al 1993; Grodstein et al 2000). However, other prospective studies that were designed to reduce many sources of bias in patient recruitment (Women's Health Initiative Study Group 1998), found the opposite: HRT may increase risk of heart disease as well as other diseases (Rossouw et al 2002). Since the 2002 Women's Health Initiative study report, the costs and benefits of HRT have been furiously debated (Lukes 2008).

The purpose of this chapter, however, is not to exhaustively review sources of bias. Instead, we leave you with a partial list of kinds of bias in clinical trials (from a review by Pannucci and Wilkins 2010).

- **Bias during study design**, e.g., use of subjective measures or poorly designed questionnaires.
- **Interviewer bias**, for example, if the interviewer is aware of the subject's condition
- **Chronology bias**, where the control subjects may not be observed in the same time frame as the treatment subjects
- **Citation (reporting) bias**, where negative results are not reported
- **Confounding**, where results may be due to factors not properly controlled that also affect the outcome. An example provided was link between income and health status, which would be confounded by access to health care.

These are but a few of the kinds of bias that even proper research may be influenced by. Solutions are to randomize, to double-blind, and to avoid reporting bias. Controls of sources of bias increase the validity of the research. How experimental design may control for bias is discussed further in [Chapter 5.3](#).

A case study

Stories we hear from the news or from our friends about health or the environment are anecdotal. A basic rule in critical thinking is to distinguish between argument that is built on anecdotal evidence and argument based on scientific evidence. Let's evaluate the following real-world example.

Airborne® is a leading dietary supplement. Here's what their website has to say about the product.

"Scientific research confirms that Airborne proprietary formulation with 13 vitamins and minerals plus a blend of health-promoting herbs, does indeed enhance immunity" (5 August 2013, <http://www.airbornehealth.com/our-story>).

This next statement appears in smaller text near the bottom of the page.

"These statements have not been evaluated by the Food and Drug Administration. These products are not intended to diagnose, treat, cure or prevent disease" (5 August 2013, <http://www.airbornehealth.com/our-story>).

So, in short, how do you evaluate Airborne? The claim is that "scientific research confirms that Airborne ... enhance immunity," and yet, there is a statement that follows that suggests that the claim has not been evaluated by the agency legally responsible in the USA for determining the efficacy of medicine and treatments. Are not these in contradiction? In the USA, supplement manufacturers need not seek FDA approval. In fact, the U.S. Congress, in its passing of the The Dietary Supplement Health and Education Act of 1994 (DSHEA), specified what needs to be reported to the FDA; statements of "enhancing immunity" are lawful, but do not require the supplement manufacture to show evidence that their product provides this support, only that ingredients in the supplement may do so.

Returning to the statements from Airborne Health, how may these statements be interpreted?

1. peer-reviewed articles testing Airborne are available
2. that experiments were conducted without bias
3. and that to "enhance immunity" is something one could measure

Let's look at #3 first. A search of PubMed (December 2013) for the phrase "enhance immunity" returned 9171 hits. I didn't look through the more than 9000 articles, but a quick look shows that the term is indeed used in research, but carries a wide range of interpretations. For one paper "enhance immune" response was the result of genetic modifications to T-cells and how they responded to a virus (PMID:24324159).

Adding "vitamin" to the search resulted in just 158 hits in PubMed; none of these are about Airborne (see claim #1 and #2). However, to be fair, go back and look at their statement — notice they are not saying *exactly* that anyone has studied Airborne (otherwise they'd probably list the papers), just that vitamins and minerals and herbs that are in Airborne have been studied.

Indeed, vitamins A, C, E and minerals like zinc and selenium that are in Airborne are among the 158 papers. To be additionally fair to Airborne, there are indeed many papers investigating benefits (harm) of supplemental vitamins and minerals — and again, being charitable here, what we can say is that results are at best inconsistent (Bjelakovic et al 2014; Comerford 2013).

The manufacturer of Airborne ran into some trouble with the FTC and ended up paying over \$20 million US dollars to settle the lawsuit. The one study, not peer-reviewed, of Airborne's effectiveness cited during these proceedings apparently was sponsored by Airborne's manufacturer. Airborne was acquired by Schiff Vitamins March 2012 ([Business Wire April 2 2012](#)). Click here for some [additional reading](#) about Airborne.

About peer-review

Evaluation of claims made, research reported, and ideas in science are generally expected to be subject to peer-review by knowledgeable persons in the field. Effective peer-review is an important “gaketeekeeper” in science (Siler et al 2015, Tennant 2018). Proper peer review gives legitimacy to research (Siler et al 2015). The traditional peer review system is a closed system, which includes elements controlled by journal editors and is based on reviewer identify remaining anonymous. The closed peer-review system is not the only option: many have argued for an open review process (e.g., Pöschl 2012, Ho et al 2013).

Does the Airborne story meet your definition of bias?

Airborne made health claims about their product; these claims were found by the federal government to be misleading but does not fit into our discussion of bias. Instead, the Airborne story fits more with critical thinking: I think it is rather a casual position to say that we should, and do, think critically of news sources or the stories our friends tell us about health.

But bias in science occurs too, and can be very difficult to control or completely eliminate.

Learning statistical reasoning will help develop critical thinking skills

In addition to considering sources of bias, i.e., can you trust the messenger and thinking about the sets of facts offered to bolster a claim, it helps to evaluate each fact to see if:

- The author is really talking about correlation and has not in fact provided evidence of causation
 - vs. experiment, which tests “cause” manipulation of X , does Y change?
- Measurement is appropriate
 - Direct or indirect
- Unmeasured important (co)variables exist

Questions

1. The xkcd comic in Figure 2.6.1 calls it survivorship bias. Provide a definition for this kind of bias and describe the effects on conclusions if it is not accounted for.
2. Doctors are encouraged to engage in “evidence based medicine,” that is, decisions about care should be based on data. How may confirmatory bias (from the patient's point of view, from the doctor's point of view) influence a discussion of “facts” when deciding on a course of medication therapy — for example, between a well-established medication with many known side-effects and a new, recently approved medicine reported to be as effective as the standard treatment?

This page titled [2.6: Statistical reasoning](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.