

Shutterstock

8 ways to be a more savvy science reader

Your guide on how to evaluate scientific evidence.

By Susannah Locke | @susannahlocke | May 20, 2015, 2:09pm EDT

How can you tell if scientific evidence is strong or weak?

The world abounds with evidence and studies, some of it good and some of it poor. How can you know what to trust? This card stack aims to guide you through tricky issues that can cloud your understanding of scientific findings.

One major problem is that scientific lingo often means something different in common parlance. And these words can insidiously sneak into media coverage. Simple words such as **theory**, **significant**, and **control** have totally different meanings in the realm of science.

Another problem is that there's no such thing as a perfect study. Experiments can suffer from issues in how they're **designed**, how they're **analyzed**— even how they're reviewed by **scientific journals**.

Read on for eight key tips that will let you more confidently assess the results from scientific and medical studies.

Know the difference between a hypothesis and a theory

Scientists and non-scientists often use these two words in very different ways. Let's unpack them:

Hypothesis

In science, a hypothesis is a proposed explanation that can be tested through further experiments and observations. It's a still-unproven idea that requires the collection of more data in order to confirm or reject it.

A hypothesis is often considered the first step of the scientific method. Many lay people often use the word "theory" here, but that's not how scientists use it.

Origin of HYPOTHESIS

Greek, from *hypotithenai* to put under, suppose, from *hypo-* + *tithenai* to put — more at [DO](#)

First Known Use: circa 1656

(Merriam-Webster)

Theory

In science, a theory is a widely accepted idea that has some serious data behind it. When scientists refer to the "theory of evolution" or the "theory of relativity," they're not saying that this is just some wild, unsubstantiated idea. Both of these concepts are backed by lots of data, observations, and experiments.

Of course, established scientific theories can later be changed or rejected if there's enough data to warrant that. But the reason theories became so broadly accepted in the

first place is because they are supported by a substantial body of evidence.

Watch out for selection bias

If a psychologist (say) could run a single test on every person in the world, that would lead to some powerful results. But that's not practical. So scientists do the next best thing: they select a smaller group to study.

However, they always have to be careful about the particular group they are selecting. Studies can suffer from selection bias, in which the chosen subset isn't random enough and therefore somehow biased in favor of a certain outcome of the study.

Selection bias can happen in a number of ways. Perhaps certain types of people are more likely to want to be involved — or are committed enough to not quit, say, a longer multi-year experiment.

Consider a yearlong study of a weight-loss drug in which half the participants dropped out of the study before it was over. The ones who stayed in the game might have all lost weight, but it's also important to consider those who quit. Maybe those people were seeing no progress. So an apparent success rate of 100 percent was actually more like 50 percent.

Another issue to think about is whether the people participating are representative enough of the group that the paper or article is talking about.

This is why, for example, polling data from a **nationally representative sample** of people (such as the ones conducted by Pew and Gallup) can be more informative about national opinion than an informal poll that's open to anyone on the internet, even if far more people participated in the latter.

Another scenario that's a common problem with psychology studies is that they tend to primarily enroll American undergraduates because they are easy to recruit on college campuses. But undergrads aren't necessarily representative of the average American.

Similarly, studies in the US or Europe that look at people from "WEIRD" countries (that is, Westernized, Educated, Industrialized, Rich, and Democratic) might not apply to people

from other cultures. For more, check out Bethany Brookshire's great **piece** about psychology's WEIRD problem at Slate.

Don't confuse correlation and causation

Often scientists will find that two different variables are correlated — for example, they both increase together over time. That's a hint that they might be related, but it doesn't necessarily mean that one is causing the other. Perhaps it's just a coincidence. Or perhaps a third variable is causing both of the other two. Further testing is typically required.

Over time, lots of correlative evidence — combined with systematically ruling out other possible causes — can lead to a stronger case that something is causing something else. But the best way to show causation is to perform a carefully controlled experiment.

THE BEST WAY TO SHOW CAUSATION IS THROUGH A CONTROLLED EXPERIMENT

Here's an example: One **study** found that doctors were more likely to overprescribe antibiotics in the afternoon. That is, there was a correlation between antibiotics prescriptions and the time of day. The authors guessed that a phenomenon called decision fatigue could be the cause — that the brain gets tired after making too many decisions. But other possible causes could be too little sugar in the body (glucose fatigue) or general fatigue.

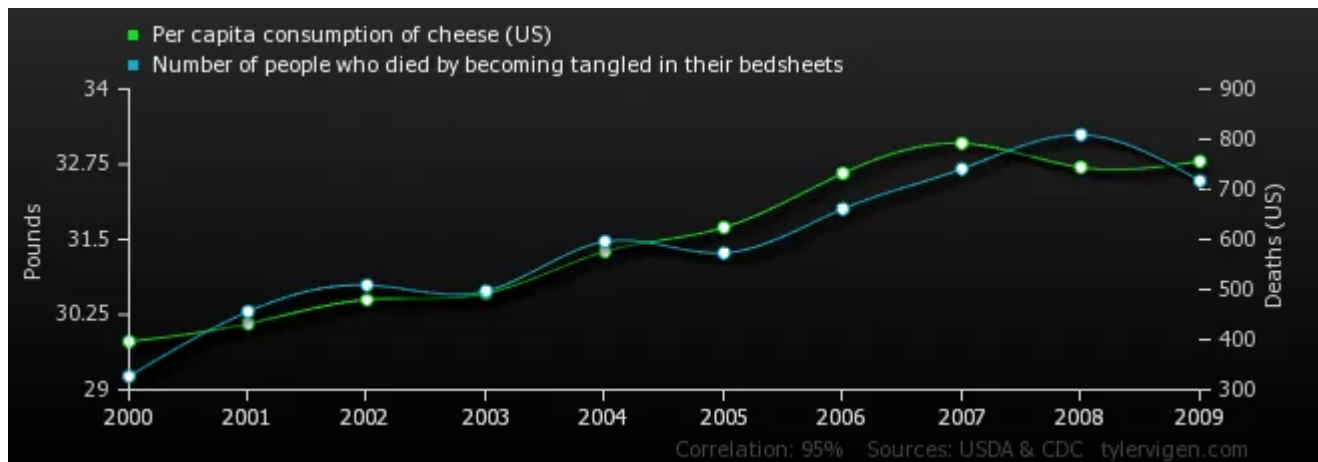
In order to find out if decision fatigue is a cause, they would need to set up an experiment in which they randomly had some doctors make more decisions than others. And to make the experiment as controlled as possible, those making fewer decisions would have to do some other kind of mentally fatiguing task, instead.

What does it mean that an experiment is controlled? In science, a control group is a group used for comparison. Control groups in medical studies often receive a placebo — a fake medicine, device, or procedure.

For example, many health problems will naturally get better (or worse) on their own. If you didn't also have a control group, you might think that, for example, you've invented a cure

for the common cold. But the truth is that the common cold gets better in a week or two anyway.

It's also worth watching out for odd correlations. If a correlation seems really weird or too good to be true, then it's possible that there's nothing meaningful behind it. Tyler Vigen creates a bunch of what he calls **spurious correlations** from real data, like this one between per capita cheese consumption and people who have died by getting tangled in their bed sheets:



(tylervigen.com)

One fun correlation that people like to cite is the decrease in the number of **pirates** worldwide and the increase in global temperatures. However, it's highly unlikely that the loss of pirates has increased temperatures or that increasing temperatures has killed pirates off. It's also unlikely that some underlying cause is affecting both. There's a correlation, but it's meaningless.

Look for the gold standard: double-blind, placebo-controlled, randomized tests

The most reliable type of study — especially for clinical trials — is generally thought to be the randomized, placebo-controlled, double-blind study.

If you are looking at a clinical trial, a psychology study, or an animal study, and it hasn't been designed like this — and there isn't a good reason that it couldn't have been — then you might want to question the results.

Let's break down this terminology:

1) Randomized: This means that the participants in the study were randomly placed into the experimental group and the comparison group. This is important because if people get to choose, they might be more likely to pick one or the other because of some unexpected factor.

As a hypothetical example, maybe people who are more optimistic are more likely to want to try a new drug for anxiety rather than an old drug that's being used for comparison. And maybe optimism is linked to better outcomes for generalized anxiety disorder. The researchers could end up thinking that these people got better because of the drug when it was actually because they were innately going to do better anyway.

Similar problems can be introduced if researchers choose who goes into which category. That's why random is best.

2) Placebo-controlled: A controlled study has an appropriate comparison group, also called a control group. In medical studies, one comparison group usually gets a placebo — a fake intervention such as a sugar pill. This is in order to distinguish what the drug actually did from what a participant's psychological expectations did. (Placebo effects can be surprisingly strong — so strong that they can oftentimes relieve pain, among other health problems. And they've been getting stronger in recent decades, according to Steve Silberman's in-depth placebo **story** from Wired.)

A good placebo group should be as similar to the experimental group as possible. So, for example, if you were testing out a drug that's a large, red pill, you'd ideally want to give the people in your comparison placebo group a large, red pill that's the same in every way, but doesn't contain the drug. (Yes, even a pill's color and size can have a placebo effect.) Some studies go as far as to do **sham surgeries**, including anesthesia, incisions, stitches — the works.

3) Double-blind: A study is "blind" if the participants don't know whether they are in the experimental group or the control group. For example, you don't want someone knowing if she's received a real drug or a fake drug because her expectations could change the outcome of the study.

A study is "double-blind" if the researchers in personal contact with participants *also* don't know which treatment they are administering. You don't want the nurse giving out pills to

know if they're real or not because then subtle differences in his behavior could influence patients — and therefore the results.

Understand "significance"

In everyday language, significant means that something is important or large. But a scientific finding that's considered "statistically significant" isn't necessarily either of those things. Scientists generally say that something is statistically significant if the effect can be picked up with a particular statistical tool called a **p-value**.

What's considered a good p-value is arbitrary and can vary somewhat between scientific fields. Often the cut-off for what's considered "statistically significant" is a p-value of 0.05.

It's important to keep in mind that p-values aren't the only relevant numbers in a study. For example, a treatment for a disease could have a statistically significant effect of changing the survival rate from 43 to 44 percent. That's a tiny change that probably isn't all that meaningful for how the disease will be treated in the future.

In fact, some people think that scientific papers should **do away** with p-values altogether and instead clearly and prominently show both the size of the effect and the range of the effect, both of which can be exceptionally important.

Another hazard: if you run a study many, many times or do a whole bunch of different statistical analyses on the same data, you could end up with results that look meaningful purely by chance. And then publishing only those meaningful-looking results would be likely to make the public draw misleading conclusions about your research. For more, Charles Seife has a good **overview** of various p-value pitfalls up at Scientific American.

Be aware of conflicts of interest

Conflicts of interest can come in many forms. The one that's generally most of concern within science and medical publishing these days is financial.

For example, this could be someone who received funding from a company that has a vested interest in the outcome of her own study. Or maybe that person has a relationship

with the company — such as sitting on its board or acting as an unpaid consultant — that could lead to benefits in the future.

For example, one type of conflict would be a council that promotes a certain type of food then funding a study about that food's health benefits. Another would be a researcher who accepted travel money for a conference from a drug company and also researches that company's drugs or a competing company's drugs.

A recent **analysis** found that from 7 to 32 percent of randomized trials in top medical journals were completely funded by medical industry sources. And that's just those with full, direct funding. Presumably, the percent with any type of conflict of interest would be far higher.

One solution might be to ban such conflicts of interest. But what many journals have opted for instead are various requirements about disclosure, such as **this one** from the journal *Science*, which instructs people submitting papers to reveal "any affiliations, funding sources, or financial holdings that might raise questions about possible sources of bias." (The actual **form** that authors fill out is even more detailed.)

The editor then determines which relationships should be printed publicly with the scientific paper. And then whoever's reading the paper can draw their own conclusions about whether those relationships have — knowingly or unknowingly — influenced the data.

What, if anything, needs to be disclosed is up to the journal (and in some cases people's **employers**, too). Many journals publish conflict-of-interest policies on their websites. And if you scrutinize a paper closely, you may find some of these disclosures, too.

Know that peer review isn't perfect

Peer review is the system in which a couple of independent experts read over a paper that's been submitted to a journal. Generally, a journal isn't considered high quality if the papers aren't peer reviewed.

It's usually the case that reviewers are chosen by the journal and kept anonymous so that the review can be as impartial as possible. These people can recommend revisions to the text, new experiments that should be added, or even that the journal shouldn't publish the

paper. Then, the paper's authors will generally look at those reviews and incorporate them into a revised paper, if necessary.

But reviewers *aren't* asked to do everything within their power to make sure that the results are absolutely correct. (That would simply take too much time and be impractical. A paper can sometimes take years to put together already.) For example, reviewers are not expected to try the experiments themselves. And they don't generally look at raw data or re-run calculations.

They *do* **look** at the manuscript to see if the experiments were properly designed, if the data supports the paper's conclusions, and if the findings seem important enough to warrant publication.

So, peer review is generally beneficial, but not perfect. The scientific process really isn't complete until someone else replicates what's in the paper. And that's something that happens (if at all) after publication, not before.

In addition, sometimes papers get retracted. It's rare, but it does happen. Ivan Oransky and Adam Marcus's blog **Retraction Watch** is a great place to hear stories of the biggest, most important, and most dramatic retractions. (And they can be quite dramatic. In 2014, the *Journal of Vibration and Control*, for example, **retracted** 60 papers all at once.)

There are a few odd exceptions to the general peer review process, including the journal the *Proceedings of the National Academy of Sciences*, which allows members of its exceptionally prestigious academy to choose their own reviewers for up to four papers a year. Peter Aldhous has a good **story** about this controversial process over at *Nature*. (*PNAS* also accepts many papers through a more traditional peer-review system.)

Realize that not all journals are good

Just because it's in a journal doesn't mean it's a fantastic study. Journals and papers both range from great to meh to downright fraudulent. And even a top-notch journal can sometimes publish a flawed study.

The most commonly used metric to assess a scientific journal's influence is the Impact Factor. The Impact Factor is essentially a measure of popularity. It counts the number of

times a journal's papers have been mentioned in other papers, relative to the journal's own volume of article output.

The more of these citations that appear, the more influence the journal seems to have on people's work. (Specifically, the IF is calculated from citations in the Thomson Reuters **Journal Citation Reports** database.)

How do you find a journal's Impact Factor? If you belong to a good library, some will have a subscription that can get you into the **Journal Citation Reports** analysis that comes out each year. If not, many journals and journal publishers will proudly list their rating somewhere on their websites. Just search for "impact factor." For comparison's sake, some of the most prestigious journals around, such as *Science*, *Nature*, and *JAMA*, have Impact Factors in the high-20s to mid-30s. (And the *New England Journal of Medicine* has an astounding Impact Factor in the 50s.)

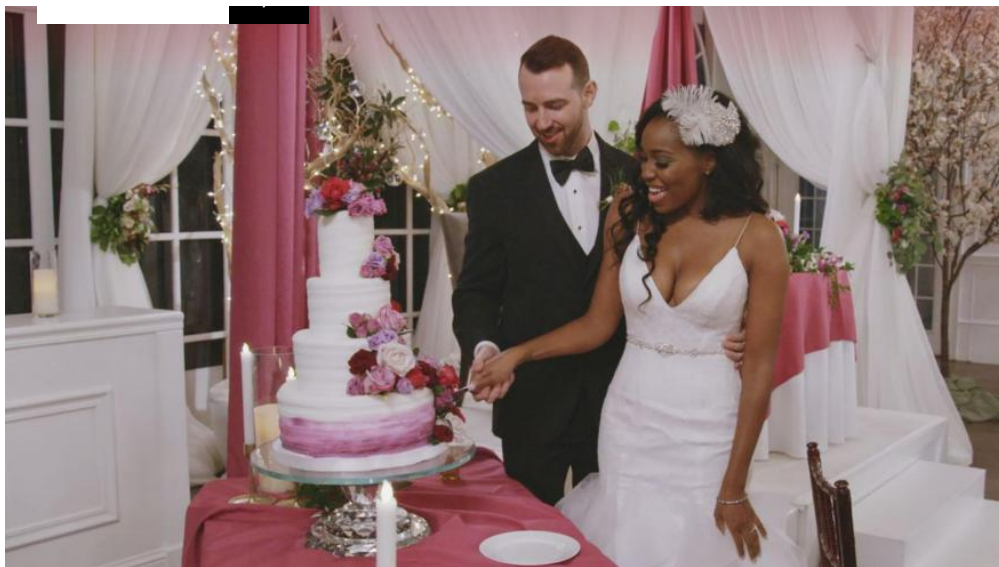
The Impact Factor is **controversial**. It's a handy tool, but it's not the only way to look at things.

Some fields of science naturally generate more citations than others, but that doesn't necessarily mean that they're really better or more influential. And at least one **study** has found that Impact Factors don't correlate well with expert opinion.

Another thing to look out for are predatory, for-profit journals that will publish just about anything (and without **peer review**). In recent **sting operations**, several people have gotten such journals interested in publishing flawed or incoherent papers.

Also keep in mind if the study is in an appropriate journal for its subject matter. Sometimes junk science can end up in a peer-reviewed journal, especially if it's outside the journal's area of expertise. Its reviewers and editors might be less able to accurately assess the paper's quality.

Sarah Fecht has a good story over at Popular Mechanics about how **bad science** ends up being published — and then covered by the media as if it were good.



Who got married in Love Is Blind's volatile season finale?

[READ MORE >>](#)

SCIENCE & HEALTH

The coronavirus diagnostic testing snafu, explained

ENERGY & ENVIRONMENT

Oregon Republicans are subverting democracy by running away. Again.

POLITICS & POLICY

What David Brooks gets wrong about Bernie Sanders



[View all stories in The Latest](#)