How to Know Whether to Believe a Health Study

AUG. 17, 2015

Photo



CreditPete Ryan



**Austin Frakt**

[**THE NEW HEALTH CARE**](http://www.nytimes.com/column/the-new-health-care)

Every day, new health care research findings are reported. Many of them suggest that if we do something — drink more coffee, take this drug, get that surgery or put in this policy — we will have better (or worse) health, or longer (or shorter) lives.

And every time you read such news, you are undoubtedly left asking: Should I believe this? Often the answer is no, but we may not know how to distinguish the research duds from the results we should heed.

Unfortunately, there’s no substitute for careful examination of studies by experts. Yet, if you’re not an expert, you can do a few simple things to become a more savvy consumer of research. First, if the study examined the effects of a therapy only on animals or in a test tube, we have very limited insight into how it will actually work in humans. You should take any claims about effects on people with more than a grain of salt. Next, for studies involving humans, ask yourself: What method did the researchers use? How similar am I to the people it examined?

Sure, [there are many](http://www.healthnewsreview.org/toolkit/tips-for-understanding-studies/) other [important questions](http://cdn.journalism.cuny.edu/blogs.dir/422/files/2012/04/Covering-Medical-Research.pdf) to [ask about a study](http://www.minnpost.com/second-opinion/2015/05/confused-drumbeat-health-news-dreck-minnesota-based-website-aims-help) — for instance, did it examine harms as well as benefits? But just assessing the basis for what researchers call “causal claims” — X leads to or causes Y — and how similar you are to study subjects will go a long way toward unlocking its credibility and relevance to you.

Let’s look closer at how to find answers. (If the answers are not in news media reports, which they should be, you’ll have to chase down the study — and admittedly that’s not easy. Many are not available without cost on the web.

It’s instructive to consider an ideal, but impossible, study. An ideal study of a drug would make two identical copies of you, both of which experience exactly the same thing for all time, with one exception: Only one copy of you gets the drug. Comparing what happens to the two yous would tell us the causal consequences of that drug for you.

Clearly, there are a few complications in the real world. We only have one of you to play with. Also, you don’t participate in most studies, if any. The people researchers examine are never *exactly* like you. So how do we extract some value from the imperfect?

Researchers employ various methods to infer what would happen to people who might be like you in two different circumstances, such as taking or not taking a drug. The most widely trusted approach is the randomized controlled trial. In the most basic randomized trial, individuals are randomly assigned to treatment (e.g., they get the new drug) and control (e.g., they get a placebo or nothing).

This random assignment is powerful. If done with enough people, it causes the two groups to be statistically identical to each other except for the experience of the treatment (or not). Whatever changes are observed can usually be attributed to that treatment with a good degree of confidence.

Though a randomized trial makes two groups statistically identical to each other — apart from treatment received — it still doesn’t mean either group is identical *to you*. If the individuals selected to participate in the trial happen to be very similar to you — similar ages, income, living environment and so forth — that increases the chances that the results would apply to you. But if you’re, say, a 65-year-old, middle-class New Yorker, a study whose subjects were poor 30-somethings in rural China may not translate to your experience.

This is one of the chief [limitations](http://theincidentaleconomist.com/wordpress/methods-angus-deaton-on-everything/) of randomized trials. They’re typically focused on narrow populations that meet strict criteria — [those most likely to benefit from treatment](http://www.nytimes.com/2015/02/03/upshot/how-to-measure-a-medical-treatments-potential-for-harm.html?rref=upshot&abt=0002&abg=0). Many drug trials [exclude older patients or children](http://theincidentaleconomist.com/wordpress/limitations-of-randomized-trials/) because of ethical or safety concerns. Many, particularly much earlier trials, [didn’t include women](http://circoutcomes.ahajournals.org/content/early/2010/02/16/CIRCOUTCOMES.110.868307.full.pdf). We know a lot less about how drugs affect groups who weren’t studied than we might like. Harm could even come if it was assumed that findings from those who were studied applied to people who weren’t.

My colleague Aaron Carroll [provided an example](http://www.nytimes.com/2014/06/03/upshot/calling-an-ordinary-health-problem-a-disease-leads-to-bigger-problems.html?_r=1) of just this problem. Based on the results of randomized trials that included only adults, prescriptions of drugs known as proton pump inhibitors to infants with[gastroesophageal reflux disease](http://health.nytimes.com/health/guides/disease/gastroesophageal-reflux-disease/overview.html?inline=nyt-classifier) grew sevenfold between 2000 and 2004. Only later, in 2009, a direct study of infants found that those drugs caused them harm, with no benefit.

A type of study other than a randomized trial is less likely to have this kind of problem. Rather than recruiting and randomizing a narrow set of patients to generate new data, researchers can turn to “nonexperimental” or “observational” database studies. These database studies use large data sets, like those available from [Medicare](http://topics.nytimes.com/top/news/health/diseasesconditionsandhealthtopics/medicare/index.html?inline=nyt-classifier), [Medicaid](http://topics.nytimes.com/top/news/health/diseasesconditionsandhealthtopics/medicaid/index.html?inline=nyt-classifier), the Veterans Health Administration or very large surveys. Some studies of this kind are large enough to allow researchers to report differences in treatment effects across groups. Perhaps women respond differently, for example.

And because they don’t have to generate new data, nonexperimental studies are typically cheaper than randomized trials and produce results more quickly.

People like you are more likely to be represented in a nonexperimental database study, so your top concern might be whether the findings are valid. After all, such a study doesn’t rely on the clean comparisons of randomized groups of people. Instead, it often compares groups of people who could have self-selected into receiving treatment or not. Maybe those who opted to receive it are [systematically different](http://www.nytimes.com/2015/04/07/upshot/alcoholics-anonymous-and-the-challenge-of-evidence-based-medicine.html?smid=tw-share&abt=0002&abg=1&_r=1) — healthier, sicker, more careful, for example — and that’s what drives the findings. If so, what might appear causal isn’t, giving rise to the familiar “correlation does not imply causation.”

That concern is why researchers [employ techniques](http://www.nytimes.com/2015/04/07/upshot/alcoholics-anonymous-and-the-challenge-of-evidence-based-medicine.html?smid=tw-share&abt=0002&abg=1) to try to adjust for differences across comparison groups in nonexperimental studies. These can get complex in a hurry, and few news media reports could describe them in detail. But that doesn’t mean they’re all sketchy or all ironclad. The key fact is that they all rely on different assumptions than a randomized trial, and [those assumptions can and should be probed](http://jama.jamanetwork.com/article.aspx?articleid=2203787) to gain confidence in causal inferences.

Most news media reports acknowledge when a study is nonexperimental, and sometimes you can find a sentence or two about how the researchers sought to adjust for differences and tested assumptions. You should also look for statements from experts about whether those adjustments and tests were sufficient. However, these rely on judgment. There is always room for doubt.

Ultimately, no single study is perfect. Whether it’s a randomized trial or a nonexperimental one, one can never be absolutely sure study findings are valid and applicable to you. The best bet is to wait, if you can, until evidence accumulates from many studies using a range of methods and applied to different populations.

Few things are miracle cures, but when one shows up, we’ll see its signature in not just one study, but in many. Yes, that can take time. But if you want solid evidence you can count on, you cannot also be impatient.