

APRIL 25, 2013, 8:00 PM

What Do Scientific Studies Show?

By **GARY GUTTING**

As any regular reader of news will know, popular media report “scientific results” nearly every day. They come delivered in news reports and opinion pieces, and are often used to make a variety of points concerning important matters like health, parenting, education, even spirituality and self-knowledge. How seriously should we take them?

For example, since at least 2004, we have been reading about [studies](#) showing that “vitamin D may prevent arthritis.” A [2010 Johns Hopkins Health Alert](#) announced, “During the past decade, there’s been an explosion of research suggesting that vitamin D plays a significant role in joint health and that low levels may be a risk factor for rheumatologic conditions such as rheumatoid arthritis and osteoarthritis.” However, in February 2013, a more rigorous [study](#) called the previous studies into serious question. Similarly, despite many studies suggesting that taking niacin to increase “good cholesterol” would decrease heart attacks, a [more rigorous study](#) showed the niacin to have no effect.

Such reports have led many readers to question the reliability of science. And given the way the news is often reported, they seem to have a point. What use are scientific results if they are so frequently reversed? But the problem is typically not with the science but with the reporting.

In both the above examples, earlier studies had shown a correlation but not a causal connection. They had not shown that, for example, taking vitamin D was the only relevant difference between those whose pain decreased and those whose pain did not decrease. Perhaps, for example, those taking vitamin D also exercised more, and this was the cause of the pain decrease. Typically, the best way to establish a cause rather than a correlation is to perform a randomized controlled experiment (R.C.T.), where we know that only one possibly relevant factor distinguishes the two groups. In both the vitamin D and the niacin cases, there was an R.C.T. that showed that the earlier results had been merely correlations.

R.C.T.s are often very difficult to set up properly and can take many years to carry out. As a result, most research we read about involves just correlational studies. John Ioannidis, in a series of highly regarded analyses, has shown that, in published medical research, 80 percent of non-randomized studies (by far the most common) are later found to be wrong. Even 25 percent of randomized studies and 15 percent of large randomized studies — the best of the best — turn out to be inadequate. (For details, see Ioannidis’s [seminal paper](#), “Why Most

Published Research Findings Are False,” and David H. Freedman’s [Atlantic article](#) on Ioannidis’s work.)

Why, then, do scientists even bother with correlational studies, most of which they know will turn out to be wrong? One reason is that such studies are excellent starting points for deciding which hypotheses to evaluate with the more rigorous R.C.T.s. (Correlational studies are also important in [a number of other ways](#).) Contrary to what many non-scientists seem to believe, the key feature of empirical testing is not that it’s infallible but that it’s self-correcting. As the physicist John Wheeler said, “Our whole problem is to make mistakes as fast possible.” Indeed, Karl Popper built an illuminating philosophy of science on the idea that science progresses precisely by trying as hard as it can to falsify its hypotheses.

The trouble with much science reporting is that it does not do enough to ensure that the public can tell just how significant a scientific result is. The better reports will implicitly hedge results that are merely correlational, saying, for example, that vitamin D “may” decrease arthritis pain or that niacin “can” prevent heart attacks. But they seldom explain how preliminary and unreliable most correlational studies are. They don’t explain the specific limited role such studies usually play in the overall scientific process.

There’s another crucial limitation that science reporting — especially in psychology and the social sciences — often ignores. Even when we have R.C.T.s that decisively establish a scientific law, it doesn’t follow that we can appeal to this result to guide practical decisions. As [Nancy Cartwright](#), a prominent philosopher of science, has recently emphasized, the very best randomized controlled test in itself establishes only that a cause has a certain effect in a particular kind of situation. For example, a feather and a lead ball dropped from the same height will reach the ground at the same time — but only if there is no air resistance. Typically, scientific laws allow us to predict a specific behavior only under certain conditions. If those conditions don’t hold, the law doesn’t tell us what will happen.

In dealing with the natural world, we are often in a position to establish conditions that are sufficiently close to those that make a law relevant. In the human (and, especially the social) world the high degree of complexity and interconnectedness makes this extremely hard to do. A method of teaching fifth-grade math that has been rigorously shown to be highly effective for the students and teachers in one school district may well not work for the students and teachers in another. As Cartwright puts it, all a randomized controlled test tells us is that “this works here.” It is another — and often very difficult — matter to conclude that “this will work there.”

It follows then that even when we have reliable results from “pure science,” we need engineers who can tell us whether and how these results apply to the situations we are dealing with. For the natural sciences (physics, chemistry, biology) we have well-established methods of engineering. But the engineering equivalent for the human world is, with few exceptions, still a long way off. Reporting of “breakthroughs” in the human sciences needs to make clear the gap between science and application.

Media tend to present almost any scientific result they report as valuable for guiding our lives, with the entire series of reports accumulating a vast body of practical knowledge. In fact, most scientific results are of no immediate practical value; they merely move us one small step closer to a final result that may be truly useful. Too many news reports present experimental results as providing good advice on which we can reliably act. In most cases those results would be better viewed as mistakes pointing to a next step that will be a bit less mistaken.

Science reporting would be much improved if we had a labeling system that made clear a given study's place in the scientific process. Is it merely a preliminary result (a small-scale heuristic study meant to suggest a hypothesis that will itself require many stages of further testing before we have a reliable conclusion)? Is it a larger-scale observational study (showing a correlation but by no means establishing a causal connection)? Is it a large-sample randomized controlled test (establishing a causal connection, given specific conditions)? Or, finally, is it a well-established scientific law that we know how to apply in a wide range of conditions?

Of course, the above categories are just an outsider's rough suggestions. The various scientific disciplines (through their governing organizations) should set professional labeling standards for material discussed in popular media. Some such system is essential because many if not most people who read popular reports of scientific work are looking for results on which they can rely in making practical decisions about personal life, work or public policy.

Unfortunately, such results are far less common than the many highly fallible preliminary studies that contribute to the complex process leading to reliable results. Media reports saying "studies show . . ." are most often giving us highly tentative results — indeed, results that are likely to be false. They need to be labeled as such.

Gary Gutting is a professor of philosophy at the University of Notre Dame, and an editor of Notre Dame Philosophical Reviews. He is the author of, most recently, "Thinking the Impossible: French Philosophy since 1960," and writes regularly for The Stone. He was recently interviewed in 3am magazine.