

Notes from the Science Lab

Some thoughts for reporters, editors, copy and news desks on evaluating medical, scientific and environmental studies.

Source: The SMASH Desk of the Philadelphia Inquirer

Not all studies are definitive. In fact, very few are. Science is most often a continuum, in which one study builds on the last and some clear proof emerges only over time.

Many studies are based on numbers so small or evidence so flimsy that they raise the question as to whether we should publish them at all. At the very least, the reporter and editor must feel comfortable that:

- 1). The research is sufficiently strong to merit publication.
- 2). The expert comments are not all simply self-serving; outside (more neutral) experts are also quoted.
- 3). He/she understands how the researchers came to their conclusions: any percentages in the story should be backed by actual numbers (known as the two-number rule). Something can double or triple and still be a meaningless increase.
- 4). He/she understands the context of the study—does it contradict or support previous findings? If so, is that explained?
- 5). Most important, if the story raises alarm, that we go some distance to make sure that the alarm is justified and that we are giving readers information as to what they should do if they are, in fact, alarmed.
- 6). Finally, does anyone really care? Is it important? Or is this just more minutia of science that's not worth writing about at this point?

Here are some clues to help determine how strong or newsworthy a study is.

1. How to judge the sources?
 - a. National Academy of Sciences, EPA, CDC and other government agency long-term studies. These are usually multi-year studies done by panels of experts put together by the agency and then reviewed by a second independent panel of experts. These studies tend to be conservative in their findings, so you can usually be sure they are not making wild claims.
 - b. Peer-reviewed journals. If the study appears in the New England Journal of Medicine, JAMA, Science, Nature or a comparable magazine, it means that it was reviewed and approved by a panel of experts. This is not a fail-safe mechanism, but marks the most creditable work put out in the various disciplines. One note of caution: not all specialty journals are peer-reviewed.
 - c. Conference papers. Scientists and doctors deliver papers at conferences which generate news. These are not peer-reviewed but usually conference organizers have asked the individual to give the paper because he or she is a recognized expert in the field. That give some backing but not much.

- d. Press conferences. Science by press conference is to be viewed warily. Cold fusion is a perfect example of bogus science by press conference. No story should be run on press conference science without additional reporting, i.e., contacting credible scientists in the field for their reaction. If that cannot be done because of either time or staff constraints, consider spiking the story.

2. What kind of study is it?

- a. Controlled clinical studies. This is the best stuff. In this type of study, one test group gets a particular treatment and another control group does not. The groups are supposed to be as nearly identical as possible (same disease, same backgrounds, same ages). If there is a pronounced effect (good or bad)—eureka—it's a story. But sometimes the results aren't completely clear (see No.3, evaluating a study). Usually such trials are done in three phases, under FDA guidelines:

Phase I: small, usually no more than a handful of people, designed strictly to look at whether the drug will cause serious complications, not whether it works.

Phase II: A larger version of Phase I, again designed to look at safety, but also looking at efficacy.

Phase III: Considered the definitive look at whether a new drug has merit, involving thousands of people at a number of hospitals around the country. It's usually the last step before a company asks the FDA for approval. Often the phases blur, with reports of earlier phases coming out while later phases are already in progress

- b. Epidemiological studies. These are studies looking for potential disease-causing or contributing variables and make up a big portion of the health stories we do.

The problem with epidemiological studies is that you can't, except in a few instances, do clinically controlled experiments, since they would knowingly expose people to a potentially dangerous element. For example, if you wanted to judge how dangerous radon is, you couldn't expose people to varying amounts of radon to see how many get cancer.

In an effort to get around this, epidemiologists have developed the following types of studies:

Cohort Studies. A cohort study begins with a group of people who do not have the disease and follows them, measuring suspected characteristics and health performance. For example, in one lead-poisoning survey, 516 children were followed from birth to age 7. Periodic blood tests were taken to measure lead exposure and these were correlated with IQ test performance. The so-called "doctors" study, which has followed Harvard doctors for decades, found that aspirin helps prevent heart attacks but lots of fish does not. Cohort studies take a lot of time and are very expensive. **Cohort studies are the best epidemiological studies.**

Controlled Studies. Epidemiologist survey a group of people with a disease (cases) and a group without the disease (controls). They try to keep age, sex, socio-economic status equal. The researchers try to find some key element in the histories of the two groups that might explain the disease. Case-control studies depend on recollections and when dealing with mortality, of a lot of second-hand information. **Case-control studies are weaker than cohort studies.**

Cross-sectional studies. Rather than dealing with individuals, these studies look at groups—one neighborhood vs. another neighborhood, blacks vs. whites, etc. These studies are very crude. They are used to indicate areas that might bear more detailed study. That's it. **Cross-sectional studies are the weakest of all epidemiological studies.**

- a. Meta-analysis. Some studies aren't original research but rather a review of all previous studies on the topic. Researchers draw on all the findings and try to reach some overall conclusion. Meta-analyses are controversial because they attempt to draw comparisons between studies of different designs. A story based on a meta-analysis should never be cast as a brand-new finding. For instance, we shouldn't say, "New study shows Vitamin C prevents colds." If it's a meta-analysis, we should say, "A review of all studies on Vitamin C shows it prevents colds."
- b. Other science studies. Scientific research in other disciplines—biology, ecology, chemistry, etc.—tend to deal with testing a single hypothesis in a single, limited problem. By themselves, they aren't news. What makes them news is the context and implications of the study. One has to be very careful of these extrapolations. A story should contain 1) the researcher's own assessment of the implications and 2) the view of some other scientist in the field. If it does not have this, consider spiking.

3. Evaluating a study.

Key elements in assessing the value of a study and its newsworthiness are:

- a. Size of the study. How many people, plants or widgets were studied? The more, the better. A small study is usually less newsworthy. A large increase in a small population is not as statistically significant as a large increase in a large population. In a small group, there's a greater chance of the results being simply a fluke—like having a coin come up heads 5 times in a row.
- b. Length of the study. How long did it observe subjects and collect data? The longer the better. This varies with the problem in question. A short-term study on air quality and asthma—looking at a few months—might be okay, because the cause-effect is believed to be immediate. A short-term cancer study is nearly worthless, since the latency period in cancer is measured in years.
- c. Quality of the researchers. When top researchers and top institutions offer research, their work is worth looking at.
- d. Do the findings confirm or conflict with previous studies? If the study advances our knowledge by building on earlier studies, this is easy. If, however, it conflicts with previous studies, that should send up caution flags. We are not, however, in a position to know which trend is correct. We should proceed cautiously and give the reader context.
- e. The findings. We ought to subscribe to the "two number rule," which forbids using an isolated number in a story, particularly a percentage. For example, frog slime increased the risk of skin cancer 100 percent. Sounds like a story—until you see that the risk rises from one case of skin cancer in 10 million to two cases in 10 million. Or how about, half of the people in California tested for dumb disease were found to have it. But actually, only four people were tested, two of whom had the disease, in a state of roughly 30 million. Also, rates should be expressed as clearly as possible. Crude rates, such as one per thousand deaths in the general population, are weaker than rates that are age-specific, sex-specific or disease-specific.

If you're still interested...

This is the most important element and the one that is consistently butchered by the press. Here are the key parameters that should be understood by the reporter and noted in the story, if possible.

- a. Confidence intervals: this is a measure of how confident the researchers are in their risk numbers. If the confidence interval is big, that means the numbers are not very good. Statisticians like to express their confidence in their numbers as a percentage, i.e., they are 95 percent confident that the study result is between X and U. The trick is knowing what X and U are. For example, researchers estimated that chlorinated water was associated with a 38 percent increase in bladder cancer. Sounds like a good story? That confidence interval was 1.01 to 1.87, a wide spread. That meant that while the researchers thought it was 38 percent, it could be anywhere from 1 percent to 87 percent. Obviously, this is a highly speculative study and not worth front-page news.
- b. Size of effect. This is particularly important in epidemiological studies. How much effect does the suspected agent cause? These are expressed in numbers such as .05 or 1 or 2.1. The number 1 represents a 100 percent increase in effect. That may sound big, but statisticians might still consider that insignificant. For example, in the overall population of women, there are 10 cases per 1,000 among women who smoke. That is an effect of 1. Generally, epidemiologists like to see an effect of 3 to find significance. Sometimes 2 is okay. But usually, anything less than 1 is considered suspect, considering the weakness of the studies in the first place.

In some studies this is expressed as a probability, i.e., how probable was it that the result was a fluke? It is expressed as a number from 0 to 1. If the P value is zero, then there is no chance that it was a fluke. A P value of more than 0.05 is considered weak.

While some stories may merit some specific confidence numbers (as in political polls), more often we should give the reader a clear understanding of just how weak or strong the study is i.e., was this "preliminary" with more testing needed; was a mild "association" strong evidence, or whatever.