# Are most published research findings actually false? The case for reform

James J. Asher, Ph.D.
Emeritus Professor of Psychology
San Jose State University
San Jose, California
tpr-world@aol.com
www.tpr-world.com
July 2007

The title of this article was used in a recent essay by epidemiologist, John Ioannidis from a 2005 article in the Journal of the American Medical Association. As reported by the Boston Globe, the piece has been downloaded more than 100,000 times making it "an instant cult classic." A key finding: "Among 45 most highly cited clinical research findings of the past 15 years, 99 percent of molecular research has been refuted."

It occurred to me that this "dirty little secret" is not confined to epidemiological research, but applies to all research studies in all fields including psychology, linguistics, sociology, and language learning. Let me explain.

## There is more than one culprit

The first is the lack of replication. One study alone does not make for breakthroughs. It takes many replications by different scientists to establish the reliability of a finding. For example, when the concept of an "atom" and a "molecule" was first proposed, physicists did not believe these were real entities. It took many different experiments by different investigators to establish the "reality" of each.

Another example: After three days of listening to research papers on brain lateralization at UCLA, the pioneer surgeon, Joseph Bogan said, "It was all very impressive but I wondered whether the differences between the right and left hemisphere are real until I walked over to the UCLA Medical Center and visited with a patient with serious damage in his left hemisphere. According to the researchers, speaking happens for most people on the left side of the brain and singing on the right. That patient I visited could not speak a word, but if you asked him to sing, he renders a heartbreaking aria."

## The second culprit - This is the most serious

Jonathan Sterne, a statistician, and George Smith, an epidemiologist from England's University of Bristol said in the British Medical Journal, "The widespread misunderstanding of statistical significance is a fundamental problem in medical research." I would add, in all research in any field.

# When is a "significant difference" not significant?

The answer is: Always. Hey, that does not make sense. Throughout our undergraduate and graduate training, we read that this study and that study showed a "significant difference." When is a significant difference not significant?

The explanation, for which most investigators seem to have amnesia, is this: When a statistical procedure such as the t test, chi-square or analysis of variance reports a significant difference, it does not literally mean "significant difference." All it says is that with a certain degree of confidence, the researcher believes that there is a real difference in the means (or averages) in the "populations" from which the samples were drawn. "Real" means that the difference is not due to chance but to the independent variable. Remember, any sample of anything represents some larger group called a population.

# A sample can be only one case

A sample is not necessarily an aggregate, but can be one case only. For instance, your physician draws a sample of your blood, but that sample is only important because it represents all the blood in your body which is the “population." Information from the sample can be generalized to the blood chemistry in your entire body.

# Why a significant difference is a stingy piece of information

Telling practitioners in any field that, for example, "There is a significant difference in cholesterol for adults who take statins compared with those who do not," is not much information to make clinical decisions. The alert practitioner will want to know the magnitude of the difference. If the difference is tiny, I may not recommend statins for my patients because nobody wants to take pills if they don't have to, and then too, there may be side effects. But if the difference is medium or large, I will give the information serious consideration. This may be worthwhile for my patients.

# Notice that only a few studies in any field report the magnitude of the difference

Researchers forget what "significant difference" is all about, but they like the sound of the word "significant" because, I believe, it implies something important or worthwhile. Consumers of the research and editors of the journals seem to be content, so why change? Further, senior colleagues who will review the work of a junior scientist for promotion seem to be content. Isn't that good enough? No it isn't.

## Here is what I recommend

First, as I advise my students, if a statistic is "significant," calculate the effect size and feature this information in your report. As a "rule of thumb," Jacob Cohen's classic book, *Statistical Power Analysis in the Behavioral Sciences* (1988) recommends: For studies with human subjects, ES of 02 is small, .12 is medium and .25 is large.

## But, what does effect size mean?

Let's start with a fundamental idea that applies to any study in any field: The scores for samples being compared such as the cholesterol for adults taking statins and those who do not - those scores will vary from person to person. That variation can be decomposed into two parts: The variation within groups and the variation between groups. This principle holds true for all studies in any field.

## The variation within samples

The variation within samples is the result of "chance." Chance is kind of a "wastebasket" term that includes reasons why one person has a different cholesterol reading from someone else. The reasons are many, such as genetics, gender, diet, physical fitness, and age. Put them all into one bundle and we call it "chance."

The variation between samples is important to us because this is the result of the "independent variable," which is the statins in my example. So, the total variation in any study is the result of only two factors: variation within the samples, which we call "chance," and variation between samples which we call the "independent variable." In the game of science, we are always playing against our opponent, chance. We are always trying to beat chance.

If the variation between samples is about the same as the variation within the samples, we conclude that the statins performed no better than chance and therefore, the results are "non-significant." If, however, the statins perform better than chance, we won the game and announce, "Statins produced a significant difference in cholesterol, (which is called the "dependent variable").

## How journal editors can help in the reform

Journal editors and reviewers can help by insisting that researchers follow-up significant statistical differences by reporting the effect size. This is especially important in review articles that examine and integrate findings from many different studies and many different investigators. The reason: It is difficult to compare results from the t test, chi-square, correlation, and analysis of variance. This is like comparing apples, oranges, pears and pineapples.

When the investigator converts "significant" findings into effect size we now have a common metric for sensible comparisons.

## A final note

Eighty years ago, the statisticians, Peters and Van Voorihs, said something like this that seems to make perfect sense: "Our colleagues continue to apply different statistics in different studies which makes comparisons between studies ambiguous. This is puzzling when one can by-pass this problem by directly finding the variability within and between samples, and then the ratio of between divided by within variance - a ratio (which is the effect size) that tells the whole story. If you want to know how to do this, see my book, "A Simplified Guide to Statistics for Non-Mathematicians." Have fun! There is nothing more exciting than discovering something no one else in the world knows.

## R E F E R E N C E S

To order the Asher books online, go to www.tpr-world.com

Asher, James J. *Learning Another Language Through Actions (6th edition).* Los Gatos, CA., Sky Oaks Productions, Inc.

Asher, James J. *Brainswitching: Learning on the right side of the brain*. Los Gatos, CA., Sky Oaks Productions, Inc.

Asher, James J. *The Super School: Teaching on the right side of the brain*. Los Gatos, CA., Sky Oaks Productions, Inc.

Asher, James J. *The Weird and Wonderful World of Mathematical Mysteries: Conversations with famous scientists and mathematicians.* Los Gatos, CA., Sky Oaks Productions, Inc.

Asher, James J. *A Simplified Guide to Statistics for Non-Mathematicians*. Los Gatos, CA., Sky Oaks Productions, Inc.

Asher, James J. *Prize-winning TPR Research.* Los Gatos, CA., Sky Oaks Productions, Inc.

Cohen, Jacob. *Statistical Power Analysis in the Behavioral Sciences*, 1988 New York: Academic Press.

Peters, Charles C. and Walter R. Van Vorhis *Statistical Procedures and their Mathematical Bases*, 1940. New York: McGraw-Hill.

Send your comments or questions about this article to the writer at tpr-world@aol.com