Science: The Weakest Link in the Mosaic of Evidence

by Paul Bergner

As clinical herbalists or nutritionists, we base our practice on a constantly accumulating and evolving body of facts, drawn from a mosaic of widely varied sources of information. We learn from teacher/mentors, from our own experience, from colleagues and classmates, from a variety of traditions, from intuition and instinct, and in the modern era, from information of institutional Science. We develop a faith and knowledge in our methods as patterns emerge in this mosaic, our experience confirms what our teacher taught us, for example, which is also part of a tradition, and we hear the same from colleagues. Ultimately as healers, we see results in our patients, or we don't, and this is the ultimate scale for judgement of evidence. Today, the mantra of the medical establishment, and in many cases of the “green allopaths” of the herbal, naturopathic, and functional medicine worlds, is “evidence based medicine.” But this model rejects most of the mosaic evidence, and purports that the only useful piece of data to guide our practice must come from complex scientific investigations. In reality, evidence from the scientific literature is the least reliable piece of the larger mosaic of clinical experience.

Science has become the Elephant in the Room of natural medicine and herbalism in the 21st Century. This elephant has already squashed several traditions of alternative medicine flat, others are tottering on the brink of extinction, and medical herbalism on more than one continent has become severed from its historical roots. . On the one hand we are intimidated by Science, or we may react against it or ignore it, feel inferior to it, or some herbalists, naturopaths, and others want to jump on the elephant and ride it to respectability. But to be effective as herbalists, as healers, or simply as creative, aware, and competent individuals, we need to lose our intimidation by or admiration of the myth of the superiority of scientific evidence. Blind belief in Science amounts to a dominant religion in our society, a religion of Scientism, built on a faith in scientific materialism that is neither warranted nor supported by the facts.

The myth here is that Science has created a great circle of truth, and that what is in the scientific literature is somehow golden and better or more real or more reliable than other facts not within their circle. The information has been vetted, so we can accept it as golden and true. In this model, and information not so vetted is suspect, is anecdotal, is unreliable superstition, and is guilty until proven innocent.

Here is how the myth of scientific objectivity is portrayed.

- Highly trained and unbiased experts with knowledge unavailable to the common person conduct experiments. They engage in practices to reduce any unconscious bias they might have.
- They produce objective and indisputable facts on topics of general benefit to the human welfare.
- This information is then analyzed by statistical methods to determine its significance.
- It is then submitted for publication.
- After an editorial selection for the most useful and valid content, an article is then submitted for peer review by a group of highly trained experts in the field of study for evaluation. This review removes gross or subtle errors from entering into the body of published literature.
- The experiment may then performed by other researchers, and either proven to be true or false. Thus the system is self-correcting for errors.
- We end up with a circle of “vetted” facts in the published science, facts that are more real, more "evidence based," and everything outside that circle is suspect, or less real.

If when you read the above, you get a little glow about you and feel proud of what the human race has produced to pull us out of the darkness of ignorance and superstition, you are probably part of

the religious cult of Scientism. Read on if you are willing to being deprogrammed.

Economic bias

First of all, what is inside that circle of vetted science is primarily determined by economic and political interests. Anything other than the smallest research trial must attract funding, and such funding is not bias-neutral. For some research, the trial must promise a financial return or it will not be financed. So most herbs are never studied in clinical trials not because they do not offer the promise of their traditional use, but because they are not patentable and will not yield a profit. Exceptions to this are trials of concentrated herbal extracts or proprietary products which, although not patentable, may be very profitable for companies trying to break into the MD market with their herbal products or to create the next new fad herb for the public. The goal of some other research must support the institutional or political agendas of an educational or government institution to receive funding. I invite you to read the article "Big Pharma Bad Medicine" by Marcia Angell, who was editor of the *New England Journal of Medicine,* one of the most conservative and prestigious medical journals in the world, at the time it was written. She describes the systemic and nearly universal corruption of academic and medical research institutions by corporate funders. Most of what is inside the circle of vetted science is there because vested interests with their own biases want it to be there. *That which remains outside, unfunded and untested may still be equally valid, true, and useful.* Repeat that to yourself three times whenever you doubt the validity of the methods of medical herbalism. Keep reading and I will argue that our supposedly untested methods are probably *more* reliable than the material in the scientific literature.

Publication bias

Sponsored or funded trials that show positive results that meet the goals of the funders, are more likely to be submitted for publication than those with negative results; the funding corporation or organization has final say about submission for publication, and often does not publish information that might not support its own commercial or institutional agenda. This is called *publication bias.* This is a major problem in medical science, and you can see a review of some of the issues and key articles at the *Wikipedia* Encyclopedia site under the topic Publication Bias. This is not just a minor issue, it has been described as a major problem in medical literature at least since the 1980s, and one author in 1992 called the practice "medical misconduct." It introduces fatal flaws in the reliability of published trials of drugs. For example, bio-statisticians in the 1990s examined clinical trials of pain medications where a head-to-head comparison was done between two similar drugs for pain. There were about an equal number of published trials favoring each drug. However, about 90% of the published trials showing the superiority of one drug over another were sponsored by the manufacturers of the "superior" drug. We could estimate then, since the drugs were about equal in effectiveness that each of the competing sides buried 90% of their research and only published the 10% that favored their product. This is the equivalent of flipping a coin repeatedly until you get a random sequence showing that heads are superior to tails. It is dishonest, and the distorted published results are worthless for making clinical decisions.

Far fewer negative results are published than positive ones in the literature as a whole. Depending on the discipline, negative trials account for between 10% and 30% of published studies. The percentage has plummeted in recent years; researcher Daniele Fanelli of University of Edinburgh evaluated 4600 papers from across various disciplines, and found that the proportion of negative results fell from 30% to 14% over the 17 years between 1990 and 2007. Now negative trials may be *very* useful for other researchers and especially for clinicians. They may help prevent us from wasting time and money on ineffective, or marginally effective methods. Fanelli, in an editorial in the journal *Science*, has proposed that failure to submit negative results for publication should be classified as scientific misconduct, with penalties equal to those for other forms of misconduct, such as fabricating or omitting data.

A recent unveiling of the level of such misconduct appeared first in a now famous study in the *New England Journal of Medicine* in 2008. Researchers reviewed the studies that led to the FDA approval of 12 antidepressant drugs between 1987 and 2004. Out of 74 studies, 38 were positive (drug better than placebo) and 36 were negative (no benefit to drug). Of the positive trials, 37 of the 38 were published. Of the negative studies only 3 or the 36 were published honestly, and another 11 with negative results were published, with a positive spin in the abstract, title, or discussion not supported by the actual data. The remaining 22 went unpublished. The same year, another set of researchers used the Freedom of Information Act to access unpublished studies of antidepressant drugs on file with the FDA. When these were all pooled, antidepressant drugs only outperformed placebo in 20 of 46 trials. The antidepressants were no better than placebo in any trial for mild to moderate depression, and in severe depression they lowered the depression score an average of only 1.7 points more than placebo on the 52 point Hamilton depression scale.

Publication bias is not just an issue with drugs, it also infects the herbal literature, especially for herbs with the potential to become mass sellers or to make a breakthrough into the conventional medical community. It is hard to detect publication bias because you have to know about trials that have never been published. How would we ever find this out? I am aware of two of these in the field of herbalism. In one an *Echinacea* product was found to be worse than placebo for preventing respiratory infections in children. The sponsor would not allow it to be submitted for publication. In the other, an herbal remedy for allergies was found to be no more effective than an alfalfa placebo. This trial was never published. That company then redesigned the trial with a milk-sugar placebo and found, unsurprisingly, that the herb was better than milk sugar for treating allergies. That trial was published. For the herbalist, we need the results of all these trials to understand how to use the herbs. We need to know that the proprietary *Echinacea* product is not effective, and that the herb tested for allergies is not better than Alfalfa. The extent of such banished and exiled trials is nearly impossible to determine, I am quite sure that negative trials of a company's product are a closely held company secrets. We can sometimes see evidence for publication bias in trials of herbal mega-products after some decades just occurred with the antidepressant trials. For instance a standardized extract of Saw Palmetto berries has an established reputation as “The Prostate Herb” because it supposedly relieves the symptoms of benign prostatic hypertrophy in men. It was introduced and marketed during the 1990s on the basis of a few small clinical trials showing it to work. These were apparently selectively published, because by 2012, when the prestigious and methodical experts at Cochrane Database in England performed a meta-analysis of all published literature to that date, it concluded that the Saw Palmetto products, even taken daily for months, had no beneficial effect at all on urinary flow or prostate enlargement, even when given in double and triple doses in some trials. The original early trials were selectively published false positive trials.

The myth of effective peer review

A journal must be *peer reviewed* for its content to be accepted into the body of scientific knowledge. A panel of experts screens all articles for errors of method, fact, or interpretation, and recommends for or against publication, often qualifying the approval only after minor or major changes are made. We need only then say the words “peer reviewed” and the true believer in Scientism then will then glaze over and put faith in the results without a further thought. After all, some expert has already done our thinking for us and we need not be bothered. A growing body of scientific literature is evaluating the effectiveness of peer review, and the results are not only shocking, they call to question the validity of the entire edifice of scientific respectability. Consider a test conducted by John Bohannon, a Harvard biologist. He invented an article, composed completely of fabricated data, on the supposed effects of a lichen extract on cancer cells. He deliberately loaded it with errors in study design, analysis, and interpretation of the results. He then submitted the paper to 304 peer-reviewed journals. One hundred fifty-seven of these journals accepted it for publication. In another such sting

operation, Fiona Godlee, the editor of the top-tier peer-reviewed *British Medical Journal* took a well written article that had already been accepted for publication, and deliberately injected 8 errors in study design, analysis, and interpretation. She then submitted it to more than 400 of the regular peer reviewers of the *BMJ*. Two hundred twenty of them reviewed and returned their comments. Not one reviewer found all eight mistakes, some did not spot any at all. On average, they reported fewer than 2 mistakes, missing more than 3/4th of the deliberate errors. Sixty percent recommended publication with major or minor changes. These are not just some occasional outlier errors in peer review, these two studies call into question the validity of the entire process of peer review, even at the most prestigious journals.

Poor or rubber-stamp peer review also affects the literature of natural medicine and herbalism in two ways. First of all, none of the jury of peer reviewers at most conventional journals are in any way qualified to comment on or evaluate natural medicine or herbal products. For instance, one peer reviewed journal accepted an article purporting to show that feverfew has no benefit in migraine headache. Previous trials had shown the fresh plant to work, and the supposed active constituents to be volatile. The trial accepted for publication used old dried powdered material, an error obvious to one conversant with the practice of herbalism, but not to this set of authors and reviewers. So there is a bias against natural medicine in most peer reviewed journals. Another way that peer review distorts the literature of natural medicine is when journals are published or primarily supported by alternative medicine companies or institutions. One journal published by a natural products company has a large panel of otherwise qualified peer reviewers listed in its masthead, yet regularly publishes fluffy pseudo-scientific articles that promote its own products. But because it fulfills the paper-requirements for peer-review, its articles are catalogued at the U.S. National Library of Medicine and available in the MEDLINE database. In this instance a company or institution started its own journal, picked a group of peer reviewers in order to look respectable, but now publishes rigged or biased articles favoring their products, and uses the published results to demonstrate that their products are “scientific.”

False positive trials

More serious than any of the above already-fatal flaws are structural problems with the very foundation of scientific trials submitted for publication. These have to do with the nature of research itself, and its statistical analysis, and the likelihood that a published trial has a “false positive” result. I invite you to read an article by researcher and bio-statistician John Ioannidis of Stanford University. He analyzed a large number of clinical trials and found the average likelihood of a false-positive results in all published clinical trials is greater than 50%. His article, entitled “Why Most Published Research Findings are False” is available online for free. In another study, Ioannidis found that the likelihood of a false positive result in the field of neuroscience was 79%. Another researcher found the likelihood of false positive finding in the literature of psychology to be 65%.

These are not just a few eccentric iconoclasts from ivory tower institutions finding this level of error. Scientists at the pharmaceutical company Amgen, aware of the problem of false positive trials, was working in the field of cancer drugs. They sought to evaluate the validity of a group of 53 studies considered to be landmarks in the basic science of cancer and chemotherapy. They needed valid information in order to conduct research on their drugs. In many of these trial-replications they worked with the original researchers to make sure that their methods and materials were identical. They were only able to replicate the results of these 53 supposedly bedrock science articles in 6 trials, or 11% of them. Their study by first author Glenn Begley, appeared in 2012 in the prestigious journal *Nature.* A similar evaluation supervised by researcher Florian Prinz at the Bayer HealthCare pharmaceutical giant in Germany, attempted to reproduce 67 seminal studies in the areas of oncology, women's health, and cardiovascular health. They could only reproduce the original data in 7% of the articles, and found similar data in only 14% more. In other words, 79% of the supposedly gold standard trials, all previously peer reviewed, which formed the bedrock of research in three fields, were not reproducible.

This research was published in 2011 in the equally prestigious *Nature Reviews Drug Discovery.*

Natural products and p-hacking

Some researchers can produce false positive results deliberately using a method dubbed "P-hacking." A p-score is a standard measurement of how likely an outcome has been due to random chance rather than to the intervention measured in a trial. A statistical standard for "significance" of the outcome of a trial is that the P-score must indicate at least a 20:1 likelihood that the results are *not* due to chance. This is indicated by a P-score of .05 or less. However, if a large number of parameters are measured, especially in a small number of participants, it is easy to find one or more of them to show statistically significant results. P-hacking is a widespread practice in natural products research, and tests of proprietary products. It is especially evident when a trial is funded by the natural products company, one or more company officials are authors of the study, and a small number of participants are tested. John Bohannon, cited above for demonstrating the flawed peer review process in hundreds of journals, deliberately rigged a clinical trial, using P-hacking, and produced a spurious correlation between eating chocolate while dieting and losing more weight. He rigged the trial by using a small number of participants, but then measuring a large number of parameters, until he found some that were false-positive. He then got the article published for a fee in a sketchy journal without peer review, and circulated the results to the European press with a press release from his made-up foundation. The news swept the headlines of the major media in Europe in several languages. Bohannon describes the whole project, including details on P-hacking, in an article entitled "I Fooled Millions Into Thinking Chocolate Helps Weight Loss. Here's How," which is available on the internet. I a recent review of a magazine which reports on natural products research in the herbal world, nearly every study cited had used p-hacking to produce false positive results to promote a proprietary product.

Selective citation of false-positive trials.

What then occurs with many of these false-positive trials is that the early "breakthrough" trials form the foundation of a meme that persists for decades through *selective citation.* Selective citation occurs when, to prove a point in an article, any previous studies which support the point are referenced, but those that contradict it are ignored. As the scientific narrative builds, the original trial or group of false-positive trials are verified by some other trials, fail but to be verified in others. Statistical analysis shows that the positive trials which support the original ideas are cited much more frequently than the negative trials which contradict them. The severity of this problem cannot be overestimated. Consider that bio statistician Steven Greenberg, researching this phenomenon, performed an analysis of 17 research papers on a topic in the area of muscle physiology. In these papers, collectively, there were 32 total citations, many of them appearing in all 17 papers. Of these, 31 citations were to a group of four studies that supported the authors' position. A total of ten similar studies were available, the other six contradicting the findings of the four more-frequently-cited ones. Thus the evidence against the positive trials was stacked 60:40. However, only *one* of these 6 negative studies was cited in *one* of the 17 research papers; the other sixteen papers ignored all six negative trials completely. Two of the negative trials were performed by the same set of researchers that produced the 4 positive trials, yet even these were ignored. None of the other groups that tried to replicate the results of the original team could do so. You can see Greenberg's article "How citation distortions create unfounded authority: analysis of a citation network" for free online. You can see this is not just a small bias in one direction – it can only be classified as severe and it penetrates into all fields of science. Note the "foundational" studies evaluated by Amgen and Bayer above, they were foundational because they were the most frequently cited. The body of science is not only unable to self-correct, but it tends to *amplify errors.* It is like termite infestation that has eroded the foundation of the body of scientific literature, leaving it thoroughly rotten.

This amplification of false results occurs with regularity in the supposed scientific herbal

medicine. In one example identical to the phenomenon above, some early trials of *Rhodiola* in Europe as an adaptogen showed positive results, and today these trials are the most frequently cited in literature about *Rhodiola.* A reviewer in 2002, however, examined all positive and negative trials of *Rhodiola* and found that only the three most poorly designed trials showed positive results, and that the better designed trials, a majority, showed that *Rhodiola* did not benefit the parameter being measured. The reviewers all but accused the original researchers of rigging data and other scientific misconduct to spin their trials to support a commercial product in which they had a commercial vested interest. Likewise with studies of *Eleutherococcus* for exercise performance, a set of early false positive trials are now the most frequently referenced, even though another 6 trials, better designed than these early positive trials, showed that it does not benefit athletic performance. One, in fact, demonstrated that it promoted athletic burnout in a group of runners who used it regularly.

And just a further note on references in those trials. During the 1990s I was a member of the National Association of Science Writers, a group primarily composed of science editors and writers at newspapers and magazines. The NASW published a quarterly newsletter with information of interest to a writer or editor. In one article, researchers took a large selection of published scientific trials in assorted fields. They then looked up the references in each article to see if they in fact supported the statements of facts in the article. In 24% of cases the citation did not support the claims in the article, and in some instances actually said the opposite of what they were purported to.

Consider then, the next time you see, or read about, or someone talks about a scientific study, uses words like “double blind” and “peer reviewed” and you begin to feel intimidated, or feel confidence, or feel anything other than a critical skepticism, consider that the study has:

1) A nearly 100% chance of having bias from its funding source.
2) A nearly 100% chance that it has not been effectively peer reviewed.
3) If positive, a 70-80% chance that it is showing a false positive result.
4) A likelihood that other equally valid but contradictory evidence exists which has either not been published or was not cited.
5) A 24% chance that any one of its references will not support what the authors say it does.
6) An 80-90% chance that if someone tries to reproduce the experiment exactly as it was performed that they will not be able to reproduce the results.

I'm herbalist and I'm proud. I am careful and critical in what I believe, practice and teach. I have a high degree of functional scientific literacy, and for the last thirty years have spent some hours most weeks reading current science in nutrition and medical herbalism. I learn what I can from the field of science, with an emphasis on pathophysiology and nutrition. I am not a magically-thinking idealist about the infallibility of my methods. But I am sure that the materials and methods I use, taught to me by my elders, and developed over my 40 year career, and reproduced repeatedly by students and faculty in the teaching clinics that I have supervised, are predictable, reliable, and reproducible at a higher percentage than those studies published in scientific journals.

Paul Bergner is Director of the North American Institute of Medical Herbalism, http://naimh.com and Editor of the *Medical Herbalism* journal.