

# The End of Biomedical Journals

## There is Madness in Their Methods

Mikel Aickin

University of Arizona

03sep05

In *The Structure of Scientific Revolutions*, Thomas Kuhn portrayed normal science as slipping into a moribund condition, in which it could no longer provide acceptable answers to its own questions. Then some kind of shift would come along to replace the dominant outlook with an improvement, which was in turn destined to become the new version of normal science, perpetuating the whole cycle. In Kuhn's version, it was the discovery of new puzzles, consisting of observations that could not be satisfactorily explained in the current paradigm, that led to an apparent shift. He did not, however, consider a situation in which the methods of normal science might simply degenerate, producing the same kind of crisis, and possibly the same kind of resolution.

There is unsettling evidence that we are now in the midst of a methodologic degeneration in biomedical science. This appears to be occurring in, of all places, our fundamental approach to inference – using observation and evidence to decide how to act or believe. That it might be happening in medical research makes it of more than just academic interest.

One of the few benefits in a degeneration of conventional methods is that the normal scientists are unlikely to recognize that it is happening, and so the process will be not only made public, but actually touted as excellent science. So it is with a remarkable article published on 27aug05 in the *Lancet* regarding homeopathy.<sup>1</sup> The editors of *Lancet* are evidently proud of their publication, since they use it as the basis for a call to end homeopathy. Does this article justify the editorial, or is it in fact a betrayal of the very principles that the *Lancet* claims to stand for? Let us see how this specific article fares in the light of the conventional criteria that are applied to articles in clinical trials and biomedical science generally.

***Treatment Comparisons.*** A clinical trial has investigated two therapies for a given condition. On a scale in which larger numbers are better, and zero stands for no effect, the effect of therapy A is 0.54 (SDE=0.196), while for therapy B it is .13 (SDE=0.154). The researchers conclude that treatment A is effective ( $p=0.006$ ) while therapy B is not ( $p=0.40$ ).

The article is, of course, not accepted for publication. The reason is that the whole point of having two groups in a study is to compare them with each other. The difference between the treatment effects in the two groups is 0.41 (SDE=0.249) with  $p=0.10$ . By the conventional criterion for making such comparisons, this result is not “statistically significant”. It means that there is no basis for saying that the two therapies have different effects. The study is null.

The data come from the abstract of the *Lancet* article. 0.54 is the negative log of the odds ratio (0.58) from conventional studies, and 0.13 is the same transformation of the odds ratio (0.88) in homeopathic studies. In the abstract, and in the comments elsewhere in the issue, the faulty analysis is treated if it were correct: therapy A (conventional medicine) is indeed effective, while therapy B (homeopathy) is not.

**Differential Compliance.** Another study has randomized 110 patients each to two therapy groups. The therapies are hard to maintain, and so only 21 patients comply in one group, while an even more disappointing 9 comply in the other group. The difference is “statistically significant” with  $p=0.018$ . The authors are surprised when their article is rejected, on the grounds that such a low rate of compliance, combined with a differential between the two groups, casts the results in serious doubt. The study has failed.

The numbers are from the abstract of the *Lancet* article. There were 21 “high-quality” homeopathic studies, and 9 “high-quality” conventional studies. The conclusion is clear; there has been a “statistically significant” demonstration that homeopathy articles are of higher quality than comparable conventional medical articles on the same topics. Unfortunately, this invalidates the rest of the paper. (As a footnote, it was only recently that the supposed poor quality of CAM research was being cited as the reason for a false excess of positive CAM studies. Now that the quality results are in the opposite direction, this argument is evidently no longer valid.)

**Intent-to-treat.** Yet another study also enrolls 110 pair-matched patients in each of two groups. One group has 8 evaluables while the other has only 6. The article is rejected on the grounds that once patients are entered into the study, they must be analyzed in their original group. This means, among other things, that if they did not contribute endpoint data, then some imputation scheme must be used. The results as presented are faulty not only because more than ninety percent of the data are missing, but because there is no guarantee that the patients actually analyzed are matched (that is, the pair matching was destroyed by the missing data, a point passed over by the authors). The process of selection that produced “evaluables” is not above question.

The data come from the abstract of the *Lancet* article. The odds ratios cited above are based on 8 homeopathic and 6 conventional articles (not 110 of each, as implied elsewhere in the article and in the *Lancet* editorial). The loss of pairing was ignored, of course. The validity of the measures used to include articles is not adequately justified, despite the fact that the results might well be almost entirely driven by them.

**Post-study power computations.** A study without a control group reports an apparent treatment effect of 0.13 (SDE=0.154). This is properly reported as not “statistically significant”. The article is only accepted subject to revision, since a negative study with a small sample size should provide a power computation (this is not, as often and erroneously thought, to justify the study in the first place, but to determine whether the results are worth anything at all). A conventional calculation gives a detectable effect of 0.462 (power 85%). The editors decide that this is too large to be reasonable, and reject the article.

The data come from the abstract of the *Lancet* article. A negative result is reported (homeopathy is no better than placebo) with a miniscule sample size, and no power calculation.

**Control of confounding.** A group of epidemiologists conduct an observational study of seven risk factors on a disease outcome. The issue is to determine whether the risk factors are the same in two groups of people. The data presentation consists of a series of univariate odds ratios, one for each risk factor, with p-values to test a null association. The article is rejected for two reasons. First, since the purpose was to compare risk factors across the two groups, the comparisons with null effects are not germane, and the obvious comparisons between the groups should be made. But more importantly, there is a known confounder that should have been controlled in the analysis (that is, there should have been a multivariate analysis), and moreover the risk factors that were analyzed are intercorrelated, so that again multivariate analyses should have been carried out.

The data are from Table 3 of the *Lancet* article. One could take quality as the confounder, or perhaps one of the other factors. There is, of course, no reason to dichotomize quality, and since this generally results in misclassification bias, there is reason not to. Obviously the analysis does not compare conventional medicine with homeopathy, but rather compares each to the null. An appropriate analysis would not only make the comparison between therapy groups, but would take the pairing into account.

**Meta-analyses.** There are, therefore, five areas in which the *Lancet* article does not meet the minimum, conventional criteria for publication in biomedicine. This is, however, not the most serious problem with the article. For this, we need to go back to recall the original aim of a meta-analysis, or overview. It is to assemble all of the obtainable, relevant literature on studies done for the purpose of comparing different therapies for a given condition. The original reasons for developing the concept were to collect scattered literature into one place, to apply uniform criteria for study selection and analysis, and to come to a conclusion about the best therapeutic approach, or to say that the evidence was not yet conclusive. Somehow this precise and useful form has degraded into an unrecognizable hash, in which any papers on any topic can be bundled together in an investigation of questions of unlimited ambiguity. A classic paper along this path has already been published in the *Annals of Internal Medicine*.<sup>2</sup> Here the authors studied a single therapy (vitamin E supplementation), breaking the first rule of meta-analysis, for multiple conditions (breaking the second rule) in studies not designed to test the therapy (breaking the third rule). There is evidence that they were not sufficiently careful about the form of the various vitamin E treatments, violating a fourth rule.<sup>3</sup> This study in effect concocted perhaps the most biased sample of human beings one could find in the biomedical literature, and then made the truly bizarre assertion that its results applied to everyone. One can only presume that the lack of a negative reaction to the *Annals* article paved the way for the *Lancet* article.

There are other examples, of a similar order of strangeness, but I will only mention the therapeutic touch article published in *Journal of the American Medical Association*.<sup>4</sup> This article was on research carried out by a nine-year old girl, under the direction of her mother, an ardent opponent of therapeutic touch. The methodology was debunked in an article in *Alternative Therapies*,<sup>5</sup> showing that it contained appalling, irremediable flaws. After the original article was published, the *JAMA* editor was criticized for poor judgment, by accepting a low-quality article to make a political statement. This could be seen as an unjustifiably beneficent interpretation, however, because no one seems to have noticed the very real possibility that the article, poor though it was, actually did meet *JAMA*'s scientific standards.

**Malpractice.** If you make a few simple assumptions, you can roughly compute the number of possible instances of malpractice that a physician might commit in a lifetime of practice. It is not a particularly large number. Now consider a journal that publishes papers which mislead health professionals and ordinary people about the effectiveness of medical practices. Surely one article, no less one researcher, can have a harmful effect through research malpractice that dwarfs the meager capacity of a single physician. The malpractice risk of a typical journal must be even larger.

If we are to see a continued degradation of methods in biomedical research, supported by “leading” journals, then perhaps it is time to think about the End of Biomedical Journals, as we know them. In the US at least, it would not seem unthinkable that the National Institutes of Health, through the National Library of Medicine, could undertake a web-based project to publish all funded, and much of the unfunded research, in all areas of biomedicine. The need for the current unregulated system would vanish. Editors and referees would continue to be needed, but they would operate under rational regulations, and would not be in a position to endanger the public health on the basis of personal whims. Incidentally, the job of meta-analyses would be infinitely easier, since the hunt for relevant articles would be all but accomplished. And, as we know because of PubMed, the technology is available, and the NLM knows how to apply it.

To return to Thomas Kuhn, we certainly have many biomedical puzzles that are worth working on, and which have not been addressed by normal biomedical science. Some of us are engaged in an experiment to see whether we can fashion research tools that will help us to understand more, by extending existing methods when feasible, and developing new ones when appropriate. For us, it is particularly discouraging to see normal biomedical scientists perverting their own tools for the evident purpose of attacking unconventional therapies.

### **References**

1. Aijing Shang, Karin Huwiler-Müntener, Linda Nartey, Peter Jüni, Stephan Dörig, Jonathan A C Sterne, Daniel Pewsner, Matthias Egger. Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy. *Lancet* 2005; 366: 726–32
2. Miller ER, Pastor-Barriuso R, Dalal D, Riemersma RA, Appel LJ, Guallar E. Meta-analysis: High-dosage vitamin E supplementation may increase all-cause mortality. *Annals of Internal Medicine* 2005;142(1):37-46
3. Neustadt J, Pizzorno J. Vitamin E and All-Cause Mortality. *Integrative Medicine* 2005;4(1):14-17
4. Rosa L, Rosa E, Sarnier L, Barrett S. A close look at therapeutic touch. *JAMA* 1998;279(13):1005-1010
5. Cox T. A nurse-statistician reanalyzes data from the Rosa therapeutic touch study. *Alternative Therapies* 2003;9(1):58-65