

Tales from the Crypt: Fables of Foibles, or RCTs That Go Bump in the Night

Anthony Rosner, PhD, LLD [Hon.], LLC

On the eve of Thanksgiving (November 19, 1999), one would assume that another Halloween season is now safely behind us. Then again, maybe not, considering what some of the most recent medical journals have to say about that holy grail of evidence-based medicine: the randomized clinical trial. Visions of ghosts, hobgoblins and skeletons that were supposed to have retired for another year are upon us once again. This is because the revelations in these journals tell us that (I) humankind's interpretation of the results of clinical trials remains far from a complete science, and (II) outcome measures are clearly susceptible to the vagaries of human nature; not only from the perspective of the human subjects tested in clinical trials, but especially from the behavior of the trial investigators as well. Some of these images border on outright skullduggery.

The first of these ghost stories appeared very recently in the *Journal of the American Medical Association*, a publication that I have found to be one of the most conscientious vehicles of the critique of medical practice and experimental design. In this particular study, Juni prepared a statistical review of the results of clinical trials (called a meta-analysis) that compared low molecular weight heparin (LMWH) to standard heparin preparations in their respective abilities to prevent postoperative thrombosis. He used no fewer than 25 previously used scales to distinguish between high-quality and low-quality trials. Depending upon which scales he used and whether he was assessing high or low-quality trials, he could demonstrate that LMWH either was or was not superior to standard heparin solutions.¹

Therefore, depending upon which scale one uses, it is possible to obtain diametrically opposing results!

The second hair-raising tale, also appearing in *JAMA*, is also devoted to bizarre events that happen en route to building a meta-analysis. This paper concerns the use of pharmaceutical agents for patients with cancer complicated by neutropenia. My reaction to this reading is easy to describe if you have seen "The Blair Witch Project." At the very least, I felt as though I had just witnessed a major miscarriage of scientific integrity. Specifically, comparisons were made in this report between fluconazole (a new antifungal agent manufactured by Pfizer, Inc.) and amphotericin B (an older pharmaceutical manufactured by Bristol-Myers Squibb to eliminate fungi from the GI tract). A cursory examination of the trials was found to yield a distinct advantage to fluconazole. However, the picture began to rapidly unravel once you took a closer look at the vexing details in this work, in which we have been told so many times the devil must surely reside:

1. In 79% of the patients randomized in the trials to receive amphotericin B, the drug was administered orally even though it is widely known to be poorly absorbed through the GI tract (it should be administered intravenously). Fluconazole, on the other hand, displays no such problems of absorption.

2. In 43% of the patients identified for meta-analysis, results from the administration of amphotericin B were inexplicably combined with those from an ineffective agent known as nystatin.
3. When asked about these irregularities in detail, the investigators seemed to take a powder whenever the going got rough. No less than 12 of the 15 authors of the studies reviewed turned out to be noncompliant after being contacted by the mail twice. One author claimed that the trial was "old" and that the data were impounded with the drug manufacturer, failing to disclose that the data actually had been published; one author professed lack of access to the data due to his change of professional affiliation; and one author maintained that the sponsoring university failed to furnish a copy of the primary data.
4. Twelve of the 15 authors polled (their studies involved 92% of the total number of patients) affirmed that they were employed or supported by grants from Pfizer, the manufacturer of fluconazole.²

In earthier terms, this phenomenon is known as asking a runner to compete in a marathon after cutting that person off at the knees. In less-than-subtle terms, we are given a clue as to the identity of the axe-wielder.

The third creepy saga also appears in *JAMA* and tells us of yet another obstacle to the proper preparation of meta-analyses, which are supposed to systematically review clinical trials which have been published regarding the treatment of a given clinical malady. Here we are introduced to the fine art of creating "sausage publications," which is what occurs when an author or group of authors publish virtually duplicate or highly overlapping (read: redundant) papers in separate places. What happens, understandably, is that the multiple appearances of the same or highly similar results in separate articles leads to their overrepresentation in meta-analyses.

Worse, authors who publish clones of their works often fail to inform people of their serial productions unless it is in their own best interest (as in the preparation of their CVs). They do this by omitting citations of their works and may even permute the order of the authors in order to cover their tracks. Unless the astute reviewer is aware of these fine points, he or she will cite and rate these papers as if they were independent studies, giving them more weight than is warranted.³ Specific examples of these shenanigans follow:

1. A meta-analysis of the use of NSAIDs for rheumatoid arthritis revealed that 31 clinical trials were represented in 44 articles, with 20 trials appearing twice, 10 appearing three times, and one appearing five times.⁴
2. A meta-analysis involving risperidone, an antipsychotic medication, revealed that 20 papers represented only seven small and two large clinical trials.⁵
3. A meta-analysis of 84 trials involving the pharmaceutical ondansetron for postoperative emesis

in 11,890 patients collapsed into 70 trials involving only 8,645 patients, the remaining patients having appeared in duplicate articles and leading to a 23% overestimation of the efficacy of ondansetron. Worse, these duplicate publications were not noticed by expert authors of at least eight subsequent articles, reviews or book chapters.⁶

Fourth in the parade of spooky RCT stories is an elegant deconstruction offered by Howard Vernon of the Balon asthma study that appeared in *The New England Journal of Medicine* a year ago. By focusing on functional tests of the lung rather than symptom relief of patients' global scores, Balon's study papered over positive trends in several outcomes, concluding that no statistically significant difference existed between active and sham treatments for pediatric patients with asthma.⁷ Even though clinical improvements were observed in both arms, the paper strongly implied that chiropractic was of little or no benefit in managing asthma in children.

In truth, sample sizes may have been too small to unequivocally demonstrate an active treatment effect, known as a type II error.⁸ The way this situation should have been approached is best exemplified by the integrity-filled efforts of Nils Nilsson, who revisited what he suspected had been a type II error in his earlier study of the effects of manipulation in the management of cervicogenic headache.⁹ It was only by increasing the sample size in his later study that Nilsson was able to obtain a statistically significant improvement obtained by high-velocity manipulations.¹⁰

The fifth scary legend is what happens when the quality of clinical trials is scored by slavishly attaching arbitrary scores (themselves rarely validated!) to the numerous features of a published RCT without thinking through their consequences. Thus, numerous methodological scores such as those provided by Koes¹¹ and others create a misleading profile of high- and low-quality studies if they place too much emphasis upon sham procedures, which we already know will seriously compromise controlled studies involving physical methods such as spinal manipulation⁷ if they are not true placebos. Clinical irrelevance remains a fatal flaw in the Cherkin low-back pain study published in *The New England Journal of Medicine*¹² and which I^{13,14} and others¹⁵ have extensively criticized elsewhere. In other instances, the mere utterance of such terms as "blinded" or "randomized" in the title of a paper may be sufficient to glean points in the rating of clinical trials, even though such terms are never defined or qualified. The proper remedy here is to demote the trial ratings if such terms are inappropriately used.¹

Other RCT fables abound. I have already presented a few with political overtones in this column previously¹⁶ and, in the interest of space, will spare those gory details for now. Suffice it to say that while the randomized clinical trial remains an important piece of the mosaic of evidence that needs to be assembled to substantiate a clinical procedure, it is not the only piece. One needs to remain vigilant in questioning the integrity of that particular piece. Armed with the stakes and garlic of insight and patience, the astute clinical researcher will hopefully be able to keep the vampires and spirits emerging from the misinterpretation of randomized clinical trials at bay. It gives a whole new dimension to the term "stakeholder"!

References

1. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association* 1999;282(11):1054-1060.

2. Johansen HK, Gotzsche PC. Problems in the design and reporting of trials of antifungal agents encountered during meta-analysis. *Journal of the American Medical Association* 1999;282(18):1752-1759.
3. Rennie D. Fair conduct and fair reporting of clinical trials. *Journal of the American Medical Association* 1999;282(18)1766-1768.
4. Gotzsche PC. Multiple publication of reports of drug trials. *European Journal of Clinical Pharmacology* 1989;36:429-432.
5. Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996;347:1024-1026.
6. Tramer MR, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *British Medical Journal* 1997;315:635-640.
7. Balon J, Aker PD, Crowther ER, Danielson C, Cox PG, O'Shaughnessy D, Walker C, Goldsmith CH, Duku E, Sears MR. A comparison of active and simulated chiropractic manipulation as adjunctive treatment for childhood asthma. *New England Journal of Medicine* 1998;339:1013-1020.
8. Vernon H. The semantics and syntax of research: critique of impure reason. Research Agenda Conference IV, Arlington Heights, IL, July 24, 1999.
9. Nilsson N. A randomized controlled trial of the effect of spinal manipulation in the treatment of cervicogenic headache. *Journal of Manipulative and Physiological Therapeutics* 1995;18(7):435-440.
10. Nilsson N, Christensen HW, Hartvigsen J. The effect of spinal manipulation in the treatment of cervicogenic headache. *Journal of Manipulative and Physiological Therapeutics* 1997;20(5):326-330.
11. Koes BW, Bouter LM, van der Heijden GJMG. Methodological quality of randomized clinical trials on treatment efficacy in low back pain. *Spine* 1995;20(2):228-235.
12. Cherkin DC, Deyo RA, Battie M, Street J, Barlow W. A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients

with low back pain. *New England Journal of Medicine* 1998;339:1021-1029.

13. Rosner AL. Response to the Cherkin article in *The New England Journal of Medicine*. *Dynamic Chiropractic* November 2, 1998;16(23).
14. Rosner AL. Rebuttal to Cherkin comments in *Dynamic Chiropractic*. *Dynamic Chiropractic* January 1, 1999;17(1).
15. Chapman-Smith D. back pain, science, politics and money. *The Chiropractic Report* November 1998;12(6).
16. Rosner AL. FCER forum. The big picture of research. The limitations and politics of randomized clinical trials. *Dynamic Chiropractic* June 14, 1999;17(13).

JANUARY 2000