

# Clinical Trials with Surrogate Outcomes Have Brought Bad Medicine(s)

Henry H. Bauer, Ph.D.

## ABSTRACT

Clinical trials seem to promise truly scientific support for medical practices. Instead, they have been applied in ways that undermine individual patient-doctor interactions. Most medical practice in developed countries now concerns attempts to lower risks of such chronic conditions as cardiovascular disease, and clinical trials relying on surrogate biomarkers have led to over-prescribing of drugs.

## Introduction

It seems blindingly obvious—once one has read *Pharmageddon*.<sup>1</sup> Clinical trials promised to bring sound science into drug testing, but instead have brought a neglect of patients' individual needs and an epidemic of over-prescribing.

There are many technical reasons for distrusting clinical trials, including inappropriate or incompetent statistical analyses and inappropriate or biased protocols.<sup>2,3</sup> But more fundamentally, clinical trials deliver only average statistical information that should not be applied automatically to the case of any individual patient. In clinical trials, exceptional cases are of no concern or interest, as the statistical analysis only identifies the most probable average outcomes. In medical practice and treatment, however, every individual is a unique case.

## Statistical Analysis and Its Applications in Medicine

The approach to statistical evaluation used in clinical trials (as well as very commonly in other situations) is known as “frequentist” or “Fisherian,” the former because of its association with a particular view of probability, the latter because its development is generally credited to R. A. Fisher.

Fisher's initial concern was plant breeding and agriculture: Which strains of wheat are most suited to a particular environment? Do fertilizers work? Which work best? Seeds, weed killers, pesticides, irrigation, etc., could all be usefully tested by looking for “statistically significant” effects. The particular experience of an individual seed or plant or ear of corn was and is of no interest. In medicine, to the contrary, it matters very much indeed what happens to each individual, and that is not revealed in clinical trials.

In addition to the impact of guidelines based on clinical trials on individual patient-doctor relations, there are purely technical reasons why Fisher's approach can deliver misleading conclusions. These concern what should be regarded as “statistically significant” and whether a “statistically significant” result supports the hypothesis purportedly being tested.<sup>4</sup> Fisher's approach evaluates only whether data differ from what random chance alone would deliver, but even if that is the case, the reason could be something other than the particular hypothesis supposedly being tested.

That two things seem to go together more often than would be expected by chance does not imply that they go together often, let alone always. Consequently, “statistical significance” can mislead by ignoring how great the observed effect is. For example, in a large trial<sup>5</sup> comparing aspirin with clopidogrel for prevention

of “ischemic stroke, myocardial infarction or vascular death,” the latter occurred at an annual rate of 5.83% in patients on aspirin and at 5.32% on clopidogrel, a “statistically significant difference ( $P = 0.043$ )”; thus clopidogrel is marketed as having been shown to be superior to aspirin. But it is far from obvious that clopidogrel is more appropriate than aspirin for any given individual. The relative risk was reduced by less than 10% (by 0.51 from 5.83); but the absolute reduction of risk is much less than that since there are other potential causes of morbidity and mortality. That reduction in absolute risk would almost certainly be smaller than the added risk of side effects from clopidogrel by comparison with aspirin, particularly in older people.<sup>6</sup>

Public health policies, in contrast to individual patient-doctor interactions, properly, indeed inevitably, depend on generalizations and statistics. Vaccination against smallpox, for example, has evidently been so effective that it is held to be valuable despite the tiny proportion of individuals who may have been harmed by the vaccination. Since there is a general public benefit, individuals who nevertheless suffer ill effects from vaccinations are eligible for compensation from public funds.<sup>7</sup>

Intermediate between global public health and individual cases is the issue of group differences within populations. Statisticians, but few others, understand that global statistics can be entirely misleading when they mask differences among sub-groups.<sup>8, pp 61-62</sup> Healy gives the example of treatment for schizophrenia: a randomized trial of a drug showed a statistically significant effect whereas the disaggregated data reveal that only one-third of patients actually benefited significantly. Another third showed minimal if any improvement, while one-third became significantly worse under the treatment.<sup>1, pp 88-89</sup> Physicians and psychiatrists need to know which types of patients might benefit and which might be harmed, not that there is globally a statistically significant effect.

## Outcome Measures in Clinical Trials

The benefits of vaccinations and antibiotics are readily confirmed in clinical trials because the desired outcomes are plainly evident: the proportion of people who did or did not become infected, or the proportion of infected individuals who did not die after treatment.

But outcomes are not plainly evident in clinical trials of the most widely prescribed “blockbuster” drugs. Today, most medical practice in developed countries is not dealing with acute or immediately life-threatening infections; it is focused on attempts to stave off the incidence of the chronic, aging-related conditions responsible for most mortality—cardiovascular (CVD) issues, stroke, diabetes, dementia.<sup>9</sup> The appropriate outcome measures here are longer lifespan, decreased mortality, or manifestly debilitating morbidity. To evaluate those directly would require large trials stretching over decades. Instead, potential treatments are approved on the basis of trials of limited duration (6 months is the Food and Drug Administration's requirement), and those trials rely on surrogate measures, biomarkers purported to be valid

indicators of the conditions to be treated. Thus blood pressure and cholesterol levels are surrogates for CVD, blood sugar for diabetes, and bone density measures for prevention of fractures. The trouble is that these surrogate biomarkers are not in fact valid representatives of the ailments for which they stand.<sup>10,11</sup>

Surrogate biomarkers are merely statistical associations, for example between CVD and cholesterol levels and blood pressure. Those associations have not been shown to be causative relationships. Association or correlation never demonstrates causation. Modern practice has slipped into an invalid semantic and intellectual progression. Associations and correlations are described as “risk factors,” which is justified only as long as the presence of a risk factor is taken as reason for clinical examination to determine whether the ailment is indeed present. Instead, “risk factor” has morphed into “risk,” as though the mere presence of the risk factor actually increased the likelihood of further progression of disease, and as though altering the biomarker could alter the chance of progression to manifest illness. This amounts to treating symptoms as though they were the underlying disease. Drugs are developed by relying on biomarkers, and treatment is based on biomarkers, creating a circular argument. But retrospective studies and meta-analyses have revealed that the presumed benefits have not been realized in practice from widespread treating of biomarkers by means of the blockbuster drugs, the “statins, antihypertensives, and bisphosphonates.”<sup>12</sup>

The confusion of biomarkers with causative factors is now being extended to vaccinations. On the basis of a statistical association, infection by some strains of human papilloma virus (HPV) has been misconstrued as showing that it actually causes cervical cancer. But the hoped-for effects of vaccination against HPV can only be evident in the far future, and in the meantime the costs are clear while the benefits are not. The incidence of cervical cancer is less than 7 per 100,000 women, while the rate of adverse events to this vaccination has been >50 per 100,000, including some deaths.<sup>13</sup> By 2009, 43 deaths had been reported in association with 25 million doses of Gardasil, involving perhaps 8-9 million vaccinated individuals since full vaccination calls for 3 doses to protect against 75% of cervical-cancer cases.<sup>14</sup>

### Undermining Personalized Treatment

As Healy puts it, reliance on clinical trials that evaluate surrogate markers has moved medical practice into treating conditions instead of individual patients.

Before clinical trials became standard, drugs came into use when they brought immediately obvious improvement; for example, chlorpromazine could return some manic or psychotic patients to normality in short order.<sup>1, p88</sup> But as statistical significance from clinical trials has become the criterion underlying official guidelines, prescribing has become an act of faith that future benefits are likely to accrue even as neither doctor nor patient may notice any signs of improving health. There are no direct ways to assess whether the risk of stroke or cardiovascular disease or diabetes, say, has been reduced in the case of any given individual.

Increasingly bureaucratic healthcare systems influence doctors to follow guidelines based on clinical trials as mandates rather than as suggestions. Physicians experience tangible personal risks if they deviate from “standard” practice, which includes the dosages stated in those guidelines. In earlier days, when doctors thought a particular drug might help a given patient, they would typically try the lowest possible dose and increase it only if no benefit was apparent. But clinical trials do not usually seek to identify the lowest useful dose; for example, clinical trials of drugs for treating human

immunodeficiency virus (HIV) seek to find the largest tolerable dose, to identify the “treatment-limiting toxicity.”<sup>15</sup> To minimize the risk of side effects, however, individual patients are best served by identifying the lowest useful dose, which will likely differ from individual to individual.

### Conclusions

The great advances in public and individual health from widespread elimination of infectious disease have changed our primary concern to such non-acute, chronic conditions as CVD, dementia, and diabetes. Clinical trials of potential treatments for staving off such conditions inevitably rely on surrogate biomarkers, which muddy the necessary distinction between correlation and causation.

Moreover, clinical trials identify what happens on average, information that may or may not be appropriately applied for a specific patient. Even worse, conclusions are drawn on the basis of statistical significance and not on what the magnitude of a possible benefit might be, and how it compares to the incidence of side effects.

The apparently “scientific” status of guidelines based on clinical trials has undermined properly individualized attention to the case of every individual patient.

**Henry H. Bauer, Ph.D.**, is professor emeritus of chemistry and science studies and dean emeritus of arts and sciences at Virginia Polytechnic Institute and State University (Virginia Tech). Website: [www.henryhbauer.homestead.com](http://www.henryhbauer.homestead.com). Address: 207 Woods Edge Court, Blacksburg VA 24060-4015. E-mail: [hhbauer@vt.edu](mailto:hhbauer@vt.edu).

### REFERENCES

1. Healy D. *Pharmageddon*. Berkeley, Calif.: University of California Press; 2012.
2. Goldacre B. *Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients*. New York, N.Y.: Faber & Faber; 2013.
3. Götzsche PC. *Deadly Medicines and Organised Crime: How Big Pharma Has Corrupted Healthcare*. Oxford, England and New York, N.Y.: Radcliffe; 2013.
4. Matthews RAJ. Facts versus fictions: the use and abuse of subjectivity in scientific research. *European Science and Environment Forum Working Paper*; 1998. Reprinted in: Morris J (ed). *Rethinking Risk and the Precautionary Principle*. Oxford, England: Butterworth; 2000: 247-282.
5. Ryan M, Combs G, Penix LP. Preventing stroke in patients with transient ischemic attacks. *Am Fam Physician* 1999;60:2329-2336.
6. Goodwin JS. Geriatrics and the limits of modern medicine. *N Engl J Med* 1999;340:1283-1285.
7. U.S. Department of Health and Human Services. Health Resources and Services Administration. National Vaccine Injury Compensation Program. Available at <http://www.hrsa.gov/vaccinecompensation/index.html>. Accessed Aug 10, 2014.
8. Bauer HH. *The Origin, Persistence and Failings of HIV/AIDS Theory*. Jefferson, N.C.: McFarland; 2007.
9. Centers for Disease Control and Prevention. Deaths and Mortality. FastStats. Available at: <http://www.cdc.gov/nchs/faststats/deaths.htm>. Accessed Aug 12, 2014.
10. Micheel CM, Ball JR (eds). *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*. Washington, D.C.: Institute of Medicine, National Academies Press; 2010.
11. Mack A, Balogh E, Micheel C (rapporteurs). *Perspectives on Biomarker and Surrogate Endpoint Evaluation: Discussion Forum Summary*. Washington, D.C.: Institute of Medicine, National Academies Press; 2011.
12. Järvinen TL, Sievänen H, Kannus P, Jokihäärä J, Khan KM. The true cost of pharmacological disease prevention. *BMJ* 2011;342:d2175. doi: 10.1136/bmj.d2175.
13. Schmidt & Clark. Gardasil® Side Effects Lawsuit. Available at <http://www.schmidtdandclark.com/Gardasil>. Accessed Aug 12, 2014.
14. Information about Gardasil®. Available at <http://www.gardasil.com>. Accessed Aug 12, 2014.
15. Farber C. *Serious Adverse Events: An Uncensored History of AIDS*. Hoboken, N.J.: Melville House; 2006: 275-309.