
THE WISDOM OF RUSSELL AND BURCH

14. *On the Proper Application of Appropriate Statistical Methods*

Michael Balls

The proper application of appropriate statistical methods is crucial to relevant and reliable biomedical research and testing

As I said in an earlier contribution in this series,¹ although Russell and Burch stated that “reduction, of all the modes of progress” [i.e. the Three Rs], “is the one most obviously, immediately, and universally advantageous in terms of efficiency”, I have always found their Chapter 5, on *Reduction and Strategy in Research*,² the most difficult part of *The Principles of Humane Experimental Technique* to understand.

Russell and Burch were strongly committed to the proper design and analysis of experiments, and, in particular, to the proper application of statistical methods. They rightly said that statistical advice is important at the outset, since “every time any particle of statistical method is properly used, fewer animals are employed than would otherwise have been necessary”. However, from where that advice is sought, and when and how it is applied, is crucial. As Russell and Burch put it, “statisticians are justly indignant, if asked to cope with the results of bad design. Of course, it is an elementary principle for any experimenter, not himself a statistician, to seek advice *before* experimenting.” I heeded that advice, and at ECVAM, I had the privilege of working with excellent statisticians, notably Graeme Archer and Sebastian Hoffmann. I have also benefited from consultation with others, such as Michael Festing and David Lovell, being in agreement with Russell and Burch that “toxicity testing is the scene of some confused thought”. Nevertheless, as Editor of *ATLA*, I am still regularly confronted by the misuse of Student’s t-test when comparing data from several experimental groups with one control, when Fisher’s analysis of variance should have been used instead.

Russell and Burch pointed out that “the science of statistics has been connected historically with three large-scale human activities: biological research, insurance, and gambling”. The connection with biology and medicine began at the beginning of the 20th century, and “has been associated especially with life insurance and two branches of biology – experimen-

tal agriculture and the theory of genetics and evolution”. They gave particular recognition to Sir Ronald Fisher (1890-1962), “who, more than anyone else is responsible for bringing statistical methods into experimental biology”. He introduced many concepts and methods, including biometrical genetics, and he was the first to use the term “variance” in statistics (i.e. how far a number in a set is from the mean value), an issue of great concern to Russell and Burch. Fisher introduced the concept of the analysis of variance, which was a considerable advance over the correlation methods used up to then, and is now a basic statistical tool in fundamental and applied biomedical research.³

Fisher had a glittering career, including a long and fruitful collaboration on ecological genetics with E.B. Ford, who taught genetics at Oxford to, among many others, William Russell and me. I remember Fisher coming to give a seminar in about 1958, and, on seeing him have a highly complicated discussion with Ford, in public, feeling that I was in the presence of two of the greatest minds of the 20th century. The outcome of their collaboration was “the general recognition that the force of natural selection was often much stronger than had been appreciated before, and that many ecogenetic situations (such as polymorphism) were not selectively neutral, but were maintained by the force of selection”.³ If I understand properly what that means, it may indicate that biomedical scientists should be more willing to take human polymorphism into account, especially when trying to translate the results of experiments on animals into meaningful conclusions about human beings. As Russell and Burch put it, researchers should take great care to avoid the high-fidelity fallacy.

Despite his eminence, Fisher did have what could now be seen as his weaknesses. For example, he was opposed to Bayesian statistics, and he had some thoughts on eugenics which would not fit very well

with the prevailing ethos of the 21st century. Also, he was opposed to the conclusion in the 1956 report by Richard Doll and Austin Hill that smoking causes lung cancer. “He compared the correlations in their studies to a correlation between the import of apples and the rise of divorce, in order to show that correlation does not imply causation.”³

Fisher might have been wrong on smoking and cancer, but his point about correlation and causation is still widely ignored today, especially by certain scientists, broadcasters and journalists, who time and again tell us of links between what we eat or don’t eat, or what we do or don’t do, and the awful fates of various kinds which await us as a result. Often, the correlation they are referring to is mere coincidence.⁴

Thinking that a correlation between two sets of data implies causation becomes especially dangerous when this is plausible, such as the postulated link between cooking in aluminium saucepans and dementia. The correlation-causation fallacy is discussed in a wonderful book on seeing the world through numbers, *The Tiger That Isn’t*, by Michael Blastland and Andrew Dilnot,⁵ and some amusing examples of meaningless correlations in US affairs have recently been given by Joe Arrigo.⁶ For example, the correlations between: suicides by hanging and spending on science; the marriage rate in Kentucky and deaths by drowning after falling from fishing boats; the *per capita* consumption of cheese and deaths from becoming entangled in bed-sheets; the *per capita* consumption of sour cream and deaths after falling from wheelchairs; and crude oil imports from Norway and drivers killed by collision with trains.

A good example of the plausibility problem came to my attention as I was writing this piece, in an article published in *The Times* on 23 December 2014, by Kat Lay, entitled *Heart patients more likely to survive if doctor away*.⁷ It referred to a study at Harvard Medical School, Boston, USA, published online by *JAMA Internal Medicine* on 22 December 2014,⁸ which indicated that, while 60% of high-risk heart attack and cardiac arrest patients died in teaching hospitals within 30 days of admission when senior doctors were away at cardiology conferences, 70% died when they were present. The survival rates of low-risk patients were not affected, nor were the survival rates of high-risk or low-risk patients in non-teaching hospitals. All kinds of plausible reasons for the difference are being suggested. For example, fewer high-tech, and possibly more risky, treatments (such as percutaneous coronary intervention) may be applied when the leading doctors are away at scientific meetings. That implies that the harms of sophisticated techniques might outweigh the benefits, or, as Anupam Jena, senior author of the report, said: “The evidence suggests that a less-is-more approach might be best for higher-risk patients with these conditions”,⁹ but proving cause, rather than mere correlation or coincidence, will not be easy.

Mark Twain popularised the saying, often attributed to Benjamin Disraeli, that “There are three kinds of

lies: lies, damned lies and statistics”, which is often used to cast doubt on the persuasive power of numbers or to cast doubt on the statistics used to prove an opponent’s point.¹⁰ That saying is, of course, nonsense, and exposes the weakness of those who use it.

In two other memorable books, both by Joel Best, *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists*¹¹ and *More Damned Lies and Statistics: How Numbers Confuse Public Issues*,¹² the author considers the simple premise that vast quantities of statistics are bandied about in all walks of life, and we frequently rely on them to form our own opinions, though we don’t understand how the numbers work. There are four kinds of people when thinking about statistics: awestruck, naïve, cynical, and critical – Best considers most people to be naïve, and that this can lead to poor policies and poor decisions, and his aim is to turn them all into critical thinkers.

Russell and Burch had a critical view of statistics, i.e. they had a balanced and sensible approach in exercising careful judgement and judicious evaluation, as they recognised that the proper application of appropriate statistical methods is crucial to relevant and reliable biomedical research and testing. We should all have the wisdom to do likewise.

Professor Michael Balls
c/o FRAME
Russell & Burch House
96-98 North Sherwood Street
Nottingham NG1 4EE
UK
E-mail: michael.balls@btopenworld.com

References

- 1 Balls, M. (2013). The Wisdom of Russell and Burch. 4. Reduction. *ATLA* 41, P24-P25.
- 2 Russell, W.M.S. & Burch, R.L. (1959). *The Principles of Humane Experimental Technique*, xiv + 238pp. London, UK: Methuen.
- 3 Anon. (2014). Ronald Fisher. *Wikipedia*, 8pp. Available at: http://en.wikipedia.org/wiki/Ronald_Fisher
- 4 Gerbis, N. (undated). 10 Correlations that are not causations. *How Stuff Works*. Available at: <http://science.howstuffworks.com/innovation/science-questions/10-correlations-that-are-not-causations.htm>
- 5 Blastland, M. & Dilnot, A. (2007). *The Tiger That Isn’t*, 184pp. London, UK: Profile Books Ltd.
- 6 Arrigo, J. (2014). *Joe Arrigo Perspective: Meaningless Correlations*, 5pp. Available at: <http://www.joearrigo.com/2014/05/14/meaningless-correlations/>
- 7 Lay, K. (2014). Heart patients more likely to survive if doctor away. *The Times*, 23.12.14, p. 22.
- 8 Jena, A.B., Prasad, V., Goldman, D.P. & Romley, J. (2014). Mortality and treatment patterns among patients hospitalized with acute cardiovascular conditions during dates of national cardiology meetings. *JAMA Internal Medicine*, doi:10.1001/jamainternmed.2014.6781

- ⁹ Anon. (2014). Patient outcomes when cardiologists are away at national meetings. *Eureka Alert*, 22.12.14. Available at: http://www.eurekaalert.org/pub_releases/2014-12/tjnj-pow121814.php
- ¹⁰ Anon. (2014). Lies, damned lies, and statistics. *Wikipedia*, 3pp. Available at: http://en.wikipedia.org/wiki/Lies,_damned_lies,_and_statistics
- ¹¹ Best, J. (2001). *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists*, 219pp. Oakland, CA, USA: University of California Press.
- ¹² Best, J. (2004). *More Damned Lies and Statistics: How Numbers Confuse Public Issues*, 217pp. Oakland, CA, USA: University of California Press.

The Principles of Humane Experimental Technique is now out of print, but the full text can be found at http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc. An abridged version, *The Three Rs and the Humanity Criterion*, by Michael Balls (2009), can be obtained from FRAME.