

Reply to Drummond and Vowler: statistics or the philosophy of statistics?

DOI:10.1111/j.1476-5381.2012.02135.x

I am grateful for the considered response to my letter (Ashton, 2012) by Drummond and Vowler (2012). Statistics is a powerful tool, but like all tools in science it is a tool for criticism not for confirmation. The view of statistics as a tool for confirmation leads to practices in science that may limit opportunities for discovery. Discoveries are often made when new methods are used, and new methods are always conjectural and often difficult to fit into the statistical model of experimentation. Some forms of statistical advice may inadvertently reduce researchers' willingness to engage in such risky science, and contribute to declining rates of innovation and discovery.

In response to Drummond and Vowler, my point can be restated thus: the hypothetico-deductive view of science is incomplete, as not only do we pose a hypothesis and test it, but at the same time (if we are doing anything new) test a methodological hypothesis about how to test the substantive hypothesis. Not only are our theories conjectural, but so also are our attempts to experimentally test them. Just as errors in our theories are uncovered by experimentation, just so problems in our experimental plans are uncovered as they are put to work. The belief that these errors can be predicted ahead of time by following best statistical design principles can indeed help in the promotion of good experimental practice, but there is also the danger that it may subvert a deeper search for error if it is imagined that uncertainty of all types has somehow been tamed and quantified.

Of course, statistical design principles must be used at all stages to attempt to balance the allocation of fixed variables and randomize the distribution of random variables, but even then conceptual flaws in the measurement techniques, manipulations, apparatus and so on that are not considered before the experiment often loom up to invalidate it as a candidate for a perfectly balanced experiment. Sometimes, these unexpected events will be dramatic, with effects so surprising and logically improbable *a priori*, that the inference of their causal connection with initiating conditions is overwhelming. An example would be the sudden and unexpected dissolution of a blood clot in the discovery of streptokinase. Although self-caricatured perhaps unfortunately as passing 'the bloody obvious test' (Kitchen, 1987), these events seem,

in the history of medical discovery, to be where many significant discoveries are made. Eliminating the possibilities for this by adhering only to established experimental paradigms that are conducive to *a priori* design principles necessarily limits the possibilities for such discovery. I conjecture that this is partly the reason for the downturn in innovation in pharmacology in recent decades despite massive increases in funding. Big science has not led to correspondingly big discovery rates, and the reliance of big science on procedures that limit the opportunity for fortuitous discovery may be partly to blame. Finding such lucky hits requires some degree of disorganization in science, with diverse research programmes following diverse approaches (Popper, 2012).

I conjecture that statistical analysis, as it is now conceived, is also partially to blame for this lack of exposure to dramatic results in another way. Apart from accidental discovery, dramatic results may also be found (if we are lucky) following bold hypothesizing. Such hypotheses not only set researchers out to look for potentially new and surprising results, but by leading the researcher to experiment in wholly novel paradigms raise the possibility of accidental discovery. Bold hypotheses are logically improbable hypotheses, but Frequentism takes no account of this whilst Bayesianism positively penalizes it. Factoring the logical *improbability* of our theories into an assessment of data is an exercise in conceptual rather than quantitative analysis, and clashes with statistics when it is used as a tool for confirmation rather than just for critical feedback. However, a reconciliation of views may be possible if it can be agreed what statistics in science is actually *for*. The falsificationist view is that statistics, like all other tools in science, can only perform a critical function. In this view, statistics acts as a check and balance on naïve pattern recognition and intuition. In another view, an inductionist view, statistics also performs a 'positive' confirmatory function. The title of my previous letter 'When Biostatistics is a neo-inductionist barrier to science' makes clear that it is the latter type use of statistics that I consider dangerous for science, not *of course* statistics as such.

John C Ashton

Department of Pharmacology and Toxicology, School of Medical Sciences, University of Otago, Dunedin, New Zealand

References

- Ashton JC (2012). When biostatistics is a neo-inductionist barrier to science. *Br J Pharmacol* 167: 1389–1390.
- Drummond GB, Vowler SL (2012). Response to 'When biostatistics is a neo-inductionist barrier to science'. *Br J Pharmacol* 167: 1390–1391.
- Kitchen I (1987). Statistics and pharmacology: the bloody obvious test. *Trends Pharmacol Sci* 8: 252–254.
- Popper KR (2012). Organisation of science. *N Z Rev Sci* 69: 22–24.