

unknowns of the experimental situation in the doing. To think otherwise is to ignore the very fallibility of the human species that experimentation seeks to correct. For example, a pharmacologist make discover part of the way through an experiment that some drug combination lead to an unexpected problem with solubility, and so the solvent concentrations are adjusted in some groups, adding another variable to the experiment. To say 'you should have thought of that before you started' is to sit on the sidelines and comment from the very hindsight provided by the actions of the experimenter. Unexpected problems are legion and quite unpredictable. If the statistician has no more to offer than to say 'start again and do it right this time', then he may get another visit from the scientist later, reporting another unexpected problem and adjustment mid-experiment.

The underlying issue was spelt out long ago by Sir Peter Medawar. His essay 'Is the scientific paper fraudulent?' (Medawar, 1964) should be required reading for all biostatisticians. Medawar was criticizing the way inductionism forces scientists to report their experiments in a misleading way. The neo-inductionism represented by the dictates of biostatistics as it is now conceived (both Frequentist and Bayesian) could very well lead to the opposite of intended results, forcing scientists to misleadingly contort their work to fit an even more inflexible inductionist schema. It needs to be remembered that the purpose of scientific publishing is to report interesting findings, not to provide credentials to a researcher or a research program. The scientist may have data that independent judgment perceives is clearly not random, qualitatively improbable observations (e.g. Ramachandran and Blakeslee, 1998) or patterns that are complex but discovered through a tortuous iterative experimental procedure; narratives that lack the symmetry beloved of mathematicians, and so are extremely difficult to quantify. If scientists stopped experimenting in this way, sticking to the inductionist fantasies of *a priori* programmatic experimental design beloved of funding agencies and statisticians, then the discovery process that is science will be severely compromised. A case could be made that where innovation occurs today, it is in spite of rather than because of statistical formalism; that progress is more likely to be made when experiments test bold hypotheses that are logically improbable in a way that is often not easily captured by statistics. Just as every inductionist attempt to make science a potentially automated and predictable set of procedures has failed in the past, so the neo-inductionist attempts of the present day by biostatisticians must similarly fail. Scientific discovery is not the same as operational research, and it is not that scientists have not caught up with statistics, but that statistics has still not caught up with the needs of scientists. While statisticians continue to lecture scientists from high alters of hindsight, as the scientists are busy getting their hands dirty in that dirty messy unpredictable real world, they may continue to find their lectures falling on deaf ears.

John C Ashton

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

References

- Drummond GB, Tom BD (2011a). How can we tell if frogs jump further? *Br J Pharmacol* 164: 209–212.
- Drummond GB, Tom BD (2011b). Statistics, probability, significance, likelihood: words mean what we define them to mean. *Br J Pharmacol* 164: 1573–1576.
- Drummond G, Vowler S (2011a). Data interpretation: using probability. *Br J Pharmacol* 163: 887–890.
- Drummond GB, Vowler SL (2011b). Show the data, don't conceal them. *Br J Pharmacol* 163: 208–210.
- Drummond GB, Tom BD (2012). Presenting data: can you follow a recipe? *Br J Pharmacol* 165: 777–781.
- Drummond GB, Vowler SL (2012). Different tests for a difference: how do we do research? *Br J Pharmacol* 165: 1217–1222.
- Fisher RA (1938). Presidential address to the first indian statistical congress. *Sankya* 4: 14–17.
- Medawar PB (1964). Is the scientific paper fraudulent? *Saturday Rev* (1 August 1964): 2: 42–43.
- Ramachandran VS, Blakeslee S (1998). *Phantoms in the Brain*. HarperCollins: New York.

Response to 'When biostatistics is a neo- inductionist barrier to science'

DOI:10.1111/j.1476-5381.2012.02034.x

We can understand some of the sentiments expressed by Dr Ashton and agree that classical statistics considered in the early perspectives in the series are indeed inappropriate for some laboratory experiments. Our defence is that we addressed first the topics we believed were known to our readers, those that are still widely used and misused by scientists. Later perspectives introduce more relevant materials. We do not carry a particular torch for Fisher and his post-mortem, but we quoted him to make a specific point: one can usually tell, in retrospect, why the desired result cannot be 'obtained' from the data, and this is often because the experiment has been badly designed. In the case of good (or appropriate) design, as with beauty, appreciation is in the eye of the beholder. In statistics, as in art and science, opinions differ, often widely (Fisher certainly disagreed with many other statisticians). Most statisticians can recount horror stories of being presented with data from experiments that could have

been resuscitated if therapy had been administered sooner. Far from being purveyors of confidence born of hindsight, practising statisticians have experience of past errors and triumphs and can see problems looming before they happen: just as can good scientists or doctors, most of the time. They can also see if a proposed study is futile, particularly if they are also versed in the science.

However, these perspectives (Drummond and Vowler, 2011a,b, 2012; Drummond and Tom, 2011a,b, 2012) have not been written by statisticians to attack scientists. One of the authors is an amateur physiologist and an even more amateur statistician, but has picked up his pen because even to him, it is evident that many scientists misuse statistics. We argue, in a later perspective, that 'the vast majority of lab studies are not based on random samples from populations' (Drummond and Vowler, 2012), and we introduce more recent and appropriate means of analysis for such studies. Old methods and habits die hard: the *T*-test may not be best. The same author, an anaesthetist, found a physiology group using gallamine as a muscle relaxant, despite its evident disadvantages for the study in progress 'because that's what we always use'. He did not ask about the statistics but suspected that he would have got the same answer. Seeking expert advice does no harm. Indeed, an expert collaboration between a statistician and a scientist is a 'dream team' if it is based on trust and respect (Berry, 2012).

We doubt that analogy is the best form of argument, neither the failed experiment nor the engineering analogy. Engineers are schooled in the principles involved in putting together a structure. The principle 'simplify and add lightness' attributed to Colin Chapman, famous designer of the Lotus sports cars, indicated his strong grasp of theory. Both a ruined project and a crashed car suggest poorly applied principles, or perhaps pushing the limits where the principles no longer hold. Some said that Lotus cars should be able to stick to the ceiling if they got up to speed! The engineer and the scientist are trained to think in terms of the relevant principles, and the associated physical laws. Not many engineers, as far as we understand, build aircraft that do not stay in the sky. If they did, we suspect that they would not ask a statistician why. Both theory and experiment are needed: the almost entirely theoretical airfoil developed by British mathematical theory proved substantially inferior to the practically tested ones used in German aircraft during the First World War (Anderson, 1997). British pilots had to be substantially more skilful to stay in the air.

Contrary to Dr Ashton's suggestions, most statisticians are firmly rooted in the real world and know that data are often dirty, messy and unreliable. Noise, iteration, error and unpredictability: these are the statistician's stock in trade: far removed from the 'inflexible inductionist schema', they are accused of adopting. How else would modern statisticians be using expressions such as Monte Carlo, bootstrap, the coupon collector's test and censoring? Modern statistical theory is more than equal to dealing with modern problems, and the agenda is set by the problems presented.

Medawar objected to the moulding of a paper to the 'story' of a standard hypothesis-experiment-conclusion sequence. Few of us would care to plough through an account that says 'first we had this idea, and that didn't work, so we tried that, and that sort of worked, but we didn't take enough samples, so we did a bit more. . . .' The account is sanitized to 'We hypoth-

esized this, did that, and the conclusion then is often "these results are highly unlikely if our samples had been taken from the same population."' This is often a gross simplification, not necessarily truthful, and sometimes misleading, but it is a formula that most people have so far agreed to accept, although 'effect' is creeping in as a useful concept. Bolstered by old-fashioned and perhaps inappropriate statistical analysis, we agree the classic paper's structure becomes even more of a travesty. It can be even more galling if these are truly the views of funding agencies. Our experience is otherwise; funding agencies are usually pragmatic and often well informed in their statistical judgement, perhaps more than most journals.

To quote a very recent statistician's comment: 'A reliance on basic biostatistical methods by many scientific researchers translates into substandard approaches to data analysis still appearing in many published studies. A tendency for researchers to then reapply these methods to their own data without first checking their validity helps perpetuate the errors' (Woodman, 2012).

In summary, statisticians are on our side, or better still statisticians and scientists are on the same side, but we can all do to improve, and collaboration will help.

Gordon B Drummond¹ and Sarah L Vowler²

¹*Department of Anaesthesia and Pain Medicine, University of Edinburgh, Royal Infirmary, Edinburgh, UK, and* ²*Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK*

References

- Anderson JD (1997). *A History of Aerodynamics and Its Impact on Flying Machines*. Cambridge University Press: Cambridge.
- Berry DA (2012). Statisticians and clinicians: collaborations based on mutual respect. <http://magazine.amstat.org/blog/2012/02/01/collaborationpolic/> accessed 10 April 2012.
- Drummond GB, Tom BD (2011a). How can we tell if frogs jump further? *Br J Pharmacol* 164: 209–212. PubMed PMID:21884044; PubMed Central PMCID: PMC3174400.
- Drummond GB, Tom BD (2011b). Statistics, probability, significance, likelihood: words mean what we define them to mean. *Br J Pharmacol* 164: 1573–1576. PubMed PMID:22022804; PubMed Central PMCID: PMC3230804.
- Drummond GB, Tom BD (2012). Presenting data: can you follow a recipe? *Br J Pharmacol* 165: 777–781. PubMed PMID:22276976; PubMed Central PMCID: PMC3312475.
- Drummond GB, Vowler SL (2011a). Show the data, don't conceal them. *Br J Pharmacol* 163: 208–210. PubMed PMID:21501140; PubMed Central PMCID: PMC3087125.
- Drummond GB, Vowler SL (2011b). Data interpretation: using probability. *Br J Pharmacol* 163: 887–890. PubMed PMID:21651509; PubMed Central PMCID: PMC3130936.
- Drummond GB, Vowler SL (2012). Different tests for a difference: how do we do research? *Br J Pharmacol* 165: 1217–1222. PubMed PMID:22324388.
- Woodman RJ (2012). Ordinary least products: let's welcome back an old friend. *Clin Exp Pharmacol Physiol* 39: 325–326.