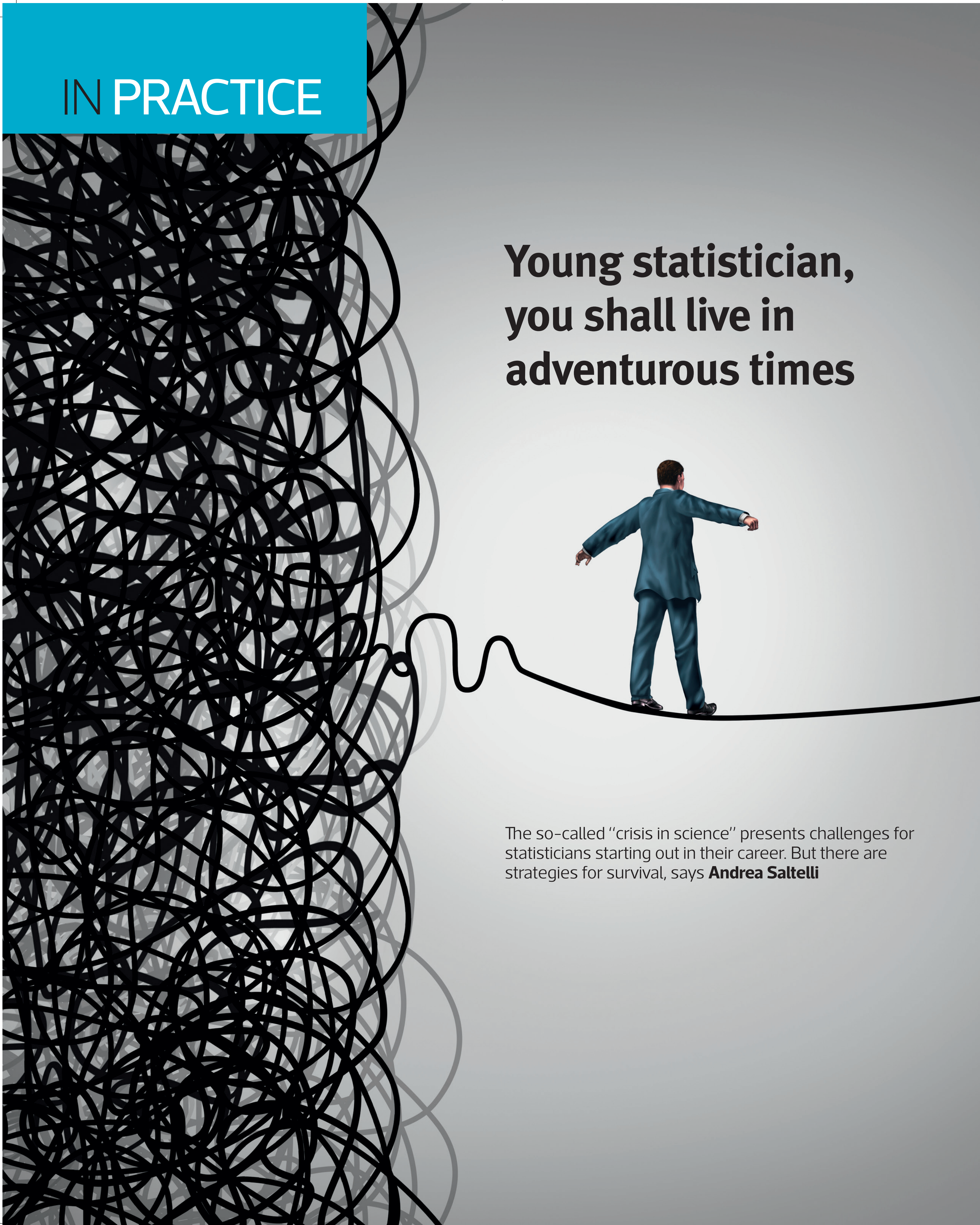
Young statistician, you shall live in adventurous times



The so-called "crisis in science" presents challenges for statisticians starting out in their career. But there are strategies for survival, says **Andrea Saltelli**

Andrea Saltelli is at the European Centre for Governance in Complexity, a joint undertaking of the Centre for the Study of the Sciences and the Humanities, University of Bergen, and of the Institut de Ciència i Tecnologia Ambientals, Universitat Autònoma de Barcelona. His main disciplinary focus is on sensitivity analysis of model outputs, and on sensitivity auditing. He tweets @andreasaltelli, and is online at andreasaltelli.eu

A couple of years ago, Philip Stark, Associate Dean of the Division of Mathematical and Physical Sciences at the University of California at Berkeley, invited me to give a commencement talk to young statisticians. Commencement speeches in previous years had been given by John Ioannidis and Nassim Nicholas Taleb, two practitioners who – in their respective styles and according to their respective interests – have been acute critics of current methodological fallacies, be it in the test for significance or in the inappropriate use of the normal distribution.

With methodological fallacies being a subject close to my own heart, it was an easy decision to continue in the same vein and to talk from my own perspective of what I saw as the emerging crisis in the quality assurance of science,[1] and specifically the role of statistics in that crisis. For it appears that today, now more than ever, statistics finds itself at the centre of controversy, and finding a way out of this controversy will fall on the shoulders of young statisticians.

What is the problem?

At first glance, the future looks rosy for those statisticians starting out in their career. They can expect to play a central role in providing a great deal of the scientific evidence that is called for to help address the critical and urgent societal challenges of our times. From climate change to genetically modified organisms, practically every policy question that demands scientific input ultimately relies on statistics and statistical methods.

Yet at the same time, statistics is held accountable for the quality of these scientific outputs, and where there are problems – say, where a result cannot be reproduced by rerunning an experiment or analysis – the fault is often laid at the door of statistics.

In discussing the problem of low reproducibility in experiments, for example, *New Scientist* magazine heaps plenty of blame on "dodgy statistics" and the "statistical sausage factory".[2] Here, one can reasonably argue that statistical skills are only part of a repertoire of crafts whose transmission mechanism seems to have jammed, and that the statistical community – in this case, the American Statistical Association – has taken swift action to issue an important statement on one particular area of concern (the appropriate use of *p*-values) to try to correct any errors and misconceptions.[3]

Poor statistical methods – or, the poor application of such methods – are a genuine cause for worry, of course. But the proper use of any statistical method calls for more than just the straightforward application of a recipe. Still, it is alarming to read that "the majority of preclinical cancer papers in top-tier journals could not be reproduced" and that poor statistics may have something to do with it, especially when, according to John Ioannidis,[4] laboratory scientists apparently cannot properly count false positives and false negatives. Ioannidis knows, of course, that more is involved than just poor methods. He observes that "a research finding is less likely to be true when … there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance".

In the view of Philip Mirowski,[5] a historian of science and economics, this crisis of quality assurance is primarily down to the rise of neoliberal ideology in all walks of life, including the management of science. In Western democracies, from the 1980s onwards, neoliberal policies led to increasing amounts of research being privatised. "Knowledge" as a monetised commodity replaced "knowledge" as a public good. In-house research labs of major corporations were closed, and research was contracted out to universities, which began to look more and more like commercial outfits. In a subsequent phase, industrial research was outsourced again to even cheaper, contract-based private organisations working on a short leash.

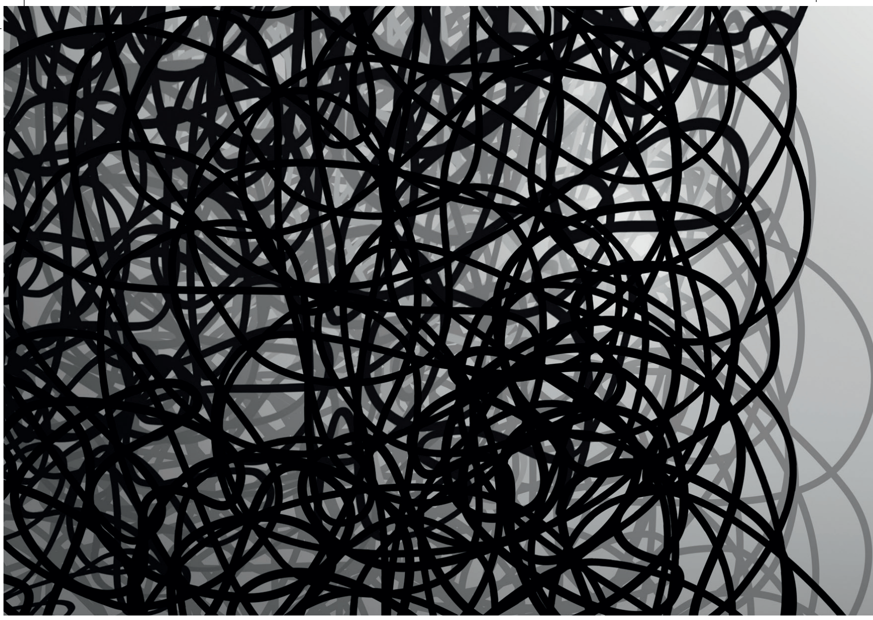
In the two years that have elapsed since my talk at Berkeley, I have come to agree with Mirowski. It had already been argued in the 1970s (by, for example, the philosopher Jean-François Lyotard) that the legitimacy of science would come under pressure as science became an instrument of profit and growth.[6] In 1971 another historian of science, Jerome R. Ravetz, published a prescient work[7] in which he foresaw and described the impending crisis in the quality control system of science in impressive detail. In Ravetz's view, science is a social activity, and he correctly predicted that the transition from "little" to "big" science would change both the social fabric and the ethos of science. "Little" science operated in restricted communities bound by personal acquaintance, by shared norms, and by a willingness to enforce them. "Big" science, or techno-science, is a vast enterprise where scientists value one another through impersonal metrics of citation and impact. It should not come as a surprise that the quality arrangements of older times would come under strain in today's realities. Ravetz also noted that reform would be hard: the quality of a scientific work is too delicate a matter to be disciplined with a set of formal rules, as any system can be gamed.

The next generation

The evolution of science – from "little" to "big", and from pubic asset to privatised commodity – has not only changed our quality assurance systems, but also challenged the trust that people have in the veracity of the evidence they are presented with. The public watch as science is wielded as a weapon in debates over policy, with "facts" being used to support contradictory courses of action favoured by opposing factions. As a citizen it can be hard to know who or what to believe. Rivals can easily sow doubts about the impartiality of evidence presented by their opponents. Attempts to simplify and explain complex issues can draw criticism that uncertainty is being overlooked. When uncertainty *is* acknowledged, it may be seized upon and inflated – to argue, for example, that if the evidence is vague, why should we regulate? Similarly, uncertainty may be downplayed: policy-makers may try to convince the public that enough evidence has been accumulated so that any doubts can be set aside.

Considering all this, what advice can we give to young statisticians to help them navigate the troubled waters of the twenty-first century?

First and foremost, I would insist that it is not wrong to take sides in any debate. Social scientists would say that more and more of



All scientists should resist hubris; for a statistician, this implies being conscious that risk is not the same as uncertainty, which is in turn not the same as indeterminacy

the issues we face have proved to be "wicked problems", the kind that are deeply entangled in webs of barely separable facts and values – whereas science was originally (and, for many, is still) predicated precisely on a strict separation (or demarcation) between the two.

Therefore, it is better to be openly passionate about social and environmental issues than to hide principles and convictions under a veil of false objectivity. We should drop the illusion that science "speaks truth to power" and accept that various models of the relationship between policy-making and science or statistics exist,[8] including:

- the rational-positivist model (good-quality statistics underpin good policies);
- the discursive-interpretive model (statistics contribute to a process of framing and focusing on a given issue among the many competing for public attention);
- the strategic model (statistics are recruited by parties competing for a given constituency).

The co-existence of all these models should be borne in mind, in contradistinction to any simplistic adherence to the precepts of the evidence-based policy model. Is it even possible, in practice, to disentangle evidence-based policy from policy-based evidence? I personally think not. In the former, policy descends from the evidence; in the latter, evidence is assembled in order to justify a policy decision. One is the flip side of the other, and they are impossible to separate – just as it is impossible to extricate facts from values at the interface between science and policy.[1]

This is best illustrated with an example concerning the first statistics ever to be collected in the field of research and development. Long before patents, citations and university rankings were ever studied, someone had the bright idea of counting the number of offspring of reputed living scientists to show that, since intelligent people reproduce less, mankind was condemned to stupidity.[9] Clearly the desired policy inference drove the collection of evidence. This example reminds us that statistics as a discipline has to take its share of responsibility for eugenics, as we know from the work of statisticians Francis Galton and Karl Pearson – both founders of modern statistics and sincere believers in the thesis that "the best breed the best and the worst breed the worst",[4] according to historian of science Ian Hacking.

As described in Hacking's book, *The Taming of Chance*[10] (see below), for these fathers of our craft the advocacy of eugenics was a mission which involved the creation of laboratories and the institutions of new journals. Galton founded the Anthropometric Laboratory at University College London, while Pearson created both *Biometrika* and the *Annals of Eugenics*. Both men believed in applying "value-neutral science and statistical techniques to the issues of the day", but such an approach could be successfully employed to sort out different classes of felons, the murderers from the less violent, to distinguish army officers from their soldiers, and, of course, to recognise Jews. This example and others throughout history show that scientists – including statisticians – are not always upholders of the true and the good.

All scientists should resist hubris; for a statistician, this implies being conscious that risk is not the same as uncertainty, which is in turn not the same as indeterminacy. Even with the power of Bayesian methods behind them, results should always be interpreted cautiously and never presented as being free of the burden of genuine uncertainty. A recent argument in this direction, which young statisticians can read, comes from Gerd Gigerenzer and Julian N. Marewski,[11] who remind us of the need to use

The Taming of Chance

Young statisticians should read Hacking's book *The Taming of Chance* – it is a page turner. This tells how the world became "numerical" between the eighteenth and nineteenth centuries, well before the "big data" era of the current generation. One can find in this book the fascinating story of Leibniz, "philosophical godfather of Prussian official statistics" (p. 18), preparing for Prince Frederick of Prussia in 1700 a list of 56 categories to "measure the power of a state". These were nothing less than the first scoreboards, presenting, for example, figures on marriageable girls, able-bodied men capable of carrying arms, disease rates and child mortality. Leibniz was also the first to propose the creation of a statistical office. It is noteworthy that as early as 1745 Jews were being counted separately in Prussian statistics, one chapter of the book being devoted to the relation between statistics and anti-Semitism.

Hacking also tells how, in the nineteenth century, probability won an epistemological war (epistemology being, in simplified terms, the study of how we go about knowing, and how we decide that we know what we know). In the process, probability became king in adjudicating on the credibility of evidence. As a result, we now look at facts mostly through the lens of statistics, in contrast to earlier, pre-Enlightenment times when chance was equated with superstition. In Hacking's view, the victory of probability was metaphysical, epistemological, logical and ethical – producing the "imperialism of probability".

statistics with judgement, and not as "recipes". They warn in particular against "the Idol of a Universal Method for Scientific Inference" and note that "[t]he application of statistics to science is not a neutral act". The reference to idols comes from none other than Francis Bacon: "In Bacon's view, it is better to have no beliefs than to embrace falsehoods, because false idols block the way toward enlightenment."[11]

A post-normal world

In the spirit of the times, James Zidek, a statistician at the University of British Columbia, suggested in a 2006 editorial for the *Journal of the Royal Statistical Society*,[12] that statisticians should become "post-normal". Post-normal science (PNS) is a concept developed by Silvio Funtowicz and Jerome Ravetz in the early 1990s[13] to help deal with situations at the science–policy interface in which the facts are uncertain, the values in dispute, the stakes high and the decisions urgent.

Statisticians should appreciate the variety and power of the methods at their disposal without being overawed and seduced by them

The developers of PNS had in mind Thomas Kuhn and his distinction between normal science and paradigm shifts in the development of scientific theories. In this way normal science would correspond to the classic vision of science as used to understand and domesticate nature, while something different – PNS – would be needed to remedy the damage inflicted by man and his technologies on nature.

Zidek notes that: "There is plenty of scope here for a post-normal statistical scientist!" I would concur. In circumstances where the evidence is contested, differences in values are substantial and the distribution of information, costs and benefits is unequal, the legitimacy of the policy-making process must be carefully established. Here the PNS approach of involving forms of extended peer communities – where different research disciplines as well as different worldviews confront one another in deliberative settings – may represent the only basis for sustainable dialogue and progress.

Zidek concludes his editorial thus: "A partial solution lies in ensuring that statistical education is sufficiently broad to acquaint statistics graduates with the challenges that are presented by PNS while equipping them with the enhanced skills that are needed to cope with them and even selectively to take advantage of the opportunities. In particular, the statistical consulting sequence that is commonly found in statistics graduate programmes might be expanded to include multidisciplinary meetings where a multiplicity of legitimate views are presented in an adversarial context."

Zidek was correct in identifying a need for scientists to be able to defend the quality and reproducibility of their work against the claims of other experts.[12] At the same time, his prescription for young statisticians falls short of what the times call for. In order to do our job as statisticians to the best of our ability, it will not suffice to train in the art of rhetoric for multidisciplinary or multi-stakeholder settings.

Philip Stark, my host at Berkeley in 2014, suggests a pledge (bit.ly/2fjKjke), specific to statisticians, among the elements of which is an injunction not to produce or to review any piece of inference for which the data are not made entirely available. The pledge – which anticipates similar initiatives being discussed by scientific institutions – has many elements which one might call Mertonian, after the scholar Robert K. Merton who first applied the scientific method to study the workings of science.

My own pledge[14] for responsible quantification focuses on an appreciation of the frames and assumptions hiding behind the crystalline purity of numbers, and insists on "a licence not to quantify" when the conditions for responsible quantification are not met. Statisticians should appreciate the variety and power of the methods at their disposal without being overawed and seduced by them. Finally, for young statisticians to fully appreciate the need for modesty and caution, training for post-normal times should include elements of the history, philosophy and sociology of science.

This is not as innocuous a prescription as it sounds. It implies important cultural transformations: the "unlearning" of the deeply ingrained conception of the privileged role of science in society; the abandoning of a blind faith in formal models; the genuine acceptance of our fallibility as scientists.

These ingredients will help young statisticians to live through the present disputes on the quality of science and its role in society. Statistics, with its engagement with big data and the algorithms that pervade people's ordinary lives (see next article), will increasingly find itself on the witness stand, if not in the dock. Challenges will need to be faced, and times will certainly be adventurous – especially for those willing to defend the quality of the craft. ■

Note

This article is adapted from a commencement speech for statisticians given at the University of California at Berkeley on 19 May 2014.

References

1. Benessia, A., Funtowicz, S., Giampietro, M., Guimarães Pereira, A., Ravetz, J., Saltelli, A., Strand, R., and van der Sluijs, J. (2016). *Science on the Verge*. Tempe, AZ: The Consortium for Science, Policy and Outcomes at Arizona State University.
2. van Gilder Cooke, S. (2016) The unscientific method. *New Scientist*, 16 April.
3. Wasserstein, R. L. and Lazar, N. A. (2016) The ASA's statement on *p*-values: Context, process, and purpose. *American Statistician*, **70**(2), 129–133.
4. Ioannidis, J. P. A. (2005) Why most published research findings are false. *PLoS Medicine*, **2**(8), 696–701.
5. Mirowski, P. (2011). *Science-Mart: Privatizing American Science*. Cambridge, MA: Harvard University Press.
6. Lyotard, J.-F. (1979). *La condition postmoderne. Rapport sur le savoir* (Chapter 10). Paris: Minuit.
7. Ravetz, J. R. (1971) *Scientific Knowledge and its Social Problems*. Oxford: Clarendon Press.
8. Boulanger, P.-M. (2007) Political uses of social indicators: Overview and application to sustainable development indicators. *International Journal of Sustainable Development*, **10**(1–2), 14–32.
9. Godin, B. (2010) The culture of numbers: From science to innovation. Communication presented to the Government-University-Industry Research Roundtable (GUIRR), US National Academy of Sciences, Washington, 21 May. www.csiic.ca/PDF/GUIRR.pdf
10. Hacking, I. (1990) *The Taming of Chance*. Cambridge: Cambridge University Press.
11. Gigerenzer, G. and Marewski, J. N. (2015) Surrogate science: The idol of a universal method for scientific inference. *Journal of Management*, **41**(2), 421–440.
12. Zidek, J. (2006) Editorial: (Post-normal) statistical science. *Journal of the Royal Statistical Society*, *Series A*, **169**(1), 1–4.
13. Funtowicz, S. and Ravetz, J. (1993) Science for a post-normal age. *Futures*, **25**(7), 739–755.
14. Saltelli, A., Giampietro, M. and Ravetz, J. R. (2016) Decalogue of the diligent quantifier. A pledge. bit.ly/2eglhmb