מנא, מנא, תקל ... (Mene, Mene, Tekel...)

Measure what is measurable, and make measurable what is not so.
Galileo Galilei

Performance metrics are nothing new: whether we take the story of Belshazzar's feast as history, of myth or as allegory, it is clear that the ideas of one's achievements being measured quantitatively, and of these measurements carrying consequences, is ancient. The writing, as the saying has gone since that feast, is on the wall. Science is full of measurements but, ironically, one of the most difficult things to measure accurately is the performance of science itself, and the impact it has on human and animal lives and behaviour.

I have recently been prompted to engage with the problem of measuring scientific achievement as part of my work as Deputy Chair of the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs). The NC3Rs is not a pressure group but a government-supported agency that funds and disseminates development of alternatives to animal use. For an organization like this, meaningful measures of impact are critical for discovering which approaches work best for which areas of science and for which type of laboratory. Since the ideal impact of research funded by NC3Rs achieves several different things at once — improving science along with reducing animal suffering in its cause — finding appropriate measures is not a trivial task. Guess who has been asked to lead a task group to develop them...

The most obvious method for tracking 3Rs impact — just count all the non-human animals used each year and interpret a fall in this number as success, and a rise as failure — is deeply flawed for several reasons. To begin with, the main '3Rs' aim is to reduce animal suffering, either by not using (as many) animals, or by using animals in a way that harms them less, or by using 'lower' animals (for example, fruit-flies) in place of 'higher' (for example, dogs). Just counting the total number of animals is a poor measure of suffering because it would treat a mouse in a harmless breeding programme equally with an immunodeficient mouse inoculated with a metastasizing human cancer. Similarly, simple counting would treat a fruit-fly, a mouse and a chimp equally, though most people do not take that view: when the dog *Laika* was launched on Sputnik 2 in 1957, for example, the

RSPCA and NCDL organized a minute's silence in the UK: no public protest was made when fruit-flies had been launched on the suborbital V2 *Blossom I* ten years earlier. At the very least, therefore, a useful counting system would need to involve the raw count multiplied by some factor that measures severity and by some species-specific factor connected with how much (we think) a species suffers from a particular treatment.

Even this idea of a corrected count, which includes factors for severity and species, is unlikely to be a very good measure of impact except in a very local context. Animal experiments are typically time-consuming and expensive and are frequently the rate-limiting step in research, especially in large scale screens of drugs or mutants. Reduction of animal use per candidate molecule, for example by using *in-vivo* imaging to follow a process through time in one animal, rather than sacrificing one of a group of animals at different time-points for the same information, will save a great deal of time and money, perhaps reducing both to 20% of their former values. The lab blessed with this saving will not just go off to play golf on the 4 days a week they have saved: they will now be able to screen 5 times as many molecules. In this case, total animal use (and suffering) will not have changed: what will have changed is that 5 times as much science is done per animal. Our corrected count therefore needs to be divided by some standard measure of benefit (scientific papers published? Drugs patented? Drugs eventually licensed?).

Even if we agree on factors for this now quite complex metric – and agreeing on them will not be easy – is a 'grand-total' figure really what we want? When we worry about the treatment of humans by repressive regimes, we usually focus on the fate of the few who suffer most: we would not consider a small improvement in the lot of the masses to be able to cancel out, arithmetically, a worsening treatment of the most abused few. If we reject a simple-minded calculus of suffering for humans, can we justify it for non-human animals, or should we again focus on improving the lot of the ones who suffer most?

Quantitative approaches, though natural to scientists, are not the only methods for demonstrating impact. Our colleagues in the sociology have developed many useful qualitative approaches and some of these – for example case studies – are very well suited to demonstrating advance in the 3Rs. Indeed, the narrative nature of a case study, especially one that illustrates improved science as well as 3Rs results, means that it can be directly useful in inspiring others to make a similar change in their own labs. Case studies are therefore not only a demonstration of impact: they can also be

the means of generating more. Case studies also have the possibility of being extended forward in time, from the initial description of impact in the practice of a lab, to the eventual outcome of the science in, say, human health, as well as the changing practice of other labs. This approach of case studies is encouraged in university science by the Research Evaluation Framework (REF): if future versions of the REF were to include a specific section for 3Rs impacts, there may be an almost instant increase in academics' awareness of the possibility of making a real contribution (whether they are funded by the NC3Rs or not).

In our bench science, we seldom rely on only one tool. In the “science” of impact metrics, we also need to use a selection of tools, some quantitative, some qualitative, some local, some national, some measuring the effectiveness of funding agencies, some of individual labs, some measuring the 3Rs impact directly and others capturing the improved science. None of this is particularly easy: good science seldom is. Please wish us luck in finding sensible ways to identify and publicize what all those who work with NC3Rs support have achieved.

Jamie Davies

Edinburgh

June 2013

Links:

The NC3Rs: http://www.nc3rs.org.uk/