Mene, Mene, Tekel… (part 2)

Do what you love, and do it well - that's much more meaningful than any metric.
Kevin Systrom

In a previous blog post in this series, with the same title, I reflected on the difficulty of assessing the success of different strategies being pursued by a research funding body on whose Board I served at the time. The blog considered different quantitative measures but ended up with more of a focus on qualitative approaches and the organization's formal report, when it was finished and presented to Parliament, made substantial use of case studies as well as numbers. I must stress that I am not claiming significant credit for that fact: the committee I chaired benefited greatly from very wise advice from a number of senior civil servants, industrialists and representatives of charities.

This business of metrics is not restricted to organizational levels. Enthusiasts for quantitative measurements have for a long time been trying to apply them to measure the 'quality' of finer and finer divisions of the scientific enterprise, some extending the idea to quantitative performance metrics for even individual scientists. In some cases, the numbers, however crass and devoid of meaning, have been used to determine career progression or even survival: I know a couple of excellent scientists in other institutions who have done amazing, inspiring work who have lost their jobs on the basis of metrics. Almost everyone I know finds the whole idea of crude quantitative measurement horribly misplaced – the kind of pseudo-scientific number worship of the kind mocked by the mathematician-satirist Tom Lehrer (listen to his song 'Sociology': see the links at the end). The problem is that it is difficult to speak up against the measurement system without sounding as if one has a severe case of sour grapes through not getting high scores. Well, it so happens that right now, for reasons outlined at the end of this blog post, part of the work of my lab is gaining literally world-leading quantitative metrics so there is probably no better time for me to write a blog that argues that all such scores mean very little individually and should, for most purposes, be ignored.

There are two kinds of problem with measuring the quality of a scientist, or any other kind of creative person, quantitatively and in real-time. One kind is 'shallow' and easy to explain, and the

other is deeper and harder to capture in words, but I'll try.

Starting with the shallow reason: what exactly should one measure? Grant income is clearly silly, although much-used. In a sane world, someone who can achieve a goal for little money would be rewarded much more than someone who uses vast resources in achieving the same thing, yet many people jump to the conclusion that a scientist with lots of grant funding must be more significant than someone with little. The number of papers produced per year is also silly, because measuring success that way simply encourages people to divide their discoveries into little sub-discoveries that can be published separately. This was recognized a long time ago, so focus turned from numbers of papers published to the number of times each paper is cited by other authors, the idea being that a paper that is read and cited by many people must be important and one apparently ignored by other people is not. Citation rates are currently the dominant measure of 'academic impact' (not my phrase, but that of bureaucrats), either of papers or, in aggregated form, of journals (as the 'Journal Impact Factor') or of individuals (for example, as the 'H' factor). At first the use of citation rates – the number of times other people make use of a piece of work – may seem a reasonable measure of quality. There is of course the problem that someone working in a popular field in which many papers are published can be cited far more times than someone doing work of the same quality in an obscure area of science. There have been various proposals to apply statistical corrections to account for this. But, even if they work, is a popularity contest a good way to judge anything other than popularity itself? Would you really want to measure the quality of a musical composition by the number of times someone uses it ('sorry John Tavener: sales metrics show that Gareth Gates' work completely outclasses yours'), or of poetry by the number of times someone sends it to someone else ('sorry William Blake, but metrics show that Purple Ronnie has a much higher cultural impact than you')?

Real scientific advances – not the incremental stuff that is the everyday output of most of us, but the world-changing ideas – tend not to be recognized immediately and, even when they are, it takes a while before people work out how to use them. A recent example of this is provided by the 1987 paper by Yoshizumi Ishino that announced the discovery of clustered repeats in bacteriological genomes. This was cited a mere eight times in the five years following its publication – a modest score indeed. Yet these clustered repeats are the basis of the CRISPR system that has become one of the 21st century's most dramatic and important biotechnologies. It just took a while for this and other basic science discoveries to be connected into a coherent story, and a while longer for them to be

harnessed as tools. The fact that Ishino's paper was barely cited in it first few years did not in any way make it unimportant or of low quality. It was an excellent piece of work that laid the foundations for technologies that, through agricultural as well as medical applications, may become a dominant theme of humanity's future. This is just one example of many, many important advances that took off slowly, often after the initial researchers had retired or died.

The deeper problem is with the whole notion that the real impact of anything creative can be measured in real time (ie within a few years of the work being done). Human cultural development is incredibly non-linear, and it is all too obvious that even those who devote their lives to studying history are incapable of using their knowledge to predict the future. Looking backwards from where we are now, it is possible to construct stories of cause-and-effect, or even of 'historical inevitability' if the historian has Marxist leanings but, when we are living in the midst of things, it is very difficult to see what is important. Although we like telling stories of the lone genius, in reality developments that have real impact on culture emerge from a whole ecosystem of science, technology, politics and economics. Taking of measure of which scientists alive now are making the greatest long-term impact may be possible looking back from 2117, but it is not possible now. And even then, when our successors have looked back and named a few names, there will be all of the other un-named researchers, publishers, microscope-builders etc. who all had to be there to make an ecosystem that could let the discovery be made. Trying to find a quantitative measure of which scientists are most significant reminds me of a question I had from a child at a recent outreach event: *which cell type is the most important?* Yes, we could invent some mad system of metrics that could show that a cadiomyocyte is worth 40 points and a Sertoli cell only worth 6 but, long term, humanity needs them all. And that is why use of quantitative metrics is so toxic: whatever system is used to give scientists scores like base-ball players, it will limit diversity of approaches to science and favour just a few ways of doing things – the ways that generate the best scores. Surely, more than ever, we need to explore the world in as many different ways as possible.

I mentioned at the start of this piece that the lab is currently finding itself in some metrics 'lime-light'. I'll tell the story of how we have found ourselves there briefly because it illustrates one additional problem of the way metrics are currently generated by computer systems: they do not measure what people think they measure.

Various bibliometric websites, and what might be called “scientists' social media” sites such as

Researchgate, send occasional unprompted e-mails to the authors of papers to tell them when one of their papers has reached a 'citation milestone' (such as being cited 1000 times). Most of us have, over the years, had occasional *you have been cited more times this month than anyone else on your department* e-mails sent unbidden by automated servers. Last year, though, I seemed to get a note like this pretty much every month which, given the excellent quality of the other labs in this building, simply made no sense. Last month I was sent a link to a Clarivate Analytics site (the site that took over metrics from Thomson-Reuters), and this site listed a paper from this lab as being the world's top “fast-breaking paper” in the whole of biology and biochemistry! What was this paper? Had we discovered an amazing new gene, or an amazing biotechnological method, or a wonder-drug? No: it was not an original discovery paper at all but was a simple, descriptive review of the GtoPdb database of drugs and targets, run from this laboratory. The story of how this very ordinary paper is coming to be a statistical outlier rocketing to the front of bibliometric counts is an interesting practical illustration of why even someone unpersuaded by the general arguments I made at the beginning of this blog entry should still take bibliometric data with a large measure of AnALaR NaCl.

I am very grateful to my colleague Dr Chris Southan, the lab's chemoinformatician and a man who knows his way round databases as a London Cabbie knows his way round the streets of the city, for his analysis of what has happened. One of the most important uses of our GtoPdb database is as an online databank to support publications in the British Journal of Pharmacology and the British Journal of Clinical Pharmacology. When research papers in these journals mention a drug or the molecular target of a drug, the editorial systems of the journal insert a table that contains hyperlinks from the paper to the relevant entries in our database for those drugs and targets. That way, anyone reading a paper on screen can click on the hyperlinks to remind themselves of the properties of the drug or target while they are reading – a very useful feature. The table of links contains a standard legend explaining what it is and how to use it and, as a quick way of explaining the GtoPdb resource, the legend cites our review of the database. This means that pretty much every paper in BJP and BJCP includes this citation to our paper. These are not 'real' citations, in the sense that the authors choose to put them in – they are simply a byproduct of the hyperlinking and data management mechanism provided by the journal. They are, however, evidently picked up by the software that bibliometric services use to make their counts. We *do* have plenty of real (author-generated) citations too, of course, from all sorts of journals, but their already pleasing number is being dwarfed by the purely mechanical kind from the tales of links. So we seem to have rocketing

metrics because other people's software counting citations is conflating deliberate, author-written citations from ones present for technical reasons. Neither we nor the journal editors and publishers (whom we told about this some time ago) can do anything to stop it, without breaking what is a very useful system.

My experience of becoming a temporary “citation superstar” for entirely spurious reasons has made my view of bibliometrics even more jaundiced than it was before. When I am forced into the position of having to judge a fellow scientist, for example because I am on an appointments committee, I will continue to ignore metrics and instead read her actual papers and come to a judgement about her scientific creativity from reading her work, and not from unreliable measures of how much attention it happens to attract.

Jamie Davies
Edinburgh
January 2017

Links:

- Tom Lehrer's sociology song - https://www.youtube.com/watch?v=mB97Qe2D4V0
- Our database review paper that is picking up so many 'citations' for spurious reasons: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702778/
- The 'fast breaking papers' web site http://ipscience.thomsonreuters.com/blog/fast-breaking-papers-december-edition/?category=science-research-connect (this website may become over-written in time with next year's data – a screenshot of the relevant edition appears below as evidence I am not making all this up).
- Chris Southan's own blog about this issue: https://blog.guidetopharmacology.org/2017/02/03/reference-citations-in-bjp/

Jennifer Minnick
Research Analytics Proposition Manager

Fast Breaking Papers December Edition

POSTED: DECEMBER 12, 2016 | UPDATED: 3 WEEKS AGO

*Essential Science Indicators*SM (ESI) from Clarivate Analytics lists Highly Cited Papers in 22 broad fields of science. These papers constitute the top 1% of papers in each field and each year. The lists are updated every two months to reflect their current citation counts and also include new papers that enter the top percentile.

Here, we identify a subset of these papers having the largest percentage increase in citations in their respective fields from one bimonthly update to the next. We call these "Fast-Breaking Papers" because they represent very recent scientific contributions that are just beginning to attract the attention of the scientific community.

The following table lists the Fast-Breaking Papers in each of ESI's 22 fields for the fourth bimonthly update of 2016, which covers January 1, 2006 to August 31, 2016.

Field	Authors	Journal	Title
AGRICULTURAL SCIENCES	Zhen, J; Villani, TS; Guo, Y; et al.	FOOD CHEM 190: 673-680 JAN 1 2016	PHYTOCHEMISTRY, ANTIOXIDANT CAPACITY, TOTAL PHENOLIC CONTENT AND ANTI-INFLAMMATORY ACTIVITY OF HIBISCUS SABDARIFFA LEAVES
BIOLOGY & BIOCHEMISTRY	Southan, C; Sharman, JL; Benson, HE; et al.	NUCL ACID RES 44 (D1): D1054-D1068 JAN 4 2016	THE IUPHAR/BPS GUIDE TO PHARMACOLOGY IN 2016: TOWARDS CURATED QUANTITATIVE INTERACTIONS BETWEEN 1300 PROTEIN TARGETS AND 6000 LIGANDS
CHEMISTRY	Sheldrick, GM	ACTA CRYSTALLOGR C 71: 3-8 PART 1 JAN 2015	CRYSTAL STRUCTURE REFINEMENT WITH SHELXL
CLINICAL MEDICINE	Rasmussen, SA; Jamieson, DJ; Honein, MA; et al.	N ENGL J MED 374 (20): 1981-1987 MAY 19 2016	ZIKA VIRUS AND BIRTH DEFECTS - REVIEWING THE EVIDENCE FOR CAUSALITY
COMPUTER SCIENCE	Wang, T; Gao, H; Qiu, JB	IEEE TRANS NEURAL NETW LEARN 27 (2): 416-425 SP. ISS. SI FEB 2016	A COMBINED ADAPTIVE NEURAL NETWORK AND NONLINEAR MODEL PREDICTIVE CONTROL FOR MULTIRATE NETWORKED INDUSTRIAL PROCESS CONTROL
ECONOMICS & BUSINESS	DiMasi, JA; Grabowski, HG; Hansen, RW	J HEALTH ECON 47: 20-33 MAY 2016	INNOVATION IN THE PHARMACEUTICAL INDUSTRY: NEW ESTIMATES OF R&D COSTS
ENGINEERING	Qiu, JB; Ding, SX; Gao, HJ; et al.	IEEE TRANS FUZZY SYST 24 (2): 388-400 APR 2016	FUZZY-MODEL-BASED RELIABLE STATIC OUTPUT FEEDBACK H INFINITY CONTROL OF NONLINEAR HYPERBOLIC PDE SYSTEMS