Science and Public Policy, 34(8), October 2007, pages 555–563
DOI: 10.3152/030234207X254413; http://www.ingentaconnect.com/content/beech/spp

Evaluating universities using simple scientometric research-output metrics: total citation counts per university for a retrospective seven-year rolling sample

Bruce G Charlton and Peter Andras

We advocate a scientometric, top-down and institution-based research-assessment methodology that is based on total citations accumulated from all publications associated with a specific university during the survey period. The exercise could be done every year using a rolling seven-year retrospective sample and should be performed by at least two independent auditors. Identification of elite 'revolutionary-science' institutions could be accomplished using a metric derived from the distribution of science Nobel Prizes.

THE PURPOSE OF THE UK Research Assessment Exercise (RAE) is to measure the quality of research in UK universities, with the aim of providing central government funding to support the long-term research capability of an institution (Higher Education Funding Councils, 2007). At present, the UK RAE is a 'bottom-up' and discipline-based expert review process determined using a common set of information provided by each disciplinary unit within each university.

By contrast, we advocate a 'top-down' institution-based research-assessment methodology based on total citations accumulated from all publications associated-with a specific university during the survey period. Such a survey could be done every year using a rolling seven-year retrospective sample and be performed by at least two independent auditors. Identification of elite 'revolutionary-science' institutions could be based on a metric derived from the distribution of science Nobel Prizes.

Dr Bruce G Charlton is in the School of Psychology, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK; Email: bruce.charlton@newcastle.ac.uk; Tel: +44 191 222 6247. Dr Peter Andras is in the School of Computing Science, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK; Email: peter.andras@newcastle.ac.uk; Tel: +44 191 222 7946; Fax: +44 191 222 8232.

The authors thank Claire Donovan and Andrew Oswald for comments that helped improve this paper.

Scientometric top-down research evaluation

Our suggested usage of a RAE based on a metric of total citations from all publications associated with a specific university is an example of top-down research evaluation, using a single macro-level variable. By contrast, the current UK RAE is implicitly a bottom-up approach to research evaluation using an accumulation and average of many micro-level evaluations.

This top-down approach to research evaluation derives from the discipline of scientometrics in which the evaluation of research is seen as a science in its own right (Bush, 1945; de Solla Price, 1986; Garfield, 1977: 313–318; Leydesdorff, 2005). Scientometric analysis is therefore typically performed by observers 'outside' the system being evaluated; in contrast to the bottom-up evaluations, which are typically performed by 'peer review' of individuals with expertise in the discipline being evaluated.

Since its origins in the 1940s, scientometrics has developed its own 'system language' including information selections and a distinctive lexicon and grammar (Luhmann, 1995; Charlton and Andras,

Bruce G Charlton graduated with honours in medicine from Newcastle University UK, completed an MD at the MRC neuroendocrinology unit, and is currently a Reader in Evolutionary Psychiatry at Newcastle. He has also held academic positions in physiology, anatomy, epidemiology and public health, and has a Masters degree in English literature. Dr Charlton is author of over 100 papers and a book: *Psychiatry and the Human Condition* (Radcliffe, 2000). He is Editor in Chief of the Elsevier journal *Medical Hypotheses*. Charlton and Andras have collaborated in the production of many papers, and a book: *The Modernization Imperative* (Imprint Academic, 2003).

Peter Andras was born in the Hungarian-speaking region of Romania. He attained a BSc in computer science, MSc in artificial intelligence and PhD in neural networks from the Babes-Bolyai University. He was the founding director of the Civitas Foundation for Civil Society 1992–1998. Dr Andras worked at the University of Maastricht, before joining Newcastle University, UK; where he is now Reader in Complex Systems and Computational Intelligence. He has published in fields such as neural systems, protein interaction systems and complex social systems. This has led to his current interest in science policy and scientometrics.

2003). For this reason, the top-down scientometric procedures of research evaluation are usually completely different from the academic work that goes on within those individual academic disciplines that are being evaluated.

The current UK RAE is therefore (implicitly) an example of bottom-up research evaluation, since it uses the internal evaluation procedures of many individual academic disciplines (approximately 70 disciplines for RAE 2008). These evaluations are performed by peer-review panels of experts with the relevant disciplinary expertise, and the criteria of evaluation are very different among disciplines such as mathematics, biochemistry, English literature, social policy and the performing arts. To generate a metric for each university, these dissimilarly-derived evaluations are averaged. The validity of the summary RAE metric is therefore a bottom-up consequence of the validity of all the discipline-specific metrics that have gone into it, and of the averaging procedures.

By contrast, the validity of a top-down scientometric evaluation procedure is not a product of its constituent parts, but emerges as a consequence of its usage. The validity does not, therefore, depend on discipline-level information; but on how well the metric performs as a measure. Scientometrics, as a science, therefore evolves by the usual scientific processes, such as making hypotheses and predictions that are tested by further observations (Hull, 1988; Ziman, 2000).

A key advantage of the top-down scientometric approach is that it provides a macro-level summary measure that can be evaluated independently and objectively allowing universities to make well-informed decisions about science and research strategies and policies. In comparison, the current bottom-up approach of the RAE facilitates the burgeoning of speculative theories about how the actual evaluation process will work. This leads to ill-informed decisions on short-term presentation tactics and long-term research strategies poised to change when the actual RAE outcome emerges. Possible ways of combined application of these two approaches are discussed in other papers in this special issue (Butler, 2007; Moed, 2007).

Analogy with economics

The difference between top-down scientometric and bottom-up approaches to research evaluation is somewhat analogous to the difference between top-down macroeconomics and bottom-up microeconomics (Mankiw, 2003; Bishop, 2004).

Top-down macroeconomics examines total national economic activity in terms of variables such as taxes and interest rates; and national behaviour such as economic growth and inflation. Bottom-up microeconomics examines the behaviour of individual people and organisations in terms of variables such as money and time; and incentives such as pay and profit.

Microeconomic analysis therefore aspires to discover, understand and predict the actual economic incentives directly experienced by individuals and organisations (even when individuals or organisations may not be consciously aware of these incentives). By contrast, macroeconomics uses variables that have proved themselves useful at the national level (or, at least, more useful than the known alternatives) in terms of tasks such as monitoring, predicting and controlling the national economy.

The validity or usefulness of a macroeconomic variable is not necessarily challenged by critiques from a microeconomic perspective. For example, although raising the central bank interest rate by a quarter of a percent is a recognised (and usually effective) method of reducing inflation, this is hard to understand from the perspective of the incentives of individual citizens who would probably not be able to detect the consequences of such a small alteration in their personal finances, given the number of other and larger influences on personal finance. Also, the gross domestic product (GDP) metric generally performs well as a longitudinal and internationally comparative measure of the size of a national economy, yet GDP can easily be criticised from a microeconomic perspective as both incomplete and biased.

Similarly, our proposed research evaluation metric of the total annual citations associated with a university can legitimately be criticised from a micro perspective as incomplete and biased (REPP, 2005; van Raan, 2005). For example, citations in books may be omitted, and the total citations metric is often dominated within a university by very high citation rates in relatively few fields within the natural sciences. Nonetheless, while correct at the micro level, none of these or similar criticisms are necessarily decisive at the macro level.

A research metric should have qualities such as simplicity, transparency, objectivity, replicability, precision and sensitivity; it should also allow both longitudinal and international comparisons

What matters at the macro level is whether the total citations metric performs well in the function for which it is intended. The proposed metric aims to reveal the macro-level research performance of universities, and it works well if it does this reliably. It does not aim to reveal any specific micro-level detail of this performance (for instance, whether the performance in top physics or biology groups are the drivers of the macro-level performance), although such details that are behind a certain macro-level research performance can be analysed and possibly determined. The proposed metric in itself does not provide any specific pointers for the analysis of micro-level factors behind the macro-level measure.

Longitudinal and international comparisons

Although the UK RAE currently presents itself as a method of evaluating the whole of academic research activity, we believe that the implicit motivation of research evaluation is focused upon 'scientific' research, that is, the mathematical and natural sciences, and the quantitative social sciences such as economics. The most useful type of RAE would enable scientific research to be compared longitudinally and internationally.

The current UK RAE uses a highly complex, non-verifiable, uncheckable, evolving, bottom-up, discipline-based, peer-review process that lacks transparency. So the RAE results cannot legitimately be used to track longitudinal changes of individual universities, nor can it be used to measure the relative strength of UK universities internationally (either individually or as a national whole). Furthermore, the RAE results are not checkable or replicable, even in principle. We regard these as very significant criticisms of any discipline-based peer-review system of research evaluation when used for measuring institutional scientific research performance.

We would argue that it should explicitly be accepted that the primary implicit focus of research evaluation is scientific. Indeed, scientometrics originated in the realm of evaluating scientific research during the period of rapid post-1945 expansion of science (including the industrial organisation of research) and of the large-scale government funding of science (Bush, 1945; de Solla Price, 1986; Garfield, 1977: 313–318; Leydesdorff, 2005). Science research metrics have a generally good track record and the scientometric research of organisations such as the Institute for Scientific Information (ISI, now Thomson Scientific) is heavily used at many levels within science, from international, through national to local, where it exerts a significant influence on the conduct of science policy including funding.

In a nutshell, science research metrics matter because scientific research matters to many powerful individuals and groups outside of science. By contrast, there are few compelling reasons for wishing to measure non-scientific research performance using metrics. To be blunt, non-scientific research is believed (by those outside it) to lack the critical national importance of science. Scientific research involves vast funding and large groupings of organised personnel; but this does not apply to non-scientific research. Scientific research is considered to influence national prosperity, security and standard of living; but the same is not widely believed to be the case for non-scientific areas of research.

Creativity may find its way relatively easily in non-scientific areas of research. However, we believe that these areas may flourish and nurture creativity primarily in the context of science and technology driven prosperity, and may contribute indirectly to advances in scientific research (by opening or expanding domains of life, for instance, the role of music in the expansion of the entertainment industry and related technological research).

Furthermore, research performance metrics in the arts and humanities are generally felt by insiders to be much less valid than in scientific fields. In sum, it can be argued that non-scientific disciplines neither need top-down research metrics, nor do academic specialists in non-scientific disciplines accord such metrics the respect they are given in the natural sciences.

The primary focus of research metrics is implicitly to measure and evaluate science, therefore we suggest that research assessment should be optimised for this function. A research metric should have qualities such as simplicity, transparency, objectivity, replicability, precision and sensitivity. In particular, research evaluation should be done using metrics that allow for both longitudinal (over time) and international (across space) comparisons. When a metric has these qualities it enables scientometrics to develop scientifically, because rival metrics can be checked to see if they perform adequately, and compared to see which metric performs the best overall. The evolution of metrics (the selection of improved metrics over earlier less perfect ones) guarantees that in the long-term valid and robust metrics can be used to measure most areas of science.

Superiority of research-output metrics

Research metrics can be based on input and output variables. Input variables are the resources that generate research, and these include research workers with varying levels of training and skill, institutional infrastructure and research funding. Output metrics include publications, estimates of the quality of these publications, such as the impact factor of journals in which publications appear, and citations generated by these publications (which are a measure of the actual impact of publications).

One suggestion from the Higher Education Funding Council for England (HEFCE) has been that future RAE allocation should be based on input metrics such as external research income (including research council funding and other sources of funding from charities and foundations) (HEFCE, 2007). For instance, the proportion of RAE money a university received might be the same as the proportion of external funding it had won in competition with other universities. Other possible input metrics suggested have included the number of research staff or PhD students.

However, we argue that research metrics for an RAE should primarily reflect research output, and not inputs. While it is true that research inputs and outputs for the UK 1997–2003 show a close statistical correlation at any given point in time (for instance, Andras and Charlton, 2006, unpublished analysis), the use of inputs such as research council income as an RAE funding metric provides an incentive for consuming resources rather than generating research, and over time this would probably lead to greater inefficiency.

A further serious problem with basing an RAE on research grant income is that funding agencies make their decisions on the basis of expert peer review, which is (as described above) a relatively non-transparent and unaccountable method of evaluation. Therefore, the use of input measures would tend to prevent the RAE results being used in longitudinal and international studies of scientific research performance: put briefly, a metric based on income is not well-suited to scientific validation by scientometric methods.

Correlation of citations with other rankings

We favour an RAE based on research-output metrics, specifically citations. The validity of citations as a measure of a university's research performance can be tested empirically by comparing it with independent measures of university research performance.

While rankings based on publication numbers and citations are closely correlated (see below), we would advocate the use of citations in preference to publications. An RAE based on simple publication counts would create a perverse incentive, being prone to manipulation by 'salami-slicing' research output into minimum publishable units (MPUs) to achieve maximum numbers. By contrast, citations are harder to manipulate. Salami-slicing would be discouraged by a citation-based metric because MPUs are of the lowest-possible quality while still being publishable: such publications would tend to attract very few citations.

A further possible manipulation of citations might be individual or small group self-citations. These could potentially be screened out using computerised methods. However, since self-citation is a potentially important aspect of the scientific process (Hull, 1988), the best way to decide whether to make adjustments or corrections to the method of citation counting would be comparative: the performance of a total metric (including self-cites) should be compared (longitudinally and internationally) with a metric screening out self-citations to see which metric yields the best results overall.

This experimental attitude is indeed the general approach we advocate for resolving most methodological debates concerning top-down metrics: the best answer to questions of validity is to run a comparison of metrics over time, and see which one seems most useful overall (see, for example, the metrics developed by the Centre for Science and Technology Studies at Leiden University).[1] Where differences between metrics are small, the simpler metric should be favoured as being more transparent and easier to replicate.

Methods

We recently analysed 30 years of publications and citations from 47 UK universities that were universities before 1975 and did not change their names between 1975 and 2004. The database was the ISI Web of Knowledge (WoK), which pools three specialised indices: Science Citation Index (SCI), Social Sciences Citation Index (SSCI), and Arts and Humanities Citation Index (A&HCI).

Excluded were Cardiff University (because of name changes) and University of Manchester Institute of Science and Technology (UMIST) (which has now joined University of Manchester). Many existing London colleges (Imperial, UCL, Kings, Royal Holloway) have during this time merged with smaller colleges, medical schools and other research institutes. We did not include separate medical schools (such as St George's Medical School) or research institutes (such as Cancer Research Institute).

All universities had unique search strings. For each considered university, we counted the number of publications (of all types) for each year between 1975 and 2004 inclusive, for the combined three domains of WoK. We also counted, for each university and each year, the number of citations that the corresponding publications received between the time of their publication and the time of data collection (February 2006). When a publication was written by multiple authors from more than one included university, the publication was counted for each university to which at least one of its authors belonged.

The rankings of both publications and citations were very similar (Table 1). Furthermore, rankings for publications and citations were also predictive of ranking in the authoritative Shanghai Jiao Tong University (SJTU) Academic Ranking of World Universities published in 2005 with citations providing a slightly closer correlation than publication numbers (Figures 1 and 2) (Institute of Higher Education, Shanghai Jiao Tong University, 2005).

The SJTU ranking sets out to measure excellence in (implicitly) scientific research and it depends on a weighted scale including: Nobel Prize/Fields Medal (10% for alumni and 20% for staff); number of ISI highly-cited researchers (20%); number of articles published in *Nature* and *Science* (20%); number of articles in Web of Science (20%); and a 10% adjustment to control (partly) for the size of institution. Since the SJTU score derives from a different set of data from citations (with the partial exception that citation counts are used to choose ISI highly-cited researchers), this provides an independent validation of citation counting as a measure of research excellence.

Correlations between citations and the SJTU ranking are especially close at the top of the rankings, where there are also large differences between the universities in terms of metrics: correlations are less impressive when the distribution curve is flatter and differences among universities are smaller. Another factor is that the SJTU ranking exhibits increasing levels of statistical 'noise' in lower ranks probably as a result of small-number effects. Only a few elite universities have numerous Nobel/Fields/highly-cited researchers and when there are few individuals in these categories just one more or less can make a significant difference to the SJTU rank.

Consistent with this UK data is that Harvard University (SJTU number one in the world) is also the most productive and most cited university we have measured, generating both more publications and more citations per year than Cambridge and Oxford combined (Charlton and Andras, 2006a; 2006b). In general, this analysis seems to indicate that citation counts are correlated with an independent measure of the scientific research quality of UK universities.

Table 1. Rank based on number of WoK citations, number of publications for five years 2000–2004 and SJTU UK rank

Rank by citations	Rank by no of publications	SJTU UK rank	Institution
1	1	1	Cambridge
2	2	2	Oxford
3	4	4	UCL
4	3	3	Imperial College, London
5	6	5	Edinburgh
6	5	6	Manchester
7	7	7	Bristol
8	8	11	Birmingham
9	9	12–15	Glasgow
10	10	9	Kings College, London
11	12	8	Sheffield
12	11	12–15	Leeds
13	13	10	Nottingham
14	15	16–19	Southampton
15	14	12–15	Liverpool
16	24	20–20	Dundee
17	17	16–19	Leicester
18	16	20–30	Newcastle
19	20	20–30	Durham
20	18	>30	Aberdeen

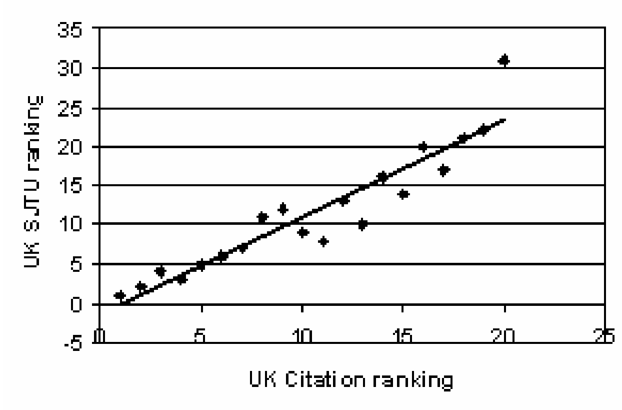

Figure 1. SJTU 2005 ranking for UK universities plotted against ranking for total WoK citations 2000–2004
Note: Spearman rank correlation: R=0.967, p<0.001

Proposal for a citation-based RAE

In the light of the above analysis, we advocate replacing discipline-based expert review with an institution-based research-output metric: total citations accumulated from all publications associated with a specific university during the survey period.

Citations could be surveyed by at least two independent auditors to be chosen by competitive tendering, and using databases such as the ISI Web of Knowledge[2] and Elsevier Scopus.[3] Indeed, Thomson Scientific (ISI) and Elsevier might be suitable organisations for commissioning to perform citation analyses for the RAE, since they have access to raw data, are in competition, and therefore each seems likely to be motivated to do a good job. This competitive system of research assessment fits with our advocacy of a scientometric approach. The

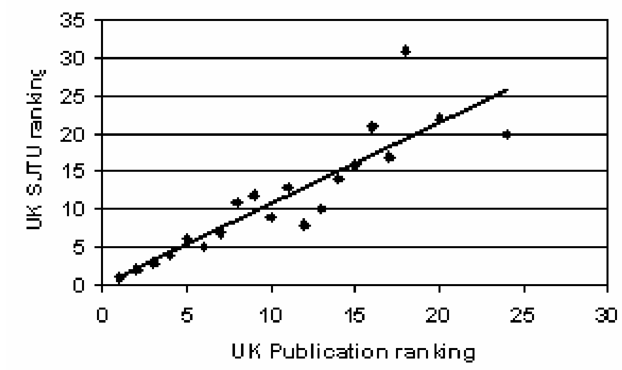

Figure 2. SJTU 2005 ranking for UK universities plotted against ranking for total WoK publications 2000–2004
Note: Spearman rank correlation: R=0.949, p<0.001

best methods for research metrical analysis will tend to emerge with time and over repeated cycles of evaluation in terms of attributes such as superior applicability, precision, predictive power and simplicity.

Simple output metrics are transparent, clear and cheap, and can be measured using independent external expertise, without involvement of those being measured. The objectivity and transparency of a simple citation metric would enable universities to manage strategically, since they can calculate for themselves in advance the approximate level of future RAE funding derived from measuring their own citation growth compared with other relevant institutions.

Because of its cheapness and simplicity, this kind of top-down RAE could be performed every year. However, it is an advantage to use several years of accumulated citations to obtain a valid and precise measurement. We therefore suggest using a rolling seven-year retrospective sample of accumulated citations; seven years is chosen on the basis that it is a reasonable approximation to the timescale of scientific activity (five years is probably too short while ten is probably too long) (Charlton, 2006). The use of a rolling retrospective sample would also smooth out year-by-year changes and avoid sudden large reductions in annual funding, which could prove needlessly disruptive to institutions.

The raw citation counts would be used to create a rank ordering of universities that may be fitted to a funding formula curve. The shape of the funding curve would be determined by strategic research priorities.

What would be the likely incentives created by this total citation metric? Probably, since total citation counts are heavily weighted in favour of natural sciences, the main incentive would be for universities to compete in attracting the most-cited scientific research teams in the leading branches of the natural sciences and quantitative social sciences. This would raise the cost of the most-cited research teams, probably improving the pay, support and conditions of group members, and further increasing competition to succeed in highly-cited fields. We believe such changes would, on the whole, be beneficial to scientific research.

The objectivity and transparency of a simple citation metric would enable universities to manage strategically, since they can calculate for themselves in advance the approximate level of future RAE funding by measuring their own citation growth compared with other relevant institutions

Elite 'revolutionary-science' universities

While in the UK the SJTU and citation rankings are very similar, the top US universities display some significant dissociations between the SJTU rankings and rankings by total WoK citations (that is, including SCI, SSCI and A&HCI) (Table 2).

Some universities are significantly higher in the SJTU and are placed significantly below the line on Figure 3, For example, Berkeley is third in SJTU and tenth in citations; Princeton is seventh and Chicago eighth in SJTU and they are not even in the top 20 for citations. Conversely, other universities are above the line in Figure 3, such as Johns Hopkins which is second for citations and only 17th in SJTU, and the University of Washington, Seattle is fourth in citations and 15th in SJTU.

This pattern is probably influenced by subject mix, since, among the top ten SJTU universities, Berkeley and Princeton do not have medical schools and Chicago only a small one. Conversely, Johns Hopkins and University of Washington, Seattle are major medical research centres. There is also a trend for private universities to rank relatively higher in the SJTU (nine out of the top ten) than in the citations ranking (five out of the top ten).

Since the US universities include the best and most complexly differentiated in the world, this pattern may represent a specialisation into Kuhnian categories of "revolutionary" and "normal" science (especially biomedical science) among the most

Table 2. Top 20 US universities ranked according to SJTU 2005 and by number of WoK citations 2000–2004

SJTU 2005		Citations 2000–2004	
Rank	Institution	Rank	Institution
1	Harvard	1	Harvard
2	Stanford	2	Johns Hopkins
3	U California Berkeley	3	Stanford
4	MIT	4	U of Washington, Seattle
5	CalTech	5	UCLA
6	Columbia	6	U Michigan, Ann Arbor
7	Princeton	7	MIT
8	Chicago	8	U Pennsylvania
9	Yale	9	U California, San Diego
10	Cornell	10	U California, Berkeley
11	U California, San Diego	11	Yale
12	UCLA	12	Columbia
13	U of Pennsylvania	13	U of Minnesota
14	U of Wisconsin, Madison	14	U of Pittsburgh
15	U of Washington, Seattle	15	Duke
16	U California, San Francisco	16	Cornell
17	Johns Hopkins	17	Washington U, St Louis
18	U of Michigan, Ann Arbor	18	U of Wisconsin, Madison
19	U of Illinois, Urbana Champaign	19	U of Colorado
20	Washington U, St Louis	20	U of North Carolina, Chapel Hill

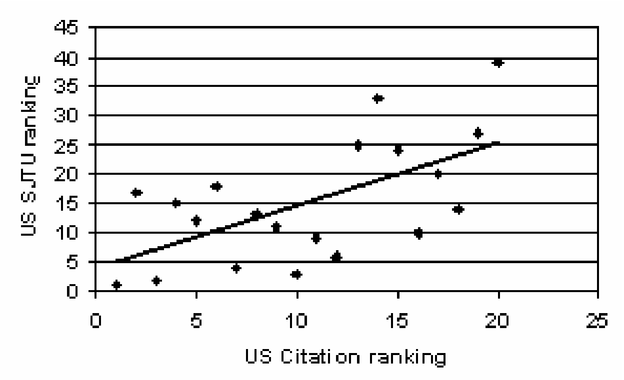

Figure 3. SJTU 2005 ranking of US universities plotted against ranking for total WoK citations 2000–2004
Note: Spearman rank correlation: R=0.564, p=0.01.

productive US universities (Kuhn, 1970). The revolutionary-science top universities (such as Berkeley, Princeton and Chicago) are most likely to gain Nobel Prizes/Fields medals, publish in *Nature* and *Science*, and contain disproportionately large numbers of very highly-cited researchers.

The more 'normal science' top universities (for instance, Johns Hopkins; Washington, Seattle; UCLA; and Michigan Ann Arbor) may be focused more on large-scale biomedical research (this implies their higher metrics-based ranking than in a ranking that takes into account the production of perhaps quantitatively less prolific, but 'revolutionary' quality science; this ranking difference is likely to cause the reduced level of Spearman rank correlation shown in Figure 3).

In its Kuhnian sense, 'normal' science entails activities such as checking and incrementally extending existing knowledge using established techniques and improving existing research methods. However, it should be noted that, by comparison with countries other than the USA, even these more normal science top-cited universities are high achievers in the indices of revolutionary science such as Nobel Prizes, ISI highly-cited academics and *Nature* and *Science* publications.

Our conclusion is that citation metrics probably do not always distinguish the highest levels of excellence in revolutionary science from less innovative but highly productive normal science. Different metrics would be required for this purpose: we suggest the use of Nobel Prizes.

Nobel Prizes as international benchmarks

Some of the components of the SJTU metrics are undesirable as measures of revolutionary science. *Nature* and *Science* are weekly-published commercial journals that contain mostly 'scientifically fashionable' papers expected to gather large numbers of citations. They have an increasing role in generating media discussion of scientific issues, which probably interferes with their status as journals of record. The inclusion of numbers of publications and of highly cited scientists again fail to pick out revolutionary science from the most successful examples of normal science.

We suggest that science Nobel Prizes are the most promising basis for a metric of high-quality revolutionary science, since they are generally regarded as having very high validity for honouring the highest level of science in their fields (in most cases the winners are awarded the prize for paradigmatic work that they did in the longer-term past, usually decades ago, and that has since shown its paradigmatic effect through the work of many followers). This would especially be the case if the Prizes could be reformed to generate more laureates per year and to provide an official allocation of credit to institutions (Charlton, 2007a). (We are aware that other non-science factors, such as political aspects, may also play a role in awarding Nobel prizes, for instance, in economics. However, we believe that these factors have a relatively small influence in most cases, and, in particular, if the number of Prizes were to increase, the role of such factors would decrease further.)

At present, there are four science Nobel Prizes — physics, chemistry, medicine/physiology and economics. Each is awarded to between one and three laureates per year generating between four and 12 laureates annually. Individuals are honoured and there is currently no official allocation of credit to the scientific institutions that have nurtured and supported the prize-winning work. The small numbers of laureates mean that only a few elite universities are able to accumulate sufficient Nobel Prizes over a short enough recent timescale to provide a measure of research quality that is both precise and useful for future policy. Nonetheless, some preliminary institutional analysis is possible, which we present here.

The Nobel Foundation lists the affiliations of laureates at the time the prize is awarded over the past century since the Nobel Prizes began in 1901[4] (many of these citations are several decades old and of doubtful contemporary relevance). A few elite universities emerge with ten or more laureates: Harvard 31; Cambridge 18; CalTech 17; MIT 17; Stanford 17; Columbia 16; Berkeley 15; Chicago 15; Princeton 10, Oxford 10 (also the Rockefeller Institute and University, New York (for graduate students only) has 14 laureates).

For comparison we can look at the evaluations generated by the authors of Wikipedia.[5] They have created a tabulation that simply counts the number of laureates associated with each university since 1901. The calculation uses four categories of association: graduate; attendee or researcher; faculty member before or at the time of the award (this category includes the Nobel Foundation listing above); and faculty member after the award; thus multiple institutions may receive credit for a single laureate. The results are broadly consistent with the more restricted official Nobel figures based on time of award, with a top rank of institutional associations as follows: Cambridge 83; Columbia 81; Chicago 79; Harvard 76; MIT 63; Berkeley 61; Stanford 50;

Oxford 47. The high degree of convergence between these two, admittedly approximate, metrics suggests that the use of Nobel Prizes to measure 'revolutionary science' has considerable promise.

As an example of using Nobel Prizes in a national RAE, we have selected only those UK universities that have had affiliated faculty who received two or more Nobel Prizes in the past 60 years (1947–2006). This picks out just six potentially elite revolutionary-science universities: Cambridge 6; Oxford 4; Imperial College, London 4; University College, London 3; King's College, London 2; and Sussex 2. The metric would be further improved if credit were primarily awarded to institutions that had supported the specific research leading to the Nobel award. So, Nobel Prizes could potentially identify the UK elite of revolutionary-science universities, whose status might justify a separate and extra funding stream to that deriving from the total citation count.

This metric could be made official and more valuable by reforms to the Nobel Prize system (Charlton, 2007a). We suggest more than doubling the number of Nobel science laureates from a maximum of 12 to a minimum of 24 per year. If there were more laureates being created, this would enable a wider range of outstanding scientific work to be recognised and a more precise estimate of the relative ability of universities and other research institutions to generate the highest quality of research.

If the maximum number of three laureates per category were awarded as a matter of course, this would increase the annual number to 12; this number could be further increased by new categories of Prize (such as mathematics and computing science) and by increasing the number of laureates in the biological category of medicine/physiology, which is now the largest and most prestigious branch of world science.

Furthermore, when awarding the Prize the Nobel committee could officially allocate differential credit for this work among research institutions, enabling a more valid metric to be developed (Charlton, 2007a). By such means the Nobel Prize might widen its role from honouring outstanding individual achievement to include a role in measuring and supporting outstanding scientific institutions.[6]

Conclusion

We suggest that the proper purpose of a government research assessment is to generate a rank ordering of universities in terms of their scientific research output. We also believe that the evaluation process should be maximally transparent and accountable.

These constraints imply that the RAE process should be based on objective measures that can be checked by third parties. To minimise distortion or corruption and avoid interference with the work of universities, it is preferable that the RAE does not require the co-operation of researchers. It is also desirable that the RAE is as cheap and quick as is compatible with validity. This suggests that an RAE should be based on top-down quantitative metrics, and we suggest total citations is the best candidate.

We suggest that an institution-based RAE, using total citations calculated on a rolling seven-year retrospective sample, is a metric that fulfils these major criteria. The validity of this scientometric method could continually be monitored on the basis of longitudinal and international comparisons of research performance, including other metrics.

The identification of elite revolutionary-science institutions could potentially be derived from a metric based on the distribution of science Nobel Prizes. A Nobel metric could be used to ensure that universities with a record of supporting a significant volume of the highest quality of scientific research would get specific financial support.

Notes

1. The Leiden ranking, <http://www.cwts.nl/cwts/LeidenRankingWebSite.html>, last accessed 14 September 2007.
2. Thomson Scientific, *ISI Web of Knowledge*, <http://portal.isiknowledge.com>, last accessed 14 September 2007.
3. Elsevier, *Scopus* <www.scopus.com/scopus>, last accessed 14 September 2007.
4. Nobelprize.org, *Nobel Laureates and Universities*, <http://nobelprize.org/nobel_prizes/lists/universities.html>, last accessed 14 September 2007.
5. Wikipedia, *Nobel Prize Laureates by University Affiliation*, <http://en.wikipedia.org/wiki/Nobel_Prize_laureates_by_university_affiliation>, last accessed 14 September 2007.
6. Note added in proof. The use of Nobel prizes, and other similarly prestigious awards, as a scientometric measure of revolutionary science has been extended in further publications (Charlton, 2007b; 2007c).

References

Bishop, Matthew 2004. *Essential Economics*. London: Profile Books.

Bush, Vannevar 1945. As we may think. *Atlantic Monthly*, **176**(1), 101–108.

Butler, Linda 2007. Assessing university research: a plea for a balanced approach. *Science and Public Policy*, **34**(8), October, 565–574.

Charlton, Bruce G 2006. Scientific life should be measured in seven year units. *Medical Hypotheses*, **66**, 1051–1052.

Charlton, Bruce G 2007a. Why there should be more science Nobel Prizes and Laureates — and why proportionate credit should be awarded to institutions. *Medical Hypotheses*, **68**, 471–473.

Charlton, Bruce G 2007b. Scientometric identification of elite 'revolutionary science' research institutions by analysis of trends in Nobel prizes 1947–2006. *Medical Hypotheses*, **68**, 931–934.

Charlton, Bruce 2007c. Which are the best nations and institutions for revolutionary science 1987–2006? Analysis using a combined metric of Nobel prizes, Fields medals, Lasker awards and Turing awards (NFLT metric). *Medical Hypotheses*, **68**, 1191–1194.

Charlton, Bruce and Peter Andras 2003. *The Modernization Imperative*. Exeter UK: Imprint Academic.

Charlton, Bruce and Peter Andras 2006a. Oxbridge versus the 'Ivy League': 30 year citation trends. *Oxford Magazine*, **255**, 16–17.

Charlton, Bruce and Peter Andras 2006b. Best in the arts, catching-up in science — what is the best future for Oxford? *Oxford Magazine*, **256**, 25–26.

de Solla Price, D J 1986. *Little Science, Big Science: and Beyond*. New York: Columbia University Press.

Garfield, Eugene 1977. *Essays of an Information Scientist*. Philadelphia PA: ISI Press).

HEFCE, Higher Education Funding Council for England 2007. *Reform of Higher Education Research Assessment and Funding*. Available at <http://www.hefce.ac.uk/research/assessment/reform/>, last accessed 14 September 2007.

Higher Education Funding Councils 2007. *rae2008 — Research Assessment Exercise*. Available at <www.rae.ac.uk/>, last accessed 14 September 2007.

Hull, David L 1988. *Science as a Process*. Chicago: Chicago University Press.

Institute of Higher Education, Shanghai Jiao Tong University 2005. *Academic Ranking of World Universities 2005: Top 500 World Universities*. Available at <http://ed.sjtu.edu.cn/rank/2005/ARWU2005TOP500list.htm>, last accessed 14 September 2007.

Kuhn, Thomas S 1970. *The Structure of Scientific Revolutions*. Chicago: Chicago University Press.

Leydesdorff, Loet 2005. The evaluation of research and the evolution of science indicators. *Current Science,* **89**, 1510–1517.

Luhmann, Niklas 1995. *Social Systems*. Cambridge MA: Harvard University Press.

Moed, Henk 2007. The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy*, **34**(8), October, 575–583.

Mankiw, Gregory N 2003. *Principles of Economics*. Mason US: South Western College Publishers.

REPP, Research Evaluation and Policy Project 2005. *Quantitative Indicators for Research Assessment*. REPP Discussion Paper 05/1. Canberra: The Australian National University.

van Raan, Anthony F J 2005. Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, **62**, 133–143.

Ziman, John 2000. *Real Science*. Cambridge UK: Cambridge University Press.