Jointly published by Akadémiai Kiadó, Budapest
and Springer, Dordrecht

Scientometrics, Vol. 72, No. 2 (2007) 325–344
DOI: 10.1007/s11192-007-1696-x

Profiling citation impact: A new methodology

JONATHAN ADAMS, KAREN GURNEY, STUART MARSHALL

Evidence Ltd, Leeds (UK)

A methodology for creating bibliometric impact profiles is described. The advantages of such profiles as a management tool to supplement the reporting power of traditional average impact metrics are discussed. The impact profile for the UK as a whole reveals the extent to which the median and modal UK impact values differ from and are significantly below average impact. Only one-third of UK output for 1995–2004 is above world average impact although the UK's average world-normalised impact is 1.24.

Time-categorised impact profiles are used to test hypotheses about changing impact and confirm that the increase in average UK impact is due to real improvement rather than a reduction in low impact outputs.

The impact profile methodology has been applied across disciplines as well as years and is shown to work well in all subject categories. It reveals substantial variations in performance between disciplines. The value of calculating the profile median and mode as well as the average impact are demonstrated. Finally, the methodology is applied to a specific data-set to compare the impact profile of the elite Laboratory of Molecular Biology (Cambridge) with the relevant UK average. This demonstrates an application of the methodology by identifying where the institute's exceptional performance is located.

The value of impact profiles lies in their role as an interpretive aid for non-specialists, not as a technical transformation of the data for scientometricians.

Received December 12, 2006

Address for correspondence:
JONATHAN ADAMS
Evidence Ltd, 103 Clarendon Road, Leeds LS2 9DF, UK
E-mail: jonathan.adams@evidence.co.uk

0138–9130/US $ 20.00
Copyright © 2007 Akadémiai Kiadó, Budapest
All rights reserved

Introduction

Scientometric analyses of research performance conventionally produce a single final metric, such as the average performance of a research group, and then use some broader reference, such as a national baseline, as a benchmark. Such metrics have significant retrospective reporting value but limited prospective management utility because they do not illustrate and so cannot explain the distribution of performance. This paper describes the application of a simple but powerful charting methodology for looking at research quality by profiling and comparing distributions. This could extend the potential accessibility of bibliometrics for research management as well as enhancing their use in evaluation.

A typical research performance indicator might be a measure of group output relative to world outputs or it might be outputs per unit input, or an average of a data sample such as income per year or citations per paper (MOED et al., 2004). UK examples are captured in the Office of Science & Innovation's (OSI) annual Public Service Agreement (PSA) Target Indicators report (EVIDENCE, 2005). An approach widely used in international comparisons, such as the EC's Science & Technology Indicators (EUROPEAN COMMISSION, 2003), is to capture research performance in terms of average citation impact (MOED, 2005). Indeed, WEINGART (2005) has noted that whereas bibliometrics were once an irritation they have now become strongly favoured by administrators and politicians. MICHELSON (2006) sees the same pattern across US research funding agencies.

Citations per paper are related to impact – as GARFIELD (1955) long ago noted – and citation averages are a useful tool for simplifying reports. They digest a lot of data and absorb outlier values. They seem easy to understand. But that understanding is often based on naïve assumptions: for example, that the data are normally distributed. If the data are actually skewed then the average (assumed to be central) will differ from the midpoint value (median) and the most common value (mode). An average can therefore mislead interpretation of the true nature of the activity.

It is our experience that many non-scientometric policy makers, aware that citation metrics show that UK research performance is 'better than world average' (ADAMS, 1998), have assumed that the median of UK research and therefore about half or more of UK outputs are also 'better than average'. In fact, as analysts should know, an average says little about the spread of activity on either side. We don't know how much is 'much better', or how much is 'about average', and so we cannot explain why a performance average might have changed.

For example, UK average performance relative to key competitors across all fields improved between 1995 (MAY, 1997) and 2004 (KING, 2004; EVIDENCE, 2005: OSI Indicator 3.08). OSI Indicator 2.03 shows that the UK published about 69,400 papers in

2004, which was a slight drop on recent years (usually in excess of 70,000). Uncited papers (OSI Indicator 3.05) decreased relative to other countries.

The 'politician' is pleased to note that the UK is publishing fewer papers, more of the published papers are being cited, and the average citation impact has improved. The 'management' problem is that conventional bibliometric indicators cannot distinguish between competing hypotheses about the dynamics behind these trends. On the one hand, we could assert that 'the UK is getting better' at research implying that there have been changes in performance at the top end, where publication impact is most likely to have research and economic consequences. On the other hand this could be simply a reduction in low impact outputs that dilute the higher impact material. If UK researchers had stopped publishing marginal contributions that had no recognised value in their field, then that would account for all three changes.

This is important because if the analysis of science and technology indicators rests on the production of averages, then the discipline remains one that is essentially focussed on reporting. This is of great value and supports both policy and evaluation. But it is also limiting. It is not an aid to management. To take Scientometrics over the threshold from reporting to management we will need further evolution in both analysis and presentation so as to allow research managers to gain insights that support action.

New indicators are needed to 'unpack' the averages to which conventional metrics refer. Research managers, as well as analysts, need something more to inform their decision making and to link bibliometrics to peer review in the mode that WEINGART (2005) proposes. They need to see where the actual spread of performance falls – and track its dynamics – so they can make clearer and more targeted decisions about effective interventions. Where to invest? Where to encourage? Where to apply performance reviews?

Michel Zitt (ZITT et al., 2005) has shown the value of categorising citation profiles in developing his thesis about the significance of the level of normalisation and its effect on both normalised impact value and its ranking. He shows (op. cit. Figure 2) how the distribution of mean impact is affected by shifting normalisation from discipline to sub-discipline levels.

We have sought to focus on the distribution profile itself. In this paper, we seek to move the development of bibliometrics (and other research indicators) forward by disaggregating existing indicators rather than inventing something completely new. The approach we have chosen is to stick with bibliometric data, because they are widely used and understood across the research base. The change is to look at the distribution of the impact values for individual articles rather than the average value for samples.

Distribution profiles will show the activity at the high and low end of the performance spectrum as well as the height of the peak somewhere near the middle. Such a picture would test our assumptions and might force a rethink of what we believe happens in the research base. Our new approach is to ask: what spread of activity lies beneath the UK average? As a first step, we:

- Develop a profiling methodology and evaluate descriptive statistics that might enable rapid comparison between profiles.
- Profile the distribution and variation of performance across the UK research base as a whole.
- Create Impact Profiles for time- and discipline-based data samples, to establish the applicability of the general outcome to disaggregated data.
- Compare the Impact Profile of an elite institution with the national profile.

Methodology

In this section we first describe the data to be used for the analyses. Using the example of UK Physics we note that raw data are highly skewed, as has been widely described in bibliometrics. We discuss why this makes interpretation complex and reduces utility for policy purposes. We then propose a simple but powerful transformation and categorisation of the data to make a rational charting process more feasible. The transformed Impact Profile follows a curve in which the average is non-central. We note additional metrics, such as mode and median, and discuss how these, while not in themselves solving the problems of scientific evaluation, can provide more information about the distribution of performance within profiled data. These additional metrics are likely to be of value when groups of curves need to be compared or when summary data need to be dropped into a table.

In the next section, we analyse total UK publication data from the 10-year period 1995 to 2004. The UK impact (citations per paper) is benchmarked against world average baselines for year and field. We then break these data down by time and by discipline.

Data source

The data used in these analyses are taken from Thomson Scientific Inc databases. Individual article records are taken from Evidence's UK National Citation Report (1995–2004) derived from ISI Web of Science. Data are grouped for analysis by journal categories which bring cognate research areas together using a journal similarity analysis developed in earlier studies (ADAMS, 1998). World average impact data are sourced from the Thomson Scientific National Science Indicators 2004.

Evidence has performed a complete address reconciliation for the Thomson Scientific databases for the UK. This enables an accurate association of journal articles with institutions, using a combination of the name, address components and post codes. We have processed around 3 million address records, and linked each of these to a named research organisation. These organisations can be grouped into functional categories such as University, Health, Research Council, Industry and so on. They can also be grouped by post-code and region. This provides a comprehensive, detailed and accurate data source for analysing the publication record of individual institutions and making appropriate peer comparisons.

Physics as a data example

There are several data issues to be addressed, and they can be illustrated by looking at a sample curve, using a sub-set of the UK data. We took a single subject category (Physics) and a single year (1995, so citations have had plenty of time to build up). We then looked at the spread of impact (actual citation counts) for individual papers and then normalised this against the appropriate world average for Physics in 1995.

There were over 2000 article records, and we grouped the individual papers into about 100 equally spaced categories according to their ReBased citation Impact (RBI). That is, we took the highest impact measure (which was over 43 times world average for the field and year) and then divided the rest of the data from 0.0 impact (uncited) up to that maximum into 100 categories with equal impact increments (Figure 1).

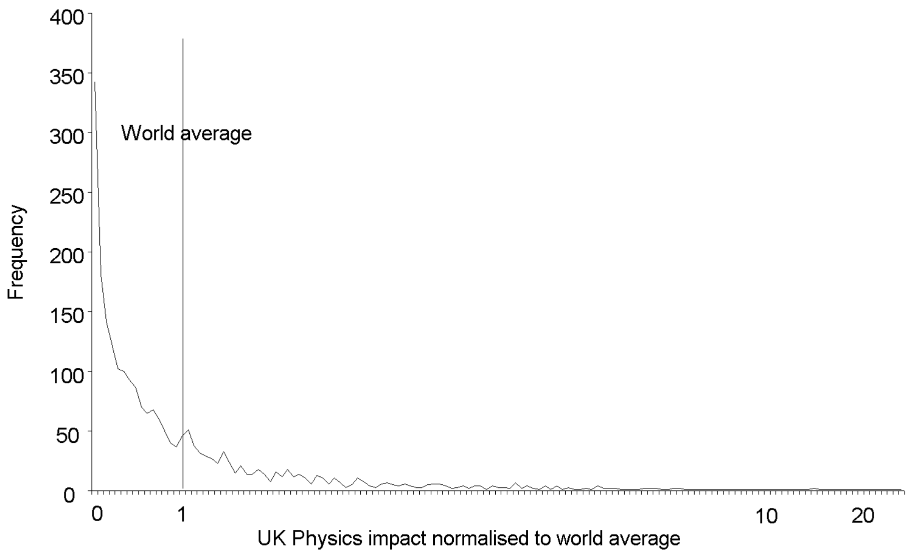

Figure 1. The distribution of rebased citation impact (target group citations per paper/ world citations per paper) for 2323 UK articles published in 1995 in the Thomson journal category for Physics. Rebased impact (RBI) is citations per paper normalised against world average for each journal category and for year of publication

Such a picture of impact indices will be unsurprising to bibliometricians and it was aptly described more than 40 years ago by DE SOLLA PRICE (1965). It is challenging to interpret, however, particularly for any non-specialist. On the one hand, there is a long trail of excellent papers going up to very high relative citation values, over forty times world average impact. On the other hand there is an apparent concentration below world average. The data are highly skewed. How can policy makers and research managers interpret this in terms of possible decisions and actions?

The problem is that the data are not normally distributed. In fact, as widely noted, citations per paper generally follow a Negative Binomial Distribution (NBD) because of both heterogeneity (variance is related to the size of the mean) and contagion (cited papers are more likely to be cited again) in the underlying data (elegantly reviewed by LEYDESDORFF & BENSMAN, 2006). Furthermore, in charts like Figure 1, each positive data point is a ratio of UK/world because citation impact is usually normalised for field and year by comparison with the appropriate world average. Impact (citations per paper) is already a ratio and is therefore not normally distributed. Dividing a ratio by a ratio takes the index further away from the original data, where the citation distribution was already skewed. The statistical distribution of ratios and percentages tends to cluster as the values approach 0 (i.e. 0%). Thus, although impact metrics are valuable indicators, they do not have an accessibly meaningful relationship to a statistically definable distribution.

Second, you cannot get fewer than zero citations so there is a cut-off at 0.0 and there are no negative values. However bad a paper, it cannot get 'less than nothing' so no scale of disapprobation exists. Furthermore, we cannot distinguish between papers uncited because they are uninteresting and papers uncited because they are, for example, important records of negative outcomes. It is therefore not feasible to create a scalar value for uncited records.

Alternative methodological approaches

Anatoly Zhigljavsky has drawn our attention to the possibility of creating formal, statistical descriptions of the data in Figure 1 and the similar outcomes that we would find in any bibliometric analysis (SAVANI & ZHIGLJAVSKY, 2006, in press a, in press b; EHRENBERG, 1988). This follows historical analyses of NBD models described by LEYDESDORFF & BENSMAN (op. cit. and see FELLER, 1943).

The advantage of such an approach is that it could provide statistically sound parameters that would accurately describe the citation data. These could be used both for reporting and for comparative tests and would therefore be supportive for expert and analytical purposes. But, as LEYDESDORFF & BENSMAN (op. cit.) note the precise mathematical model requires a crisp set which cannot be expected in research journal data. Furthermore, a continuing disadvantage would be that the graphical description of the data remains relatively difficult to interpret. This would therefore be a gain technically but not in terms of utility for policy and management purposes where simplicity of interpretation is critical.

Data categorisation

How can we provide straightforward profiles, accessible to general interpretation, without distorting the data? We have done two things to overcome the problem. First, it is necessary to treat the uncited papers as a wholly separate class. Second, it is necessary to categorise the impact values of the cited papers not by their absolute value but by their value relative to the world average.

Table 1. Schematic structure of data categorisation for profiling the distribution of research impact
Cited < or > world average refers to the world average appropriate to the field and year of publication of each article analysed

<table>
<tr><td colspan="7">All papers</td></tr>
<tr><td>Uncited papers</td><td colspan="6">Cited papers</td></tr>
<tr><td></td><td colspan="3">Cited < world average</td><td colspan="3">Cited > world average</td></tr>
<tr><td></td><td colspan="2">Cited < ½ world</td><td rowspan="2">Cited < world but > ½ world</td><td rowspan="2">Cited > world but < 2 x world</td><td colspan="2">Cited > 2 x world</td></tr>
<tr><td></td><td>Cited < ¼ world</td><td>Cited < ½ world > ¼ world</td><td>Cited > 2 x world < 4 x world</td><td>Cited > 4 x world</td></tr>
<tr><td></td><td colspan="6">And so on</td></tr>
</table>

The methodology is summarised in Table 1. It produces a data structure which is readily implemented, powerful in its outcomes, yet simple to understand. The structure is transparent and easily interpreted. Categorisation means the modified values approximate to a normal distribution. The profile needs only to 'approximate' for presentational purposes, because we are not proposing to carry out any statistical analyses. The advantage of a graph of a normal-like distribution is that it produces the familiar 'bell shaped' curve with which people are likely to feel comfortable and for which they will have a model in mind, so they will be able to pick out marked skews and divergences. Figure 2 shows how well this image works.

The transformation is equivalent to drawing the ReBased Impact on a log (2) scale. It then immediately becomes obvious that one cannot display zero (uncited) as part of the continuous distribution. Indeed, on such a scale 'zero' is not possible.

In practice, we have used eight categories for the cited papers, doubling or halving the impact relative to world average for each interval and using the first and eighth categories to collate all cited records with relative impact below 0.125 (1/8th world average) and above 8.0 times world average. The uncited papers form a ninth group, alongside the categorical scale but not part of the profile.

In the first iteration of the methodology, we took the UK National Citation Report for 1995–2004, which extracted, from the journals indexed by Thomson, all articles with at least one UK-based address. We collated 750,376 individual article records at the level of the detailed (100+) Thomson journal categories. We then rebased the specific citation counts (raw impact) for each record against the appropriate world benchmark for each journal category and for each year of publication. This produced a ReBased Impact (RBI) for each article, with the index scaled appropriately against a relevant world average. We then categorised the RBI values according to the grouping scheme set out in Table 1.

In subsequent iterations, the same methodology was followed but the data-set was structured by time frame, by journal category or by named institution within the UK.

Additional statistics

It is evident that the untransformed data are not normally distributed (Figure 1). Statistical measures based on normality, such as mean and variance, are therefore inappropriate. However, it is worthwhile considering how value can be created for tabulated descriptions of the distributions in the transformed profiles. To do this we reflect on the questions that policy makers are likely to ask when reviewing evaluation outcomes for any sample of research outputs.

What is the commonest group? Mode is the category which has the greatest number of values. It will therefore be a strong indicator of where the centre of the distribution falls without the misleading precision that a specific average would tend to imply.

Where is the midpoint of the distribution? Median is the value below and above which lie half of the ranked data values. This is likely to be the value that many people have in mind when they are told about the 'average', and in a normal distribution the median and the average will be approximately the same. However, in a skewed distribution they will depart significantly. We have used the profiled data to calculate a new reference value of the ReBased Impact (RBI) index. This is the median RBI which is the value in the midpoint of the distribution, below and above which lie half of the actual world rebased impact values for individual papers.

How does the spread of research impact relate to world? Where the mode is below the average then the median will also be below the average. Where the median is below world average, then it indicates how far below world average half of the sample output lies.

What percentage of the research is above world average impact? This is the obverse of the information gained from the median, because that usually lies below world average. Whereas the median provides a performance indicator of the relative status of half of the sample output, this percentage throws light on the relative output above a certain indicator value.

How much of the research is high impact? 'Highly-cited' is a criterion of publication excellence used by the Research Services Group at Thomson Scientific (SMALL, 2004). These papers are typically the top 1% of papers for category and year. In developing the methodology reported here, we found that the RBI value that would demarcate the top 1% of UK papers lay around RBI = 9–11 in five large (10,000 to 50,000 paper) category-based samples across the discipline spectrum. The percentage of papers in our category RBI > 8 is slightly less excellent, usually covering the top 1.5–2% of UK papers, but has robust comparability. We propose to use this as a marker of an equivalent grouping to Thomson's 'highly cited' within the categorical structure we have developed.

We have not provided all of these values for every analysis. We believe that future developments will need this tabulated translation of the profiles in order to provide compact summaries. The relationship between specific types of profile – which may change by discipline and by country and will certainly vary at institutional or programme level – and such values will need to be explored in detail in further analyses.

Results

UK output profile 1995–2004

The categorisation methodology was applied to the total UK output for the ten year period 1995–2004, with the citation count for each article normalised (rebased) against the relevant world benchmark for journal category and year of publication (Figure 2).

The vertical axis scales the percentage of total output that falls in each of the categories in the horizontal axis. An alternative would be to scale the actual count of papers. The graph would look exactly the same, for any single sample, whether we used actual count or share of total. However, if we put two different data-sets together then the column heights would be affected by the number of papers in each sample. Using percentage of total allows us to compare profiles irrespective of sample size.

Figure 2 is much easier to interpret than Figure 1, but how can we relate the profile to the UK average impact? The UK average impact is, at 1.24 for the same period, well above world average yet in Figure 2 the modal (most common) group of cited papers is that where RBI = 0.5–1 (between world average and half world average). In absolute terms, the commonest group is actually the uncited papers, but this is not on the same 'scale' as the other categories.

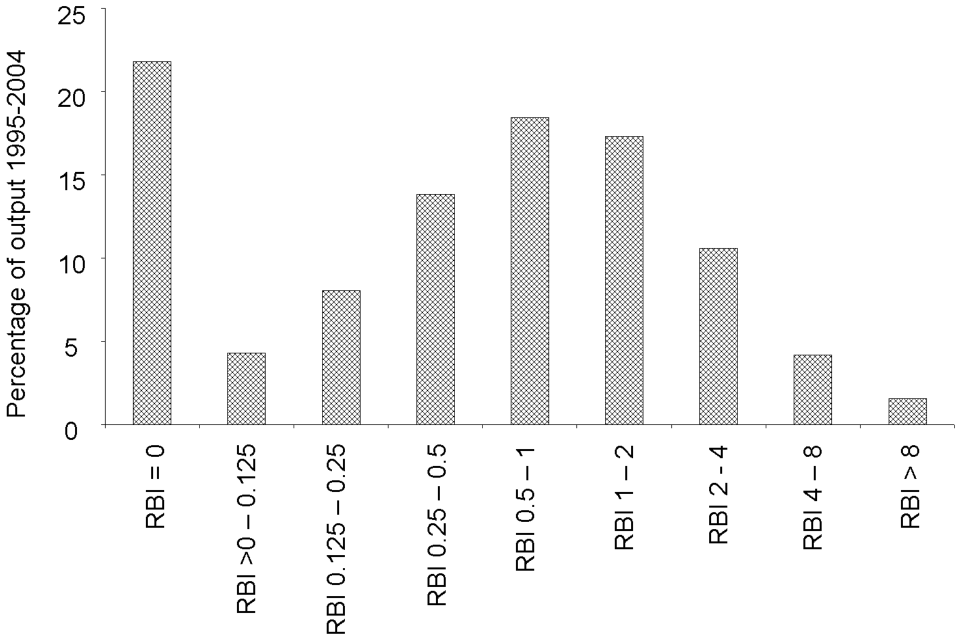

Figure 2. The distribution of relative citation impact for total UK research 1995-2004.
Ten-year UK average impact rebased against world = 1.24.

Does this analysis say that the UK is doing less well than we thought? It does not. Conventional impact analysis is unaffected by this novel presentation. Similar analyses for France, Germany and other key competitor nations would produce pictures of research performance that looked like Figure 2. What the new methodology says is that everyone actually has a different – and probably more concentrated – distribution of excellence than some analysts may have assumed.

It may surprise people that more than half the UK's output is uncited or has a citation count less than the world average. Specifically, about two-thirds of the UK's papers are in these categories. This seems incongruous given the expectations built up by years of looking only at indices of average impact. The UK average can be greater than 1.0 despite the position of the mode and the number of uncited papers because of the value of the papers in the citation groups where RBI > 4. These papers, with many times the world average citation count for their year, pull up the UK overall position.

It is more and more clear that the critical part of the UK's performance is not the bulk activity around the centre but the outstanding high-end performance.

Time: UK output profile by three-year window

Research profiling can be further tested by splitting the data into individual years or by categorising the data by selected time periods shorter than a decade but greater than a year, such as a 3-year research grant cycle or a 5-year moving window. Our conclusion was that a year-by-year approach might have its place in a detailed analytical study, but that grouped years were simpler for comparative work in evaluating the utility and robustness of the basic methodology.

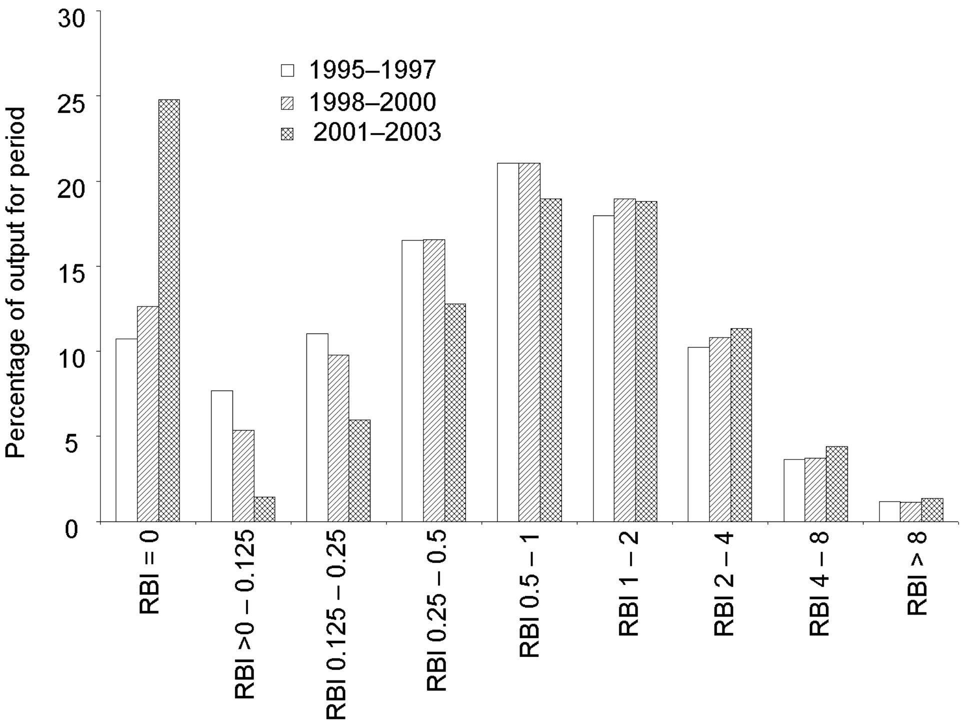

Figure 3. Tracking the distribution of relative citation impact for all UK research using three-year windows for the period 1995–2003

Figure 3 uses three 3-year windows, dropping the most recent year with its high proportion of 'not yet cited' papers. The shift across years is easy to discern. The category RBI = 1.0–2.0, just above world average, changes little as a percentage of output and acts as a 'pivot'. The categories with lower impact all fall in relative frequency over the period while other categories proportionately increase, which is in accord with indicators of a trend of improvement for the UK.

Discipline: UK output profile by journal category

A diversity of 'discipline' charts can be created in the same format as Figure 2. To evaluate the general applicability of the methodology, we chose to produce analyses at the level of Thomson's Standard journal categories (rather than the finer level 'de-luxe' categories). These broadly equate to groups equivalent to 'Schools' in a University.

Table 2 summarises key statistics for these categories, ranked by the proportion of output with normalised or ReBased Impact (RBI) greater than world average.

Table 2. Summary of research impact profiles for the UK at the level of 24 Thomson Scientific journal categories. Categories are ranked on percentage of articles with impact greater than world average for that field in their year of publication

Journal category	Total output 1995–2004	Percentage of output				% > World Average	RBI Average 2000–2004	RBI Median 1995–2004
		Uncited	RBI < 1	RBI 1–8	RBI > 8			
Plant & Animal Science	38582	19.9	39.4	38.0	2.7	40.7	1.51	0.73
Agricultural Sciences	11256	21.8	39.2	36.8	2.2	39.0	1.48	0.75
Pharmacology	13750	15.4	46.2	36.2	2.3	38.5	1.38	0.69
Space Science	13107	15.7	46.2	36.2	1.9	38.1	1.32	0.68
Chemistry	60022	19.0	44.6	34.6	1.8	36.5	1.23	0.63
Ecology/Environment	16884	20.9	42.8	33.9	2.4	36.4	1.40	0.64
Geosciences	22939	21.4	42.2	34.7	1.7	36.4	1.33	0.65
Mathematics	9596	35.7	29.6	31.0	3.7	34.7	1.29	0.52
Neurosci.& Behaviour	29141	13.3	52.2	33.1	1.4	34.5	1.18	0.60
Molec. Biol. & Genetics	23805	10.0	55.7	32.8	1.5	34.3	1.27	0.61
Biology & Biochemistry	53465	13.8	52.1	32.6	1.5	34.1	1.22	0.58
Multidisciplinary	11473	23.9	42.0	31.4	2.7	34.1	1.24	
Physics	61205	23.2	42.6	31.2	3.0	34.2	1.42	0.54
Microbiology	17705	12.3	54.0	32.4	1.3	33.7	1.20	0.62
UK total & average	**750376**	**21.8**	**44.6**	**32.1**	1.5	**33.6**	**1.28**	
Engineering	55236	35.4	31.6	28.6	4.3	32.9	1.08	0.39
Psychology/ Psychiatry	22499	24.7	43.7	29.9	1.8	31.7	1.15	0.49
Materials Science	17623	29.4	39.1	29.5	2.0	31.5	1.17	0.47
Immunology	12604	12.1	56.9	29.9	1.2	31.0	1.08	0.54
Clinical Medicine	179247	20.8	48.3	28.6	2.4	30.9	1.21	0.45
Social Sciences, general	36108	35.6	34.1	28.4	1.8	30.2	1.05	0.37
Economics & Business	15236	36.9	37.2	22.8	3.1	25.9	0.93	0.31
Computer Science	6117	43.6	30.7	21.6	4.1	25.7	1.01	0.22
Education	3852	45.4	29.3	23.0	2.4	25.4	1.02	0.26
Law	1020	50.3	41.1	8.4	0.2	8.6	0.36	0.00

Two impact values are given for each journal category. First, the UK's average recent (last 5 years) RBI is calculated for each category. Second, a median RBI is also calculated to determine the midpoint of the distribution, i.e. that below and above which lie half of the normalised impact values for individual papers. The Table also summarises the proportion of papers in the most highly-cited category where RBI exceeds 8 times world average.

There is considerable variation in the size of the categories, from Clinical Medicine at 179,247 articles to Mathematics at 9,596. The proportion uncited is usually in the range of 10–25% but some fields exceed this. No field has much more than 40% of its papers above world average but most have 5% or more in the group that is more than 4 times world average.

Average RBI – the traditional metric – is correlated with median RBI (Spearman rank correlation, r = 0.90) but the link is dependent on the shape of the profile, so they are not as closely tied as might be supposed. Both median RBI and average RBI are very strongly correlated with the proportion of papers with impact above world average (r = 0.94 and 0.96). It is clear that these analyses do not show that average RBI, the traditional indicator of research performance, is in any sense a misleading reference point. Any past analyses remain entirely valid. However, the additional information reveals the extent to which average RBI tells only part of the story.

Pair-wise comparisons show that all the components affect overall standing. For example, Chemistry has a lower recent impact than Physics (1.23 cf 1.42) but its median impact (0.63) is higher (Physics is 0.54). This is because Physics produces more uncited papers (23% vs. 19%) but has more papers in the most cited category, where RBI > 4 (6.7% vs. 5.4%). By contrast, Engineering has a greater proportion of output above world average (33%) than does Clinical Medicine (31%) but also has more uncited papers (35% vs. 20%). The recent average RBI of its total output is consequently lower (1.08 cf 1.21) although it has the strongest performance for any discipline in the category where RBI > 4 (7.4%).

The graphs in Figures 4–7 profile a spread of categories. The standard bell-shaped profile is common to most but not all of the categories that we have looked at. Some curves are flatter because more papers are uncited (e.g. Mathematics), or asymmetric because there is a relative excess or deficit above world average (e.g. Molecular Biology, excess low RBI; and Geosciences, excess above RBI = 1.0). There is variation in the distribution between those categories where the UK average impact is high and that where it is low. As noted in the examples above, the specific components contributing to the outcome can be teased out via these profiles where previously they were hidden by average indices.

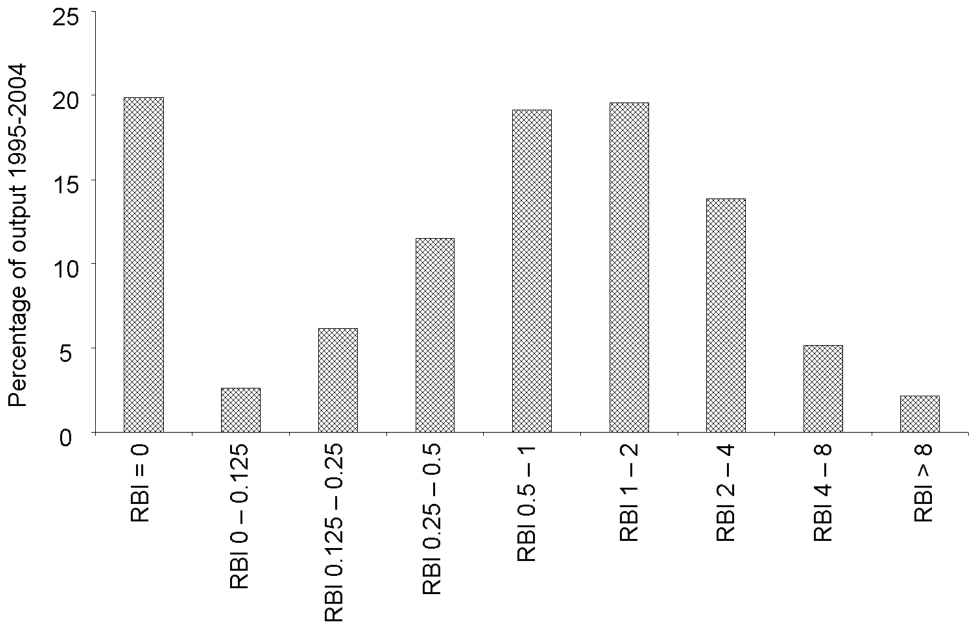

Figure 4. The distribution of citation impact for UK research in Plant & animal science 1995–2004

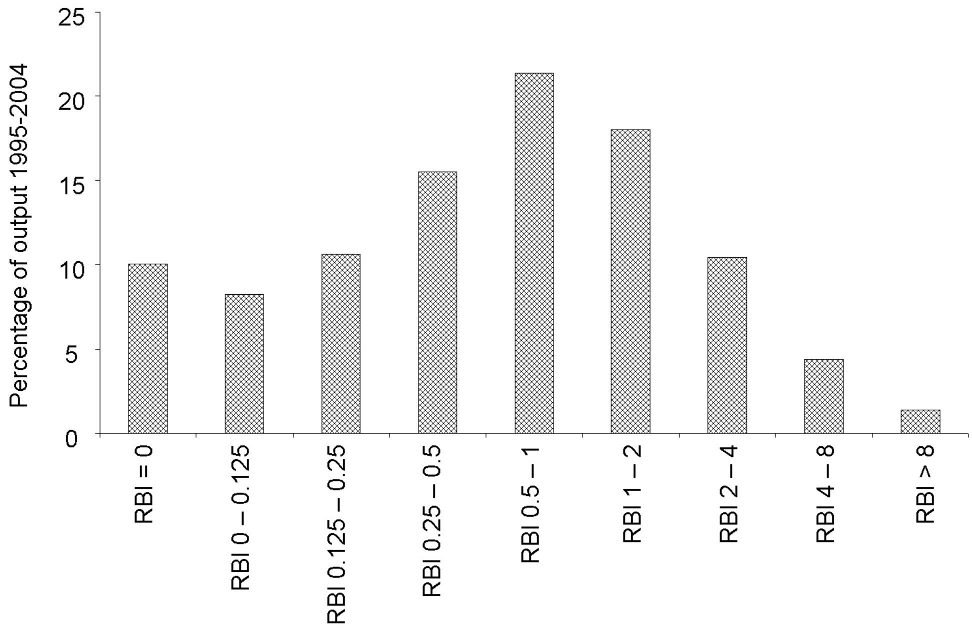

Figure 5. The distribution of citation impact for UK research in Molecular biology & genetics 1995–2004

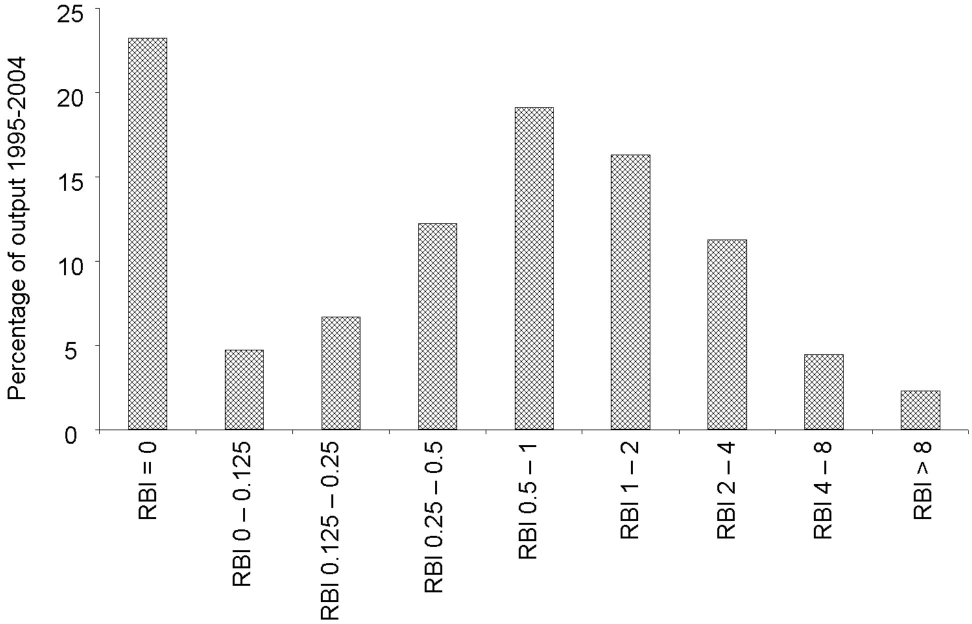

Figure 6. The distribution of citation impact for UK research in Physics 1995–2004

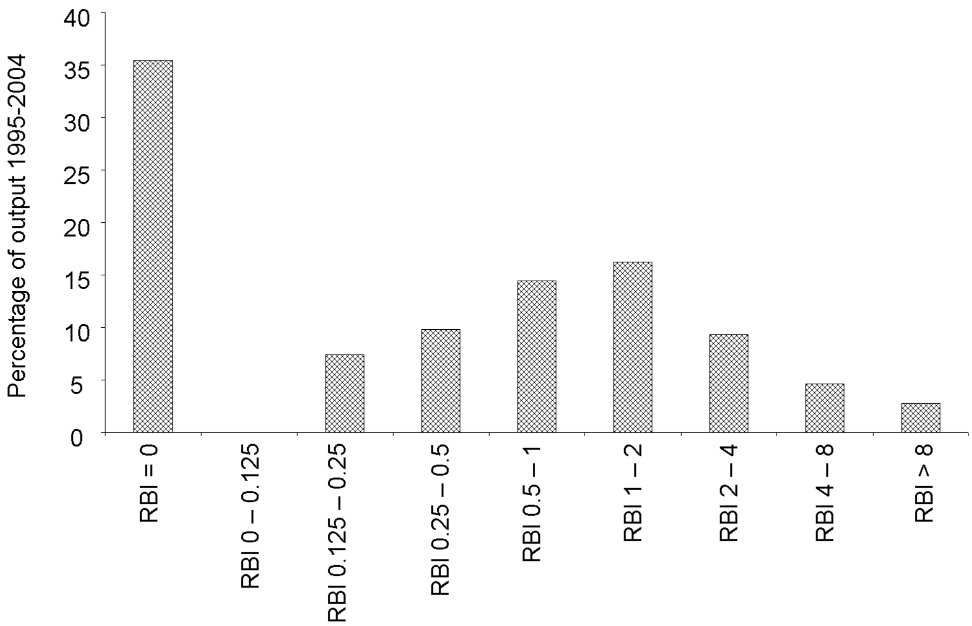

Figure 7. The distribution of citation impact for UK research in Engineering 1995–2004

Institutional profiling

The purpose of the profiling methodology is to support management evaluation of research performance, which will most often be at the level of an institution or a programme. The specific application of the methodology was therefore evaluated by examining the research activity of an elite UK research institution in its core research area and comparing this with the background profile of UK research performance across the same field.

The institute selected for this analysis was the Laboratory of Molecular Biology, Cambridge (the LMB). The institute, which is a dedicated research unit funded by the Medical Research Council, has an international reputation for its research and has won an exceptional number of Nobel prizes. It would therefore be predicted that its profile would be significantly better than the general UK profile in all respects. It should have fewer uncited and low cited papers, a higher impact mode and more papers in the most cited group (more than 8 times world average). This should be tested by examining its relative performance in the core journal category of Biology and Biochemistry research.

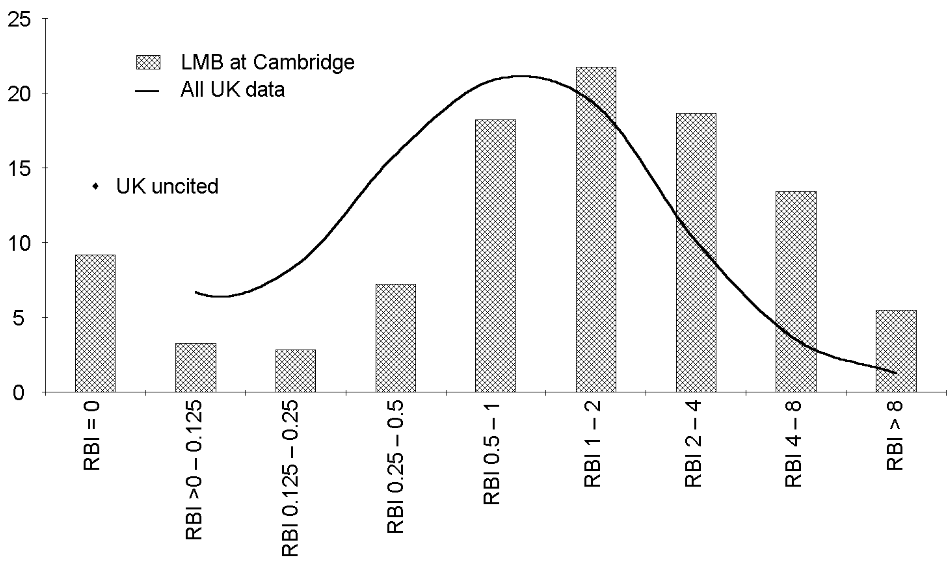

Figure 8. The profiled citation impact (1995–2004) of the UK Laboratory of Molecular Biology compared with the UK as a whole. Data are for the Thomson journal category of Biology and Biochemistry.

The relative excellence of the LMB would, of course, be identifiable from the traditional measure of average citation impact. Its ReBased Impact is about 2.2 compared to an average for the UK of 1.22 (Table 2), but this does not describe how the LMB achieves that high relative impact. The LMB's full performance profile is illustrated by the analysis in Figure 8, which reveals that all the predicted outcomes occur in practice:

- as a percentage of **its** total output, the LMB has many fewer uncited papers than the UK generally;
- the LMB has relatively fewer papers in groups cited less than world average;
- its mode is above world average where the UK's overall mode is below;
- it has relatively more of its output in the categories with impact above world average;
- it has an exceptionally high percentage of its output in the most highly cited category (6%) compared with UK average (2%).

Discussion

The methodology developed and presented in this paper does not challenge any prior work in Scientometrics but it is likely to challenge some of the conclusions that policy analysts may have drawn after looking at bibliometric reports without examining the detail. However, the critical value of the technique is not in shaking up past assumptions. The value lies in a simple profiling approach that can take bibliometrics forward from a reporting mechanism for past performance towards a management tool to support decision making about future investment. It makes the link to per review, as discussed by WEINGART (2005), much easier to achieve because a profile makes the data more transparent and accessible.

The basic methodology creates a simple profile that reveals the spread of performance across a group of outputs. The methodology applies a transformation to the data not for statistical purposes – which LEYDESDORFF & BENSMAN (2006) suggest is nugatory – but for interpretative purposes. For the UK, the profiled distribution over a ten-year period (Figure 2) shows that the majority of papers are below world average impact. The nature of the data that produce this profile as well as a UK average well above world average may not have been widely appreciated, so these impact profiles will almost certainly change policy perceptions about the interpretation of bibliometric indices.

The methodology also has immediate management value. The time progression (Figure 3) reveals that the rise in UK average impact is due to a small but significant increase in the numbers of highly-cited papers relative to the rest of the world.

In relation to the 'management problem' we outlined in the Introduction, this confirms the hypothesis that there has been real improvement in the UK rather than an improvement in effectiveness through a decrease in poorer quality publications.

The category-based profiles work well and demonstrate differences between journal categories (Table 2). That such differences are present will have been apparent from average impact indices, but the profiles show how the composition of the research base varies between disciplines in terms of the numbers of uncited papers and the relative spread below and above world average.

Measures such as the proportion of papers above world average and of median RBI have not previously been used to index research performance. The low values for most of the medians will probably cause some surprise and concern. We must point out that the same pattern would likely emerge for any country analysed in this way. In most fields more than half of the UK's papers have a citation impact below world average, as they do for the UK as a whole. Even in the best performing categories the median only stands at 0.75 world average RBI and no more than 41% of papers have an impact exceeding world average RBI.

Most of the profiles conform to a bell-shape (Figures 4–7) so the disaggregated research areas broadly follow the pattern of the total national curve. This is important in affirming the assumptions made about the general model of performance and the validity of the method. As we reduce the analysis to finer details we can move towards identification of exceptions with confidence that these are meaningful.

The profiling of a single institution against a national average works well (Figure 8). It demonstrates just where the average citation impact of the LMB is delivered, in its output of truly excellent papers as much as in the small number of infrequently-cited and uncited papers. Similar analyses for other institutions can be expected to reveal other features of their performance and thereby guide research managers as to how performance can be improved.

The method used to transform and profile the data is simple but powerful. The major changes to previous data visualisation are twofold. First, to set aside the uncited papers as a separate category. Second, to group the cited papers into categories that are scaled relative to world average rather than to an absolute index, effectively using a log (2) scale (Table 1).

This will lead to the development of new single-figure metrics – some of which we have indicated in Table 2, such as mode, median and percentage greater than the world average – that provide a more direct reference to distribution characteristics. A measure such as "the percentage of outputs that exceed eight times the world average impact" is likely to be of particular interest because it is output with high-end impact that is most likely to contribute to research and economic innovation and competitiveness.

The outcome is that average citation impact relative to world, the traditional index of research performance, is still of enormous summary value but may no longer be sufficient as a pivotal indicator. This is a transitional problem for UK analysts because we are the first to see the patterns in this format. Once the same methodology is applied to other countries' data we will be able to use new indices, such as mode, median, and percentages in different parts of the distribution, to create a much more informative picture of research activity.

How good is the UK research base? Well, it's exactly as good relative to France or Germany as we always thought (ADAMS, 1998; MAY, 1997). But none of these countries has a distribution of research quality that conforms to what we had assumed when we only knew about the average.

What are the implications for policy of the percentage of the UK's output that lies above and below world average? The relative volume of papers below world average is likely to provoke discussion about an appropriate management response. Where is it located? Is this an essential platform to the peak of higher quality work? What is the intellectual link between the peak and platform? How does it vary between disciplines and institutional types? Crucially, how has it changed over time?

Similarly, there will be renewed interest in analysing how the UK has achieved its overall improvement. The balance between reduced activity at the low end and increased activity at the high end can be drawn out, by time and discipline, and the extent to which each has affected the average which current indicators track can be determined.

*

We are grateful to Louise Jackson and Tom Letcher at *Evidence* Ltd for supporting the background development of this work. Henry Small and Nancy Bayers at Thomson Scientific commented on drafts. The work was funded by the UK Office of Science and Innovation as part of the development of the UK's science and engineering base target indicators, and John Kirk, Roger Louth and Keith O'Nions at OSI all asked difficult questions that guided the outcomes.

References

ADAMS, J. (1998), Benchmarking international research. *Nature*, 396 : 615–618.

EHRENBERG, A. (1988), *Repeat-buying: Facts, Theory and Applications*. New York: Oxford University Press.

EUROPEAN COMMISSION (2003), *Third European Report on Science and Technology Indicators*. D G Research, Brussels. ISBN 92 894 1795 1.

EVIDENCE (2005), *PSA Target Metrics for the UK Research Base*. A report by Evidence Ltd for the Office of Science and Innovation. http://www.ost.gov.uk/research/psa_target_metrics.htm

FELLER, W. (1943), On a general class of 'contagious' distributions. *Annals of Mathematical Statistics*, 14 : 389–400.

GARFIELD, E. (1955), Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, 122 : 108–111.

KING, D. A. (2004), The scientific impact of nations. *Nature*, 430 : 311–316.

LEYDESDORFF, L., BENSMAN, S. (2006), Classification and powerlaws: the logarithmic transformation. *Journal of the American Society for Information Science and Technology*, 57 1470–1486.

MAY, R. M. (1997), The scientific wealth of nations. *Science*, 275 : 793–796.

MICHELSON, E. S. (2006), Approaches to research and development performance assessment in the United States. *Science and Public Policy*, 33 : 546–560.

MOED, H. F. (2005), *Citation Analysis in Research Evaluation*. Springer, Dordrecht. ISBN 4020 3713 9.

MOED, H. F., GLÄNZEL, W., SCHMOCH, U. (2004), *Handbook of Quantitative Science and Technology Research.* Springer, Dordrecht.

PRICE, D. DE SOLLA (1965), The scientific foundations of science policy. *Nature*, 206 : 233–238.

SAVANI, V., ZHIGLJAVSKY, A. (2006), Efficient estimation of parameters of the negative binomial distribution. *Communications in Statistics: Theory and Methods*, 35 : 1–17.

SAVANI, V., ZHIGLJAVSKY, A., Efficient parameter estimation for independent and INAR(1) negative binomial samples. *Metrika* (in press a).

SAVANI, V., ZHIGLJAVSKY, A., Asymptotic distributions of statistics and parameter estimates for mixed Poisson processes. (in press b).

SMALL, H., Why authors think their papers are highly cited. *Scientometrics*, 60 (2004) 305–316.

WEINGART, P. Impact of bibliometrics upon the science system: inadvertent consequences? *Scientometrics*, 62 (2005) 117–131.

ZITT, M., RAMANANA-RAHARY, S., BASSECOULARD, E., Relativity of citation performance and excellence measures: from cross-field to cross-scale effects of field normalisation. *Scientometrics*, 63 (2005) 373–201.