

evidence

How good is the UK research base?

June 2006

Jonathan Adams

Chief Executive

Evidence Ltd



Higher Education Policy Institute

How good is the UK research base?

Jonathan Adams

Evidence Ltd

Introduction

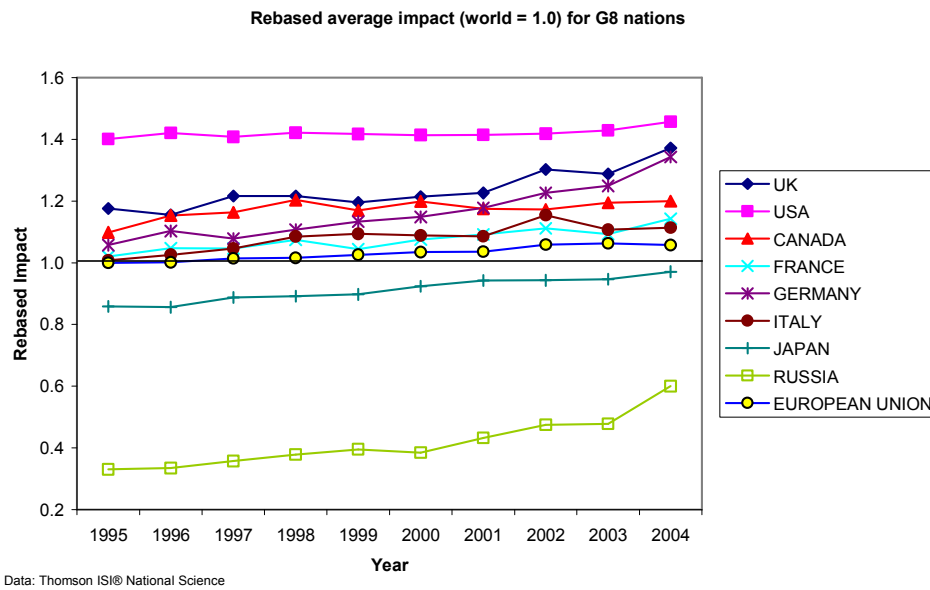
1. Conventional indicators of research performance provide a simple measure of the average performance of the national research base but no explanation of the distribution of quality. This report concerns a new way of looking at research quality by profiling that distribution and thereby gaining novel insights.¹

2. We need new indicators to unpack the averages to which conventional metrics refer. Research managers and policy makers, as well as analysts, need something more to inform their decision making. 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?

3. Most research performance analyses produce a single final metric and then use some broader reference as a benchmark. Depending on the benchmark, this might result in a measure of output relative to world outputs or it might be an average such as income per year, outputs per unit input, or citations per paper. An approach widely used in international comparisons, such as the EC's Science & Technology Indicators, is to capture research performance in terms of citation impact. A UK example of the use of a similar measure is the Office of Science & Innovation's (OSI) annual Public Service Agreement (PSA) Target Indicators report (Figure 1), which measures the impact of UK research relative to the world average.

¹ The work described in this report was developed under a contract for the UK Office of Science & Innovation.

Figure 1. Taken from Indicator 3.08 in the OSI's PSA Target Indicators report (2005 edition).



4. Averages are useful. They digest a lot of data and absorb outlier values. They seem easy to understand. But that understanding is generally based on assumptions: for example, that the data are normally distributed. If the data are actually skewed then the average will differ from the median (midpoint value) and mode (most common value). An average can therefore mislead interpretation of the true nature of the activity.

5. Most people, aware that metrics show that UK research performance is 'better than world average', will assume that the median of UK research is also around world average and that about half or more of UK outputs are therefore 'better than average'. In fact, an average says little about the spread of activity on either side. We do not know how much is 'much better', or how much is 'about average'. We know performance has changed, but we do not know whether that is due to real improvement, or to a reduction in poor performance.

6. For example, the UK's average performance across all subject fields improved between 1995 and 2004 relative to key competitors and the rest of the world (OST Indicator 3.08 in Figure 1). OST Indicator 2.03 (not shown here) 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 (OST Indicator 3.05, not shown here) decreased relative to other countries.

7. The UK is publishing slightly less, more of the papers being published are getting cited, and the average citation impact has improved. Our 'management' problem is in distinguishing hypotheses about the dynamics behind these trends. On the one hand, we could assert that 'the UK is getting even 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, if UK researchers had stopped writing marginal contributions that had no recognised value in their field then that would account for all three changes. There would be a reduction in "waste", not a high-end gain. The present metrics do not shed any light on this.

8. In this study, we have sought to move 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. The change is to look at the distribution of individual 'impact' values rather than the average value.

9. 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. They help us to analyse and interpret the data in a more sophisticated way than has been possible hitherto.

Methodology

Data source

10. The data are supplied by Thomson Scientific (formerly ISI) which maintains the most complete international data on research journal publications and their citations. Citations are references subsequently made to an article by later publications, so giving an indication of its impact. More highly-cited work is recognised as having a greater impact and high citation rates are correlated with other measures of research excellence.

11. In this section we show why the raw data make it difficult to deliver a simple interpretation, useful for policy purposes. We propose a transformation and categorisation to make rational data profiling more feasible.

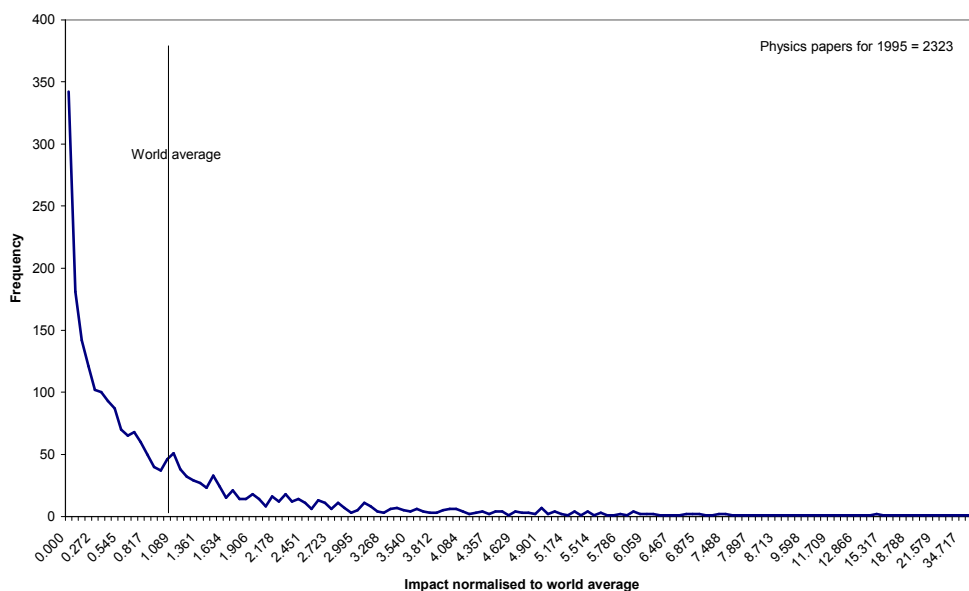
12. In the following section, we analyse total UK publication data from the most recent 10-year period 1995 to 2004. The UK impact (citations per paper) is benchmarked against world average baselines for year and field. We then break these data down by time and then by discipline.

Physics as a data example

13. The problem with raw data is illustrated by a sample from 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 citation counts for individual papers and normalised this against the appropriate world measure.

14. There were over 2000 UK article records. We grouped the papers into about 100 categories equally spaced according to their citation impact. That is, we took the highest impact measure (which was over 43 times world average for the field and year) and then divided the data into categories of equal impact increments.

Figure 2. The distribution of citation impact for UK Physics in 1995



15. Figure 2 is challenging to interpret. The data are highly skewed. There is a long tail of excellent papers going up to citation values over forty times world average (for most values beyond about 8 times world average there seem to be only one or two papers). There is also a concentration below world average.

Data categorisation

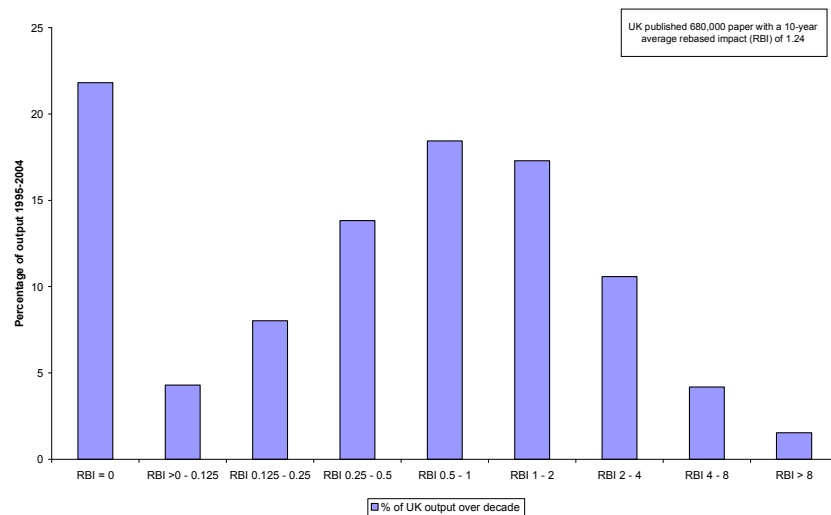
16. The approach we have adopted to take account of this uneven distribution, and to shed new light on our research performance is, first, to treat the uncited papers as a wholly separate class but keep them in the presentation; and, second, to group the impact values of cited papers according to their value relative to the world average. We have used eight categories into which to group 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 ($1/8^{\text{th}}$ world average) and above 8.0 times world average. Uncited papers form a ninth category.

UK output profile 1995-2004

17. The categorisation methodology was applied to the total UK output for the ten year period 1995-2004, with the citation count for each paper normalised (rebased) against a world benchmark for the year of its publication (Figure 3). This categorisation is a practical transformation that means the modified values approximate to a normal distribution².

² This is the equivalent of transforming the ReBased Impact on a log (2) scale. On a log scale 'zero' cannot be plotted, so it is obvious that the uncited are not part of the continuous distribution.

Figure 3. The distribution of citation impact for total UK research 1995-2004. World average impact = 1.0. Rebased impact (RBI) is average citations per paper normalised against world average.



18. The vertical axis scales the percentage of total output that falls in each of the categories in the horizontal axis.

19. Figure 3 is much easier to interpret than Figure 2, but how can we relate this profile to the UK average in Figure 1? UK average impact is well above world average but the modal (most common) group of cited papers (Figure 3) 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.

20. This analysis does not say that the UK is doing less well than we thought. Similar analyses for France, Germany and other key competitor nations would almost certainly produce pictures of research performance that looked like Figure 3. What the new methodology says is that we actually have a different – and probably more concentrated - distribution of excellence than analysts have assumed.

21. 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 reason why the UK average is greater than 1.0 despite the

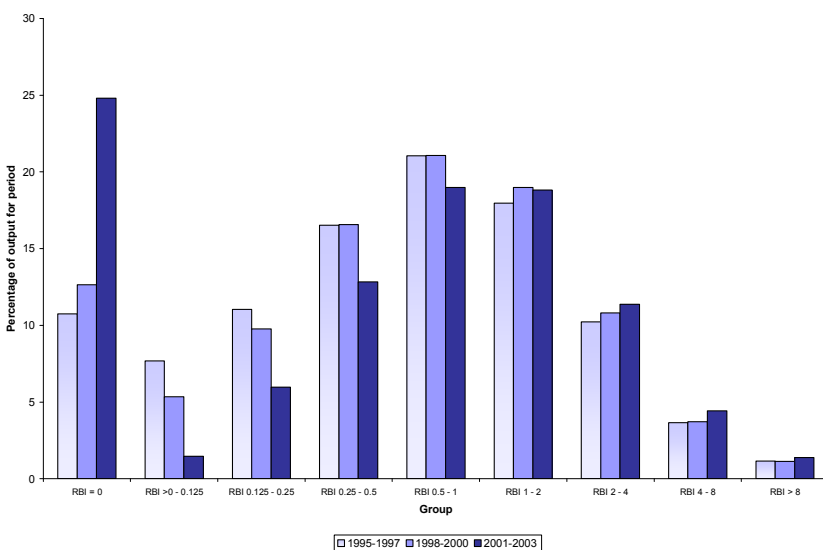
position of the mode and the number of uncited papers is the real value of the papers which are cited more than four times the world average. These papers, with many times the world average citation count for their year, pull up the UK overall position. It is more and more clear that the critical part of the UK's performance is not the bulk activity around the centre but the outstanding high-end performance.

22. 'Highly-cited' is a criterion of publication excellence used by Thomson Scientific, to which Evidence Ltd has drawn attention in other analyses. These papers are typically the top 1 per cent of papers for each subject in a year, and their threshold RBI value would generally be over 10. The percentage of papers in the category $RBI > 8$ is slightly less excellent, usually covering the top 1.5-2 per cent of UK papers, but has robust comparability.

UK output profile by time and discipline

Time

Figure 4. Tracking the distribution of citation impact for all UK research using three-year windows for the period 1995-2003



23. Figure 4 uses three 3-year windows to explore the data further, dropping the most recent year with its high proportion of 'not yet cited' papers. The shift across years is easy to discern. The category $RBI = 1-2$, just above world average, is pretty level and acts as a 'pivot'. The

categories with lower impact all fall over the period while other categories rise, in accord with indicators of a trend of improvement for the UK. The uncited category increases over the period, but that reflects the time it takes for citations to be generated: nothing sinister should be read into this element.

Discipline

24. A diversity of 'discipline' charts could be created in the format shown in Figure 3. For testing, we chose coarse-level categories (here called disciplines) which equate to groups equivalent to 'Schools' in a University.

25. Table 5.a and 5.b. summarise the profiles and key statistics for these disciplines, ranked by the proportion of output with normalised or ReBased Impact (RBI) greater than world average. There are two impact values. We have calculated the UK's average recent (last 5 years) RBI for each category. We have also calculated a median RBI which is the value in the midpoint of the distribution, below and above which lie half of the normalised impact values for individual papers.

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

27. The analyses do not show that average RBI, the traditional indicator of research performance, is 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.

28. 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. Again, we must point out that the same pattern would very

likely emerge for any country analysed in this way. As for the UK as a whole, in most fields more than half of the UK's papers have a citation impact below world average. Even in the best performing disciplines the median only stands at 0.75 world average RBI and no more than 41 per cent of papers have an impact exceeding world average RBI.

Table 5 Summary of research impact profiles for the UK at the level of Thomson journal categories (disciplines)

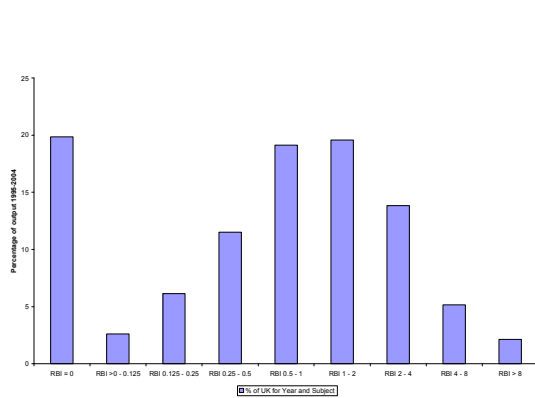
	Total output 1995-2004	Percentage of output				% > World Average	RBI Average 2000-04	RBI Median 1995-04
		Uncited	RBI < 1	RBI 1-4	RBI > 4			
Plant & Animal Science	38582	19.9	39.4	33.4	7.3	40.7	1.51	0.73
Chemistry	60022	19	44.6	31	5.4	36.5	1.23	0.63
Ecology/ Environment	16884	20.9	42.8	30.5	5.8	36.4	1.4	0.64
Geosciences	22939	21.4	42.2	30.8	5.6	36.4	1.33	0.65
Mathematics	9596	35.7	29.6	28.3	6.4	34.7	1.29	0.52
Molecular Biology & Genetics	23805	10	55.7	28.5	5.8	34.3	1.27	0.61
Physics	61205	23.2	42.6	27.5	6.7	34.2	1.42	0.54
UK total & average	750376	21.8	44.6	27.9	5.7	33.6	1.28	
Engineering	55236	35.4	31.6	25.5	7.4	32.9	1.08	0.39
Clinical Medicine	179247	20.8	48.3	25	6	30.9	1.21	0.45
Social Sciences, general	36108	35.6	34.1	24.7	5.5	30.2	1.05	0.37
Economics & Business	15236	36.9	37.2	21.4	4.5	25.9	0.93	0.31

29. The following four graphs profile a spread of subjects. The standard bell-shaped profile is common to most but not all. Some curves are flatter because more papers are uncited (e.g. Mathematics), or asymmetric because there is a relative excess (e.g. Geosciences) or deficit (e.g.

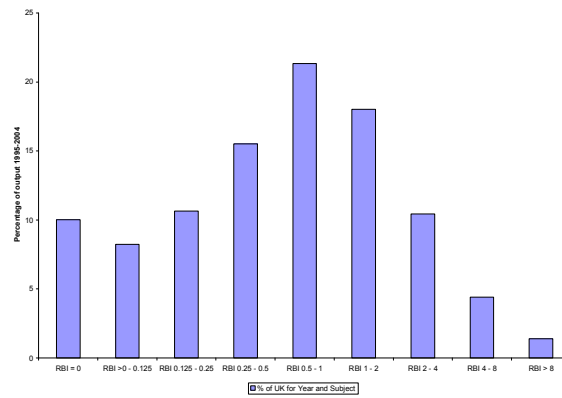
Molecular Biology) above world average. There is variation in the distribution between those disciplines where the UK average impact is high and where it is low. As noted in the examples above, the specific factors that contribute to the outcome can be teased out via these profiles where previously they were hidden by average indices.

Figures 6-9. The profiled distribution by journal category of citation impact for UK research 1995-2004.

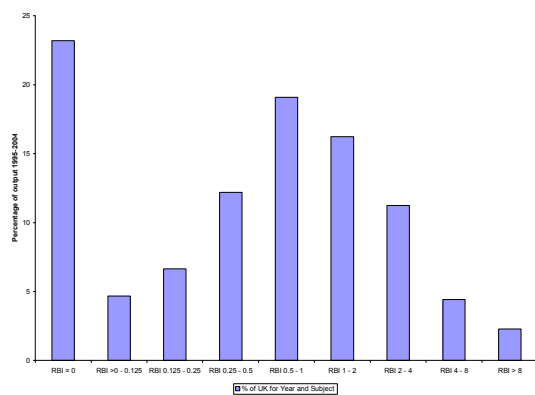
Plant and Animal Sciences



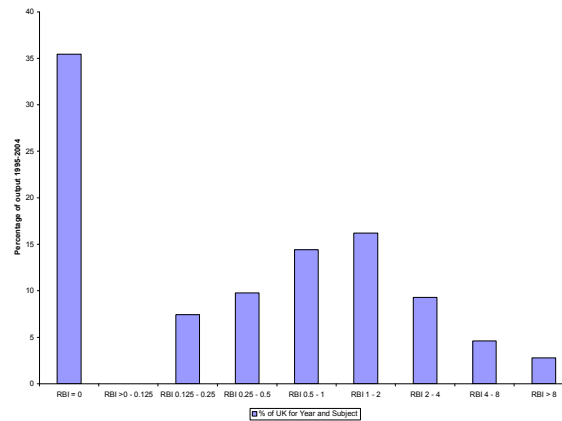
Molecular Biology & Genetics



Physics



Engineering



Conclusion

30. How good is the UK research base? Well, it is exactly as good relative to France or Germany as we always thought. 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.

31. The profiled distribution for the UK over the ten-year period (Figure 3 and Table 5) 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 the profiles will almost certainly change perceptions about the interpretation of bibliometric indices.

32. The relatively large proportion below world average is offset by the smaller but critically important component that is more than 4 and more than 8 times above world average. This is the reason for the UK's share of the world's highly cited papers, where it is second to the USA.

33. 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 policy and management response. Where is it located? Is this an essential platform for 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?

34. Similarly, there will be renewed interest in analysing how the UK has achieved the overall improvement shown in Figure 1. 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.

35. For the future, how can the new methodology be applied? It takes bibliometrics over the threshold from historical indicators to a tool that can support policy judgements and management action. The category-based profiles work well and demonstrate differences between areas (Table 5). 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 nationally and institutionally. Most of the profiles conform to a bell-shape, so the disaggregated research areas broadly follow the pattern of the total national curve.

36. There is also a timely target for these metrics. The methodology provides a route to creating a subject-based profile for every UK institution

submitting to the RAE2008. This can be benchmarked against the panel grades. After 2008, each institution can be tracked against its departure point and against the generic UK profile. Exceptional improvements, or decline, in research performance can be spotted. The methodology could provide an 'assessment-lite' system with a dipstick to trigger more intensive reviews.