Commentaries

The geography of scientific citation

As the marketplace for research and education becomes ever more global, there is growing interest in international comparisons of standards and quality in different countries and places. Rankings of universities according to various criteria, ranging from the quality of the educational experience imparted, to their research prowess, have become regular news items in English-speaking countries (University of Illinois Library, 2002). In some countries such as the United Kingdom where the core funding of universities comes from central government, such assessments are now used routinely for resource allocation (HERO, 2002).

International comparisons are difficult, however, with few published rankings despite rapidly increasing international migration to pursue research at the graduate level. In the USA 20% of all full-time graduate students are non-US citizens (NCES, 2002); in the United Kingdom the comparable figure is 25% (HESA, 2002). As there are no global rankings, most decisions to pursue research at a particular institution must be based on casual perceptions of quality, cost, and overall value for money. To examine the research quality of universities worldwide, citation indices provide a first approach to the problem (Oswald, 2002). The ISI's HighlyCited database (http://www.isihighlycited.com) which is currently (December 2002) composed of the top 100 or so cited individuals in fourteen scientific fields is a manageable source for classifying scientists not only by their field but by their institution, their location, and the country in which they work.

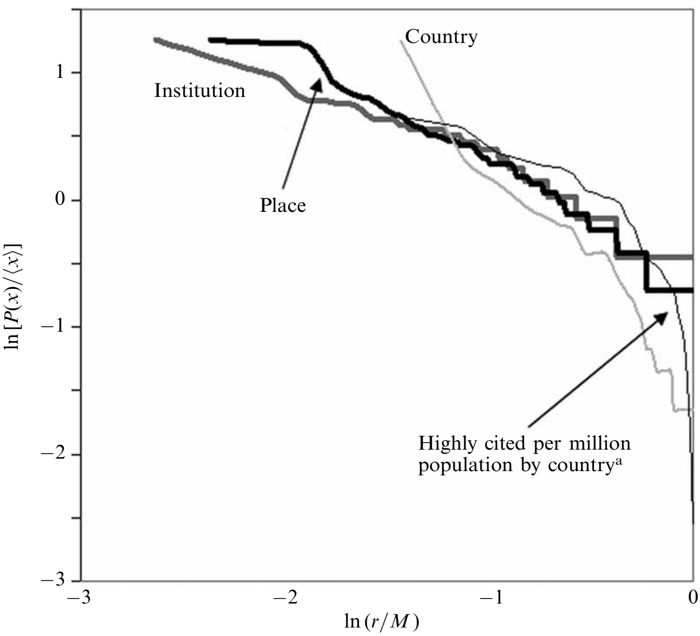
This source has many limitations, for it excludes mathematics (other than physics), the social sciences, and the humanities, and is thus biased towards the medical sciences. Moreover, it is under rapid development with the size of the database planned to increase to twenty-one subject area categories with over 4500 names by the end of 2003. There are also limitations to the 'institution' data with respect to joint, part-time, and related appointments which clearly complicate any indices we may derive (M McVeigh, private communication, 2003, Institute of Scientific Information, Philadelphia, PA). Nevertheless, I consider that a preliminary analysis is worthwhile and I have taken data from twelve of the fourteen categories listed. From a detailed scrutiny of each entry, I have used data on 1222 scientists. A significant minority of the scientists cited—some 30%—work in research institutes, hospitals, and private firms, albeit many connected to university institutions, but to maintain comparability between cities and countries, I have retained these data.

The pattern of concentration that this analysis reveals is remarkable: 1222 scientists work in 429 institutions which are located in 232 places in 27 countries. Almost half these scientists are in 50 institutions in 5 countries, most being in the United States. In table 1 (see over) I list the top 20 institutions in terms of the number and percentage of scientists cited. These institutions contain nearly 30% of the scientists, and are all located in the USA with the exception of University College London and the University of Cambridge. The concentration increases as the data are aggregated from institution to place and thence to country. In figure 1 (see over) I show these aggregated data sets as Zipf plots where I have plotted the logarithm of the number of scientists for each institution, place, and country, normalised by their means, against the logarithm of their normalised ranks. Collapsing each data set in this way shows quite clearly how the

Table 1. Top twenty ranking of instituting by number of highly cited scientists.

Rank	Research institution	Number of highly cited scientists	Percentage of highly cited scientists
1	Harvard University	52	4.3
2	Stanford University	36	2.9
3	University of California, San Diego	30	2.5
4	MIT	26	2.1
5	NIH[a] National Cancer Institute	19	1.6
6	University of California, San Francisco Cornell University	17	1.4
8	University of California, Berkeley University College London (UK)	16	1.3
10	CalTech	15	1.2
11	NIH[a] Allergy & Infectious Diseases	13	1.1
12	Johns Hopkins University University of Cambridge (UK) University of Washington, Seattle Washington University, St Louis	12	1.0
16	University of California, Davis University of Texas Cancer Center	11	0.9
18	Michigan University Northwestern University Yale University	10	0.8

[a] NIH National Institute of Health.



[a] This is the plot for countries, normalised by population in millions, which illustrates a different pattern of concentration from the basic data. I have fitted linear plots to the basic data using $\ln[P(x)/\langle x\rangle] = k - \alpha \ln(r/M)$, where for institutions $\alpha = 1.049$ ($R^2 = 0.962$), for places $\alpha = 0.816$ ($R^2 = 0.938$), and for countries $\alpha = 1.997$ ($R^2 = 0.949$). All these values are significantly different from zero at the 99% level.

Figure 1. Rank-size distributions of highly cited scientists.

concentration increases as the data are aggregated into places and countries. I have fitted power laws to these plots based on $[P(x)/\langle x \rangle] \sim (r/M)^{-\alpha}$, where $P(x)$ is the number of cited scientists at rank r, $\langle x \rangle$ is the mean number of cited scientists, and M is the number of institutions, places, or countries for each of the three respective aggregations (Redner, 1998). The value of the power α is related to the degree of concentration.

In table 2 I show the top ten countries in terms of the number of scientists and places where they work. The ratio of scientists to places provides another measure of the concentration, with the implication that the larger the number of highly cited in each country, the more likely they are to be highly concentrated in a small number of places. If we normalise the data by population, we get a slightly different picture; the top five are now Switzerland, United States, Sweden, United Kingdom, and Israel with smaller countries becoming more significant. We might even consider normalising by the square of population, reflecting the potential interaction within a population although, for the scientific enterprise, this probably is not a good measure of where such interaction is possible. All this does is to sharpen the index even further, with small countries dominating. Similar analyses for institutions and places are more complicated as the choice of a population for the normalisation is uncertain. College towns begin to dominate, for example.

Table 2. Top ten ranking of countries by highly cited scientists.

Rank	Country	Number of highly cited	Number of places	Concentration: scientists/places	Highly cited per million population
1	United States	815	90	9.06	3.16
2	United Kingdom	100	24	4.17	1.72
3	Germany	62	21	2.95	0.78
4	Canada	42	15	2.80	1.53
5	Japan	34	14	2.43	0.27
6	France	29	11	2.64	0.50
7	Switzerland	26	5	5.20	3.78
8	Sweden	17	2	8.50	1.96
9	Italy	17	10	1.70	0.29
10	Australia	17	9	1.88	0.96

A particularly graphic indication of the basic pattern is illustrated in figure 2 (see over) where I have mapped the main locations of places by circles proportional to the number of cited scientists. Of the most highly cited scientists 40% work in 10 places of which 9 are in the USA. These locations bear out our perceptions of where the world's top institutions are most heavily concentrated: on the west coast of the United States, the Boston – Washington megalopolis on the east coast, central London, Chicago, and interestingly in the cluster of towns around Research Triangle Park in North Carolina. I have not yet examined the local detail of where these institutions are located, but casual knowledge suggests that these are even more highly clustered at ever finer scales. For example, the institutions in Boston are all within a two-mile radius of the MIT Museum whereas in London they are within a three-mile radius of the British Museum. At an even more local scale in central London, for example, the majority of the scientists cited are located within half a mile of Euston station in Bloomsbury.

This analysis is of course limited by the bias in the ISI data to English-speaking countries, to the medical sciences, and to full-time research rather than education. Although for US institutions, there is only a 40% correlation with the top 50 universities in terms of doctoral programs most recently ranked by *US News and World Report* (2002), this simply indicates the fact that size is all important in the rankings produced

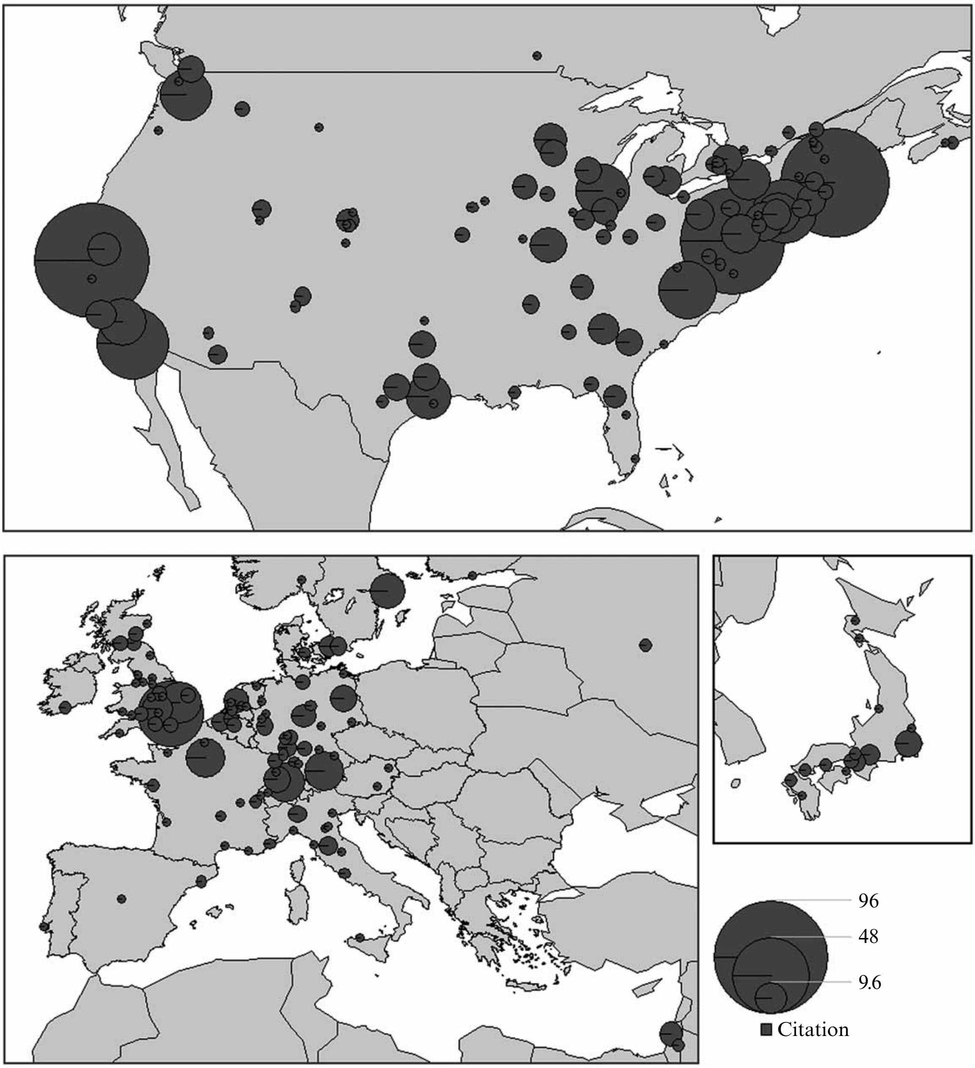


Figure 2. The geographical distribution of the highly cited.

from the ISI data. The correlation in the United Kingdom with *The Times* newspaper (2002) ranking is much the same at 43%. What this analysis reveals is a pattern of much greater concentration than I originally anticipated from other literature on the geography of the modern economy, notwithstanding the influence of history and the effects of national policy on the location of research centres (Matthiessen and Schwarz, 1999).

I consider there are important implications from these findings for national educational policy, and the distribution of research resources, especially during a period when governments and institutions are competing ever more intensely to gain and retain the best, and to build critical mass. There are issues involving the choice of the best graduate schools implied in the analysis. This analysis puts the geographical distribution of scientific wealth in perspective (May, 1997). In a British context it will be surprising to many academics and politicians that, of the 1222 scientists cited, only 100 (about 7%) are located in the United Kingdom based in 38 institutions (about 9%) of the 429 associated with these citations.

Considerably more work can be done with this source for within the next 12 months much more data over a wider number of fields will be available and it will be possible to mine the data for changes in citations at the margin. We do not yet know how robust these indicators of geographical concentration actually are, although I suspect that they will not change very much on an annual basis. I also suspect that from year to year there may be considerable volatility in the actual names of those who form the HighlyCited database but that once we aggregate these across institutions, places, and countries, such volatility will begin to disappear. However, what we are most interested in is how different places and countries are changing over decades rather than years for this will give us some index of how patterns of global research quality are changing which is of central importance to science policy everywhere. These analyses will be forthcoming in due course.

Michael Batty
Centre for Advanced Spatial Analysis at University College London, 1 – 19 Torrington Place, London WC1E 6BT, e-mail: m.batty@ucl.ac.uk

Acknowledgements. The Institute for Scientific Information gave permission for this analysis to be published. Further details of this analysis are given on the author's website at http://www.casa.ucl.ac.uk/citations/

References

HERO, 2002, "UK Research Assessment Exercise", Higher Education and Research Opportunities, http://www.hero.ac.uk/rae/index.htm
HESA, 2002, "Student tables", Higher Education Statistics Agency, http://www.hesa.ac.uk/holisdocs/pubinfo/stud.htm
Matthiessen C W, Schwarz A W, 1999, "Scientific centres in Europe" *Urban Studies* **36** 453 – 477
May R M, 1997, "The scientific wealth of nations" *Science* **275** 793 – 795
NCES, 2002, "Enrolment in postsecondary institutions, Fall 2000 and financial statistics, fiscal year 2000", National Center for Education Statistics, Institute of Education Sciences, US Department of Education, Washington, DC, http://nces.ed.gov/pubs2002/2002212.pdf
Oswald A O, 2002, "A crisis of quality" *Education Guardian* 15 November, http://education.guardian.co.uk/higher/comment/story/0,9828,840892,00.html
Redner S, 1998, "How popular is your paper? An empirical study of the citation distribution" *European Physics Journal B* **4** 131 – 134
The Times 2002, "Good university guide", http://www.times-archive.co.uk/news/pages/tim/2000/04/14/timguggug01002.html
University of Illinois Library, 2002, "Graduate and research program rankings", Education and Social Science Library, University of Illinois at Urbana-Campaign, http://www.library.uiuc.edu/edx/rankgrad.htm
US News and World Report 2002, "College rankings", http://www.usnews.com/usnews/edu/college/rankings/ranknatudoc_brief.php