

Profiles, not metrics.

Jonathan Adams, Marie McVeigh,
David Pendlebury and Martin Szomszor

January 2019

Web of Science

 **Clarivate**
Analytics

Authors

Professor Jonathan Adams is Director of the *Institute for Scientific Information (ISI)*, a part of *Clarivate Analytics*. He is also a Visiting Professor at King's College London, Policy Institute, and was awarded an Honorary D.Sc. in 2017 by the University of Exeter, for his work in higher education and research policy.

Marie McVeigh is Head of Editorial Integrity as part of the Editorial team within the *Institute for Scientific Information*. Originally a cell biologist from the University of Pennsylvania, she has been working and publishing on journal management and intelligence with *ISI* and its predecessor bodies within *Clarivate* since 1994. Her recent work on JCR enhancement added article-level performance details and data transparency to support the responsible use of journal citation metrics.

David Pendlebury is Head of Research Analysis at the *Institute for Scientific Information*, a part of *Clarivate Analytics*. Since 1983 he has used *Web of Science* data to study the structure and dynamics of research. He worked for many years with *ISI* founder, Eugene Garfield. With Henry Small, David developed *ISI's* Essential Science Indicators.

Dr. Martin Szomszor is Head of Research Analytics at the *Institute for Scientific Information*. He was Head of Data Science, and founder of the Global Research Identifier Database, applying his extensive knowledge of machine learning, data integration and visualization techniques. He was named a 2015 top-50 UK Information Age data leader for his work with the Higher Education Funding Council for England to create the REF2015 Impact Case Studies Database.

About Institute for Scientific Information (ISI)

ISI is the 'university' of the Web of Science Group at *Clarivate Analytics*: it maintains the knowledge corpus upon which *Web of Science* and related information and analytical content, products and services are built; it disseminates that knowledge internally through reports and recommendations and externally through events, conferences and publications; and it carries out research to sustain, extend and improve the knowledge base.

About Web of Science

Web of Science is the world's most trusted and largest publisher-neutral citation index, powering global discovery and citation analytics across the sciences, social sciences and art & humanities. With over 1.4 billion cited references going back to 1900 and millions of users per day – from leading government and academic institutions and research intensive corporations – *Web of Science* citation network serves as the foundation for the *Journal Impact Factor*, *InCites* and other powerful and trusted citation impact measures. The *Web of Science* helps researchers, research institutions, publishers and funders discover and assess the citation impact of over a century of research publications found in the most prestigious journals, books, and conference proceedings.

To learn more, visit: clarivate.com/products/web-of-science

In this report, we draw attention to the information that is lost when data about researchers and their institutions are squeezed into a simplified metric or league table. We look at four familiar types of analysis that can obscure real research performance when misused and we describe four alternative visualizations that unpack the richer information that lies beneath each headline indicator and that support sound, responsible research management.

We are surrounded by analyses that claim to measure relative performance among people and organizations. University managers evidently use them, disregarding counter-arguments offered by informed analysts and to the dismay of researchers. Critical discussion about the credibility of university rankings is endless, but they continue to be published. We ask: why are simplistic analyses, such as single-point metrics and linear rankings, so popular?

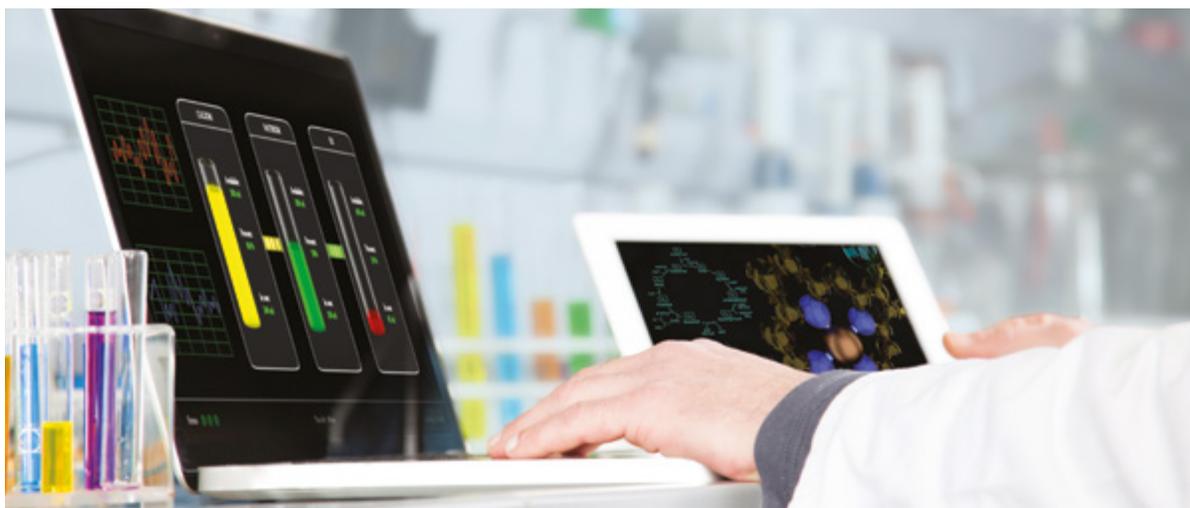
Summary statistics and league tables have innate appeal. We want ‘to see who does best’, taking an analogy from sports. But a sports’ league table is the product of a series of matches between similar members of a defined group, with the table headed up by whichever currently has the better balance of victories in direct and explicitly matched competition. A league table is a one-dimensional ranking based, sensibly for its specific purpose, on the single dimensions of the paired matches.

Research is not one-dimensional: the process is complex and no two projects are identical. Nor do research organizations have a single mission: they teach as well as research; their research may be blue-skies, analytical, applied, collaborative, societal or industrial; and their activity is spread across many disciplines, each with its own academic characteristics.

Single-point metrics have value when applied in properly matched comparisons, such as the relative output per researcher of similar research units in universities. That may tell us about real differences in ‘similar’ research. But the information is limited and an individual (or isolated) metric can be misused if it is a substitute for responsible research management, for example in academic evaluation without complementary information, or even as a recruitment criterion.

University rankings take a set of variables to ‘picture’ an organization, using proxy data spread across activities and disciplines. Each variable is indexed: scaled to link counts, money, impact, time and other incompatible items; and then weighted to bring different items together in a final score. Without well-informed data management, that number may have only a distant relationship to the rich diversity of university life.

For every over-simplified or misused metric there is a better alternative, usually involving proper and responsible data analysis through a graphical display with multiple, complementary dimensions. By unpacking the data and placing the metric against a background or setting it in a wider context, we see new features and understand more. The examples that follow show how easy this is and how much it improves our ability to interpret research activity.



Individuals: h-index vs the beam-plot

A widely quoted but poorly understood way of characterising a researcher's publication and citation profile is the h-index, created by physicist Jorge Hirsch (2005). It reduces a list of publications and their citation counts to a single number: a researcher (or group or even country) with an index of h has published at least h papers each of which has subsequently been cited at least h times.

The h-index depends on career length and discipline because citation counts accumulate over time at rates that vary between research fields, so it provides no proper comparability between individuals; it usually excludes non-journal publications; and it is mathematically inconsistent (Waltman and van Eck, 2012).

An alternative approach proposed by Lutz Bornmann and Robin Haunschild, Max Planck Institute (Bornmann and Haunschild, 2018), puts a researcher's articles into a context suitable for comparison. Each paper's citation count is 'normalized' by the average for journals in their same category and publication year, and that value is converted to a percentile. This provides a better measure of central tendency than an average because citation distributions are so skewed. A percentile of 90 means that a paper is among the 10% most cited and the other 90% have achieved less citation impact. The median score is 50: the average impact among publications ranged within a common scale between 0 and 100.

The beam-plot can be used for a fair and meaningful evaluation. It quickly conveys information that the h-index never suggested. This researcher's average percentile is significantly better than 50, the central impact in the fields where they published. Their median annual impact was below that benchmark in early years but can be seen to move above the average over time.

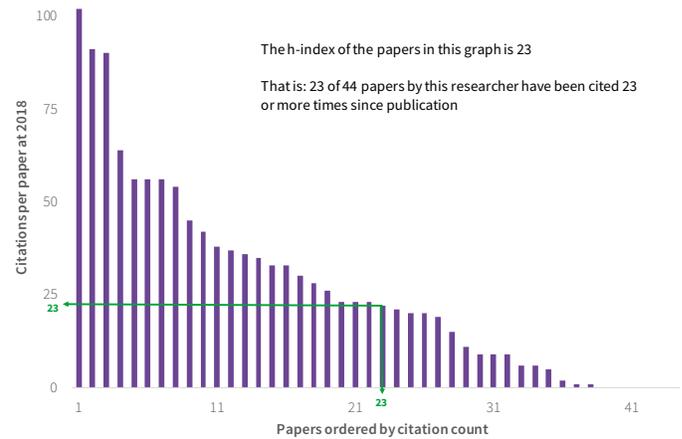


Figure 1. In this example h -index = 23 for a researcher who is an author or co-author on 44 citable journal articles over a 15-year period. Output included reports and proceedings that cannot be analysed in this way. Graphing the data reveals the spread, skew, and presence of relatively highly cited items buried under the 'h' value. Uncited items disappear.

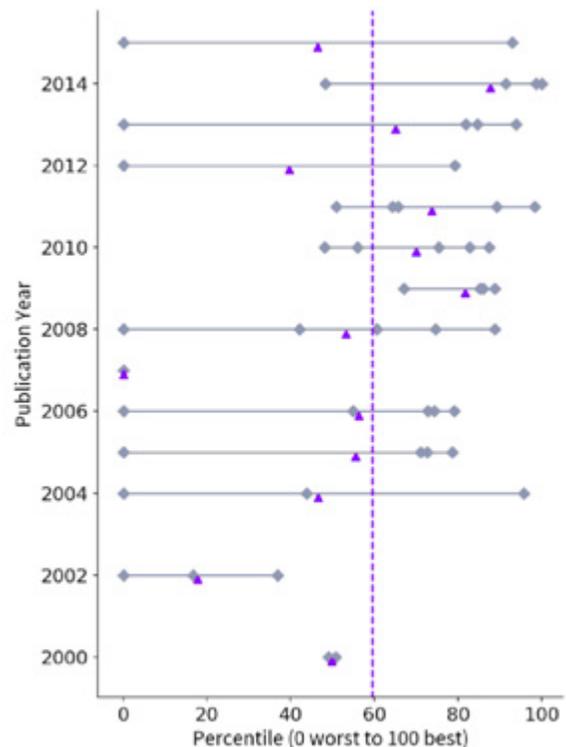


Figure 2. A beam-plot of the data in Figure 1. Each article is compared to its own reference set but all use a common 0-100 percentile scale. The ranges of each year's article percentiles are shown (grey marks, across the beam) with their annual median (purple mark, a pivot). The benchmark line is the researcher's overall average: the 59th percentile.

Journals: JIF vs JCR distributions

Quantitative research evaluation usually looks at collections of publications and compares average citation counts with category-based benchmarks. Evaluators may also look at the journals in which the articles are published.

The *Journal Impact Factor (JIF)* is a familiar indicator in general use. It was developed by Eugene Garfield, the founder of the *Institute for Scientific Information* Garfield (1955) raised the idea of publication ‘impact’ and created a ‘journal’ impact factor (Garfield and Sher, 1962) to help select journals for the new *Science Citation Index (SCI)*. That anticipated the first *Journal Citation Reports* in 1975.

JIF2 (that is, based on two years of journal data) has two elements: the numerator, the number of cites in the current year to any items published in the journal in the previous two years; and the denominator, the number of substantive articles and reviews published in those two years. The basic elements of it can be adapted to consider either shorter or longer time intervals. A *JIF* based on the previous year’s articles alone would give weight to rapidly changing fields; a *JIF* using 5 or 10 years of items and their citations could include a greater proportion of the journal’s citation performance in a given year.

Quantitative journal comparators are used by librarians who need to manage many subscriptions within a limited budget and by publishers who track the performance of their serials. The problem is that *JIF*, developed for responsible use in journal management, has been irresponsibly applied to wider research management.

To offset this, the 2018 *Journal Citation Reports* revised journal profiles with a richer data context. For example: a bar graph gives the *JIF* value as Percentile Rank in Category allowing quick visualization of the quartile rank; and, essential for researchers, the citation contribution for individual items are shown within the overall spread of citation frequency across items.

The new journal profile makes it clear that *JIF* is a summation of a larger and more complex data pool. It may be a useful tool for journal managers but by itself it only tells the research manager part of what they need to know about the journal or any article’s value.



Figure 3. Left: *Journal Impact Factor* Trend graph for *EMBO Reports* shows *JIF* and percentile rank in category. Right: *Citation distribution 2017* shows medians and overall spread.

Institutes: Average Citation Impact vs Impact Profile™

The limitations created by packing research activity data into a single point metric become even more evident when moving from individuals and journals to research groups and institutions.

We looked at the research publications of two biomedically orientated research units. Their specific missions differed but their portfolios were broadly similar; they were supported by a similar range of research funding organizations; and they had a similar historical background. To inform a review group looking at the two units, we gathered their research publication record over a recent 5-year period, which gave a total of about 1,250 journal articles: Unit A had published 845 papers while Unit B had published 403 papers.

Because citations accumulate over time at rates that are discipline-dependent, the citation counts for the units' papers need to be 'normalized' against the world average for the relevant journal category and publication year. This gives a Category Normalized Citation Impact value (CNCI - also referred to as Re-Based Impact RBI). The average CNCI was 1.86 for Unit A and 2.55 for Unit B (compared to a world average of 1.0).

These CNCI values have no statistical power, but a typical management assumption would be that the smaller unit is delivering greater 'impact' than the larger unit. Average impact indicators are deceptive, however, because several hundred individual publications contribute to the spread of separate CNCI values that lie beneath each average.

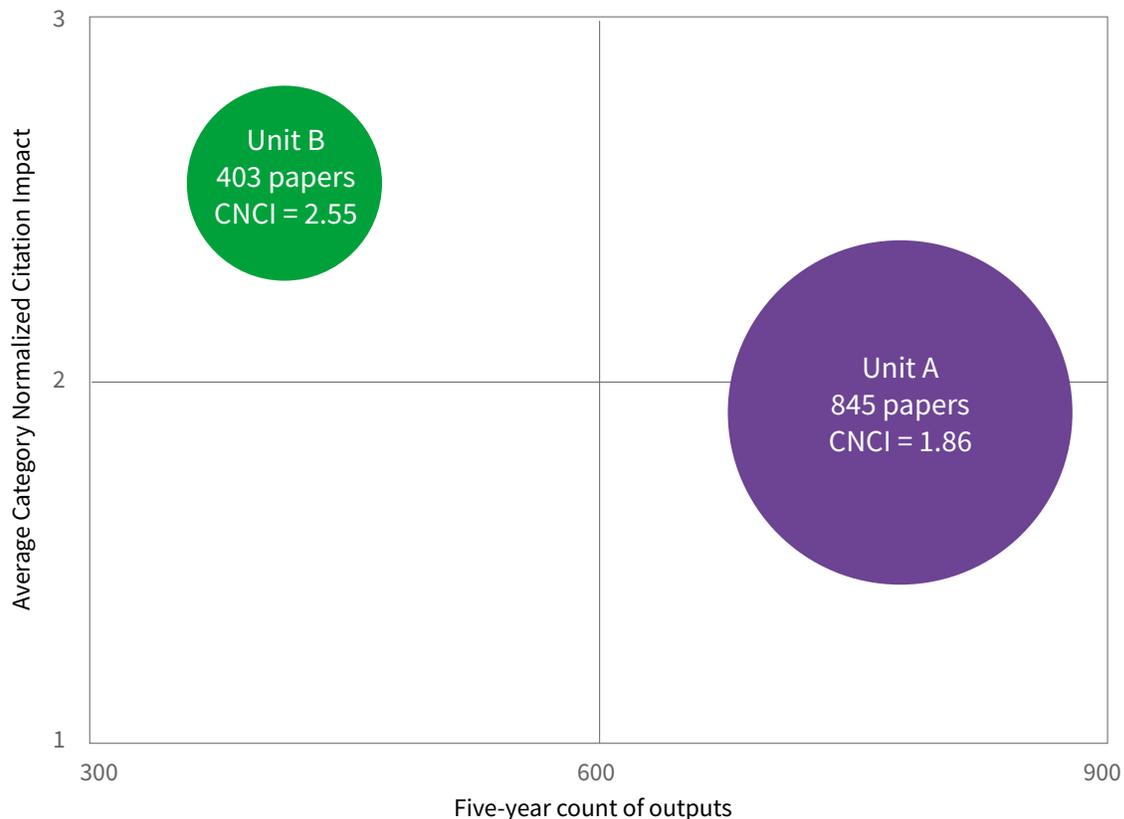


Figure 4. Relative five-year volume of publication output and average Category Normalized Citation Impact (CNCI) of two UK biomedical research units. Unit B has about half the output but a much higher average normalized citation impact than Unit A.

Citation counts are very skewed, with many low and a few high values in almost any sample. So, to visualize the spread we categorised the counts relative to the world average: first, above world average by summing up four categories or bins that cover from 1 to 2x world average citation impact, then 2-4x, 4-8x and over 8x.

At the same time, we take the counts from 1.0 to 1/2, then 1/2 to 1/4, and so on to create four bins below world average by impact range. The uncited papers we set in a separate ninth bin. This reveals the overall Impact Profile™ of each dataset, showing the real spread of more and less well-cited papers (Adams, Gurney and Marshall, 2007).

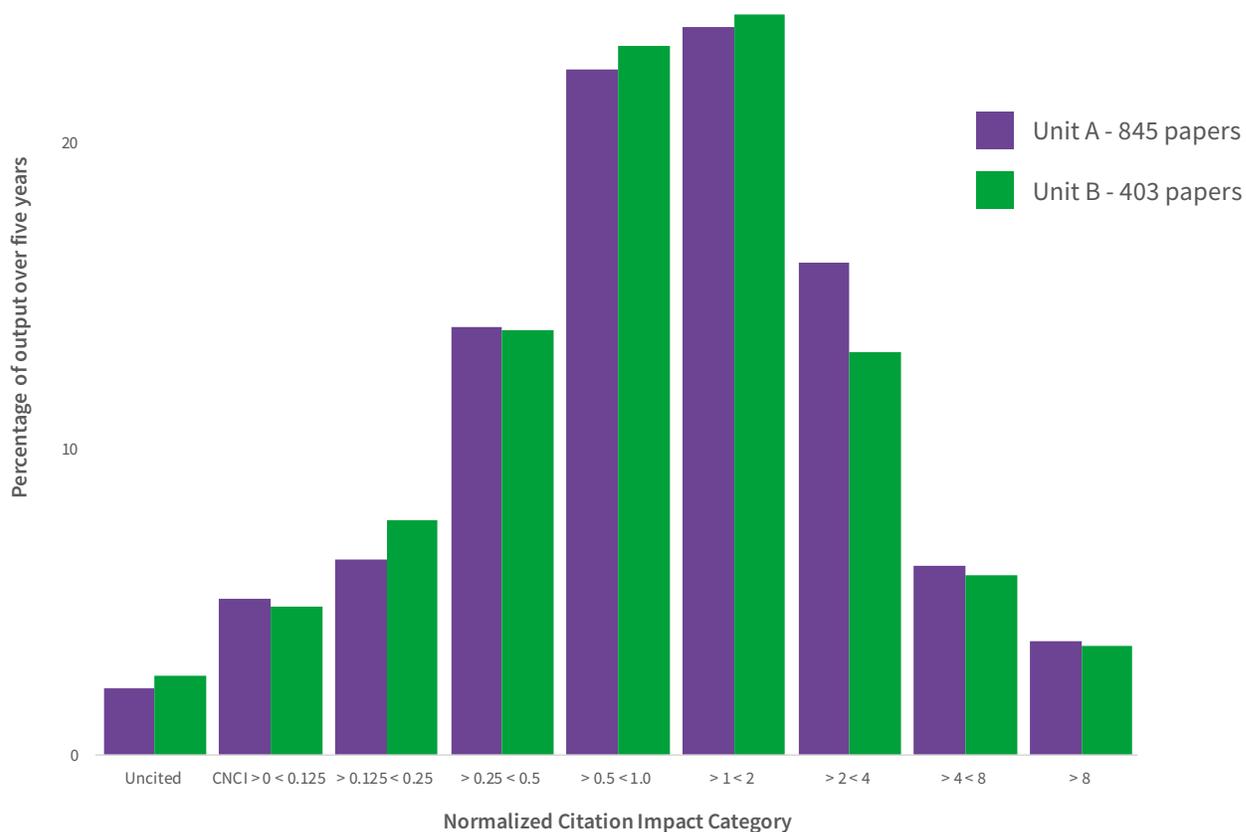


Figure 5. The Impact Profile™ of two UK biomedical research units over five years. The citation count of each paper is ‘normalized’ by the world average for that publication year and journal category (CNCI: see text) and allocated to a series of bins grouped around that average (world average = 1.0; uncited papers grouped to the left). Counts are shown as percentage output for each unit.

This procedure gives us a far more informative picture than the summary values in Figure 4. The profile looks rather like a normal (Gaussian) curve, distributed either side of the world average. We could ‘locate’ each unit’s overall average within that and check how much of their output is actually above and below that metric: more would be below for both.

Most importantly, we can immediately see that there is no substantive difference in the two Impact Profile™, which effectively visualize their research performance. By checking back to the original data, in fact, we find that the high average impact for Unit B is influenced by a single, very highly cited review in a leading journal.

Universities: Ranking vs Multifactor Research Footprint

A summary indicator such as ‘average citation impact’ compresses only one type of data. Even so, the research manager suffers significant information loss in using the average as compared to the picture in an Impact Profile™.

The information loss is even greater and useful content is much less readily unpacked when a league table brings together many types of data about diverse university activities in a single, ranked number.

Of the familiar annual ranking systems, the data collection is relatively broad and the algorithms are well-informed and balanced in *Times Higher Education’s* World University Rankings, which are also moderated for factors like institutional size so they avoid dominance by the largest old universities. Nonetheless, most people reading the rankings ‘know’ which institutions they expect to see at or near the top.

Scholars who have studied at any of these universities know their complexity and the diversity of activity — and achievement — that actually occur on campus. The difference is emphasised in the UK by two well-known London institutions: Imperial College, London, with a strong STEM tradition; and the London School of Economics, with a global reputation in the social sciences.

No one would normally seek to make a direct comparison between the two, let alone rank them on the same, global scale.

Even when comparisons are made between more similar well-established, multi-faculty universities with large medical schools, such as Edinburgh (29th) and Manchester (57th), the outcome of a ranking is uninterpretable. What do these positions mean?

The underlying reality is that any institution scores better on some parameters and less well on others, continuously varying its position relative to others. A global university ranking may be fun, but it is only a reference point. It hides far too much detail even for careful short-listing for students, let alone as a tool to inform management.

A more informative approach comes from the Research Footprint used in *Evidence* Ltd UK Higher Education Research Yearbooks between 2002 and 2014. A Research Footprint is a radar diagram, using multiple axes for multiple indicators, where each entity is compared to a common reference footprint, or a series of institutional ‘footprints’ share one picture.

Global universities	WUR position		UK universities
University of Oxford	1	1	University of Oxford
University of Cambridge	2	2	University of Cambridge
Stanford University	3	9	Imperial College London
MIT	4	14	University College London
CalTech	5	26	London School of Economics
Harvard University	6	29	University of Edinburgh
Princeton University	7	38	King’s College London
Yale University	8	57	University of Manchester
Imperial College London	9	78	University of Bristol
University of Chicago	10	79	University of Warwick

Table 1. The global league table position of the universities that were ranked highest in *Times Higher Education’s* World University Rankings (WUR) for 2018.

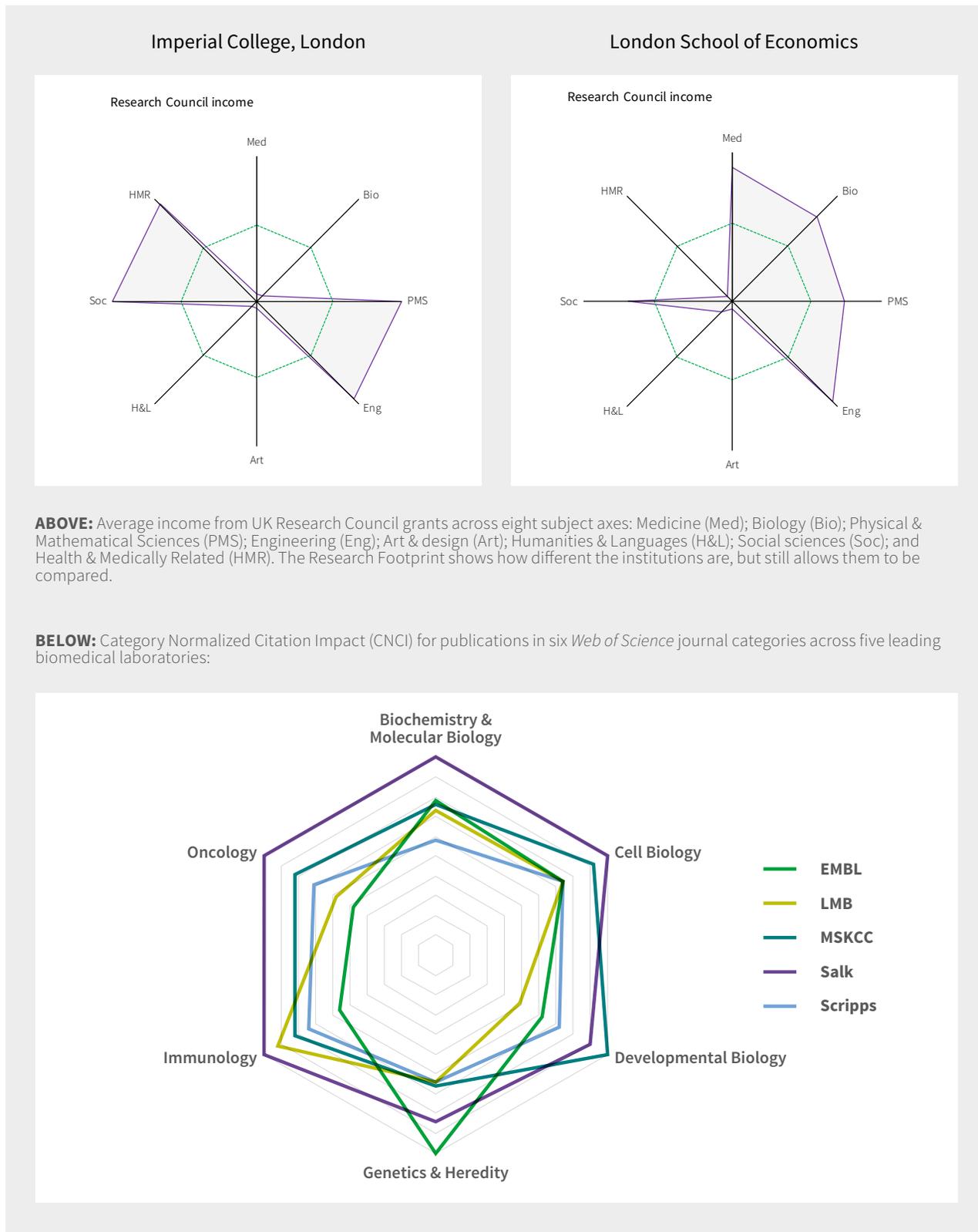


Figure 6. Research Footprints for the two UK higher education institutions (upper row) displays publication output by major discipline (similar diagrams could be used for funding, student and staff count, or citation impact) with a reference benchmark from an appropriate comparator group. The Research Footprint in the lower row compares the output of leading biomedical institutes in specific research categories in which they are active: in this case, no benchmark is needed.

Discussion

The point metrics (h-index, *Journal Impact Factor*, average citation impact) and the university ranking discussed in this report are all potentially informative but all suffer from widespread misinterpretation and irresponsible and often gross misuse. The alternative visual analyses are ‘picture profiles’ of research activity. They are graphical illustrations that: are relatively simple to produce; unpack a spread of much more valuable information; and support proper and responsible research management.

A beam-plot, not an h-index

The beam-plot is a single ‘picture’ of a researcher’s output and impact, showing how it varies within a year and evolves over time. The use of percentiles means that citation impact, which is highly skewed, can be seen in a context appropriate to both discipline and time since publication. Reducing this to the single value of an h-index may be an intriguing summary but it tells us nothing we can properly use in evaluation.

A Journal Profile Page, not just the JIF

The *Journal Impact Factor (JIF)* suffers from misapplication. It isn’t about research evaluation but about journal management. Putting *JIF* into a context that sets that single point value into a profile or spread of activity enables researchers and managers to see that *JIF* draws in a very wide diversity of performance at article level. *JIF* may be a guide but the full context is needed for real information outside the library and publishing house.

 *The old proverb says that a picture is worth a thousand words. Visualizing a data distribution is worth a thousand single-point metrics.*

An Impact Profile™, not an isolated CNCI

A summary index of the average Category Normalized Citation Impact (CNCI) can also be misleading, because it submerges a diverse data spread which, as at individual and journal level, is highly skewed and subject to outlier values. The Impact Profile™ shifts that skew into a more digestible form and reveals the underlying distribution. It shows that the spread around a world average and an institutional average means that many outputs are inevitably cited more and others less often. Whereas the summary value told us nothing more than X had a higher average than Y, the Impact Profile™ points up a whole series of questions, but also provides routes to answers for research management: where are the collaborative papers; do the same people produce both high and low cited material; did we shift across time?

A Research Footprint, not a university ranking

The ranking table of universities suppresses far more information than most analyses. The Research Footprint can unpack performance by discipline or by data type. It can compare two institutions or countries, or it can compare a series of target organizations to a suitable benchmark. Critically, it shows that there is no sensible way to compare two complex research systems with a single number: it’s a bit more complicated than that!

The old proverb says that a picture is worth a thousand words. Visualizing a data distribution is worth a thousand single-point metrics.

References

- Adams J, Gurney K A and Marshall S. (2007). Profiling citation impact: a new methodology. *Scientometrics*, **72**, 325-344
- Bornmann, L and Haunschild, R. (2018). Plots for visualizing paper impact and journal impact of single researchers in a single graph. *Scientometrics*, **115**, 385-394. DOI <https://doi.org/10.1007/s11192-018-2658-1>
- Bornmann, L and Marx, W. (2014). Distributions instead of single numbers: percentiles and beam plots for the assessment of single researchers. *Journal of the Association for Information Science and Technology* **65**, 206–208. DOI: 10.1002/asi
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *PNAS*, **102**, 16569–72.
- Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, **122**, 108-111.
- Garfield, E. (2006). The History and Meaning of the Journal Impact Factor. *Journal of the American Medical Association (JAMA)*, **293**: 90-93, January 2006.
- Garfield, E., & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, **14**, 195-201
- Waltman, L., & Van Eck, N. J. (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*, **63**(2), 406-415.
- Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, **94**, 851–872. doi: [10.1007/s11192-012-0775-9](https://doi.org/10.1007/s11192-012-0775-9)

Web of Science