Profiles, not metrics

Why it is important to drill into the data that feed any ‘single point’ metric

Jonathan Adams

Director, Institute for Scientific Information

April 30th, 2019

Web of Science
Trust the difference

Clarivate
Analytics

Profiles, not metrics

- information is lost when data about researchers and their institutions are squeezed into a simplified metric or league table
- ISI looked at four familiar types of analysis that can obscure real research when misused
- These analyses seek to describe individuals, journals, research units and whole universities
- I will describe four alternative visualisations that unpack the richer information that lies beneath each ‘headline’ indicator
- The visualisations may seem complex but they lead to additional questions about the data, which supports more responsible research management and more confident decision making

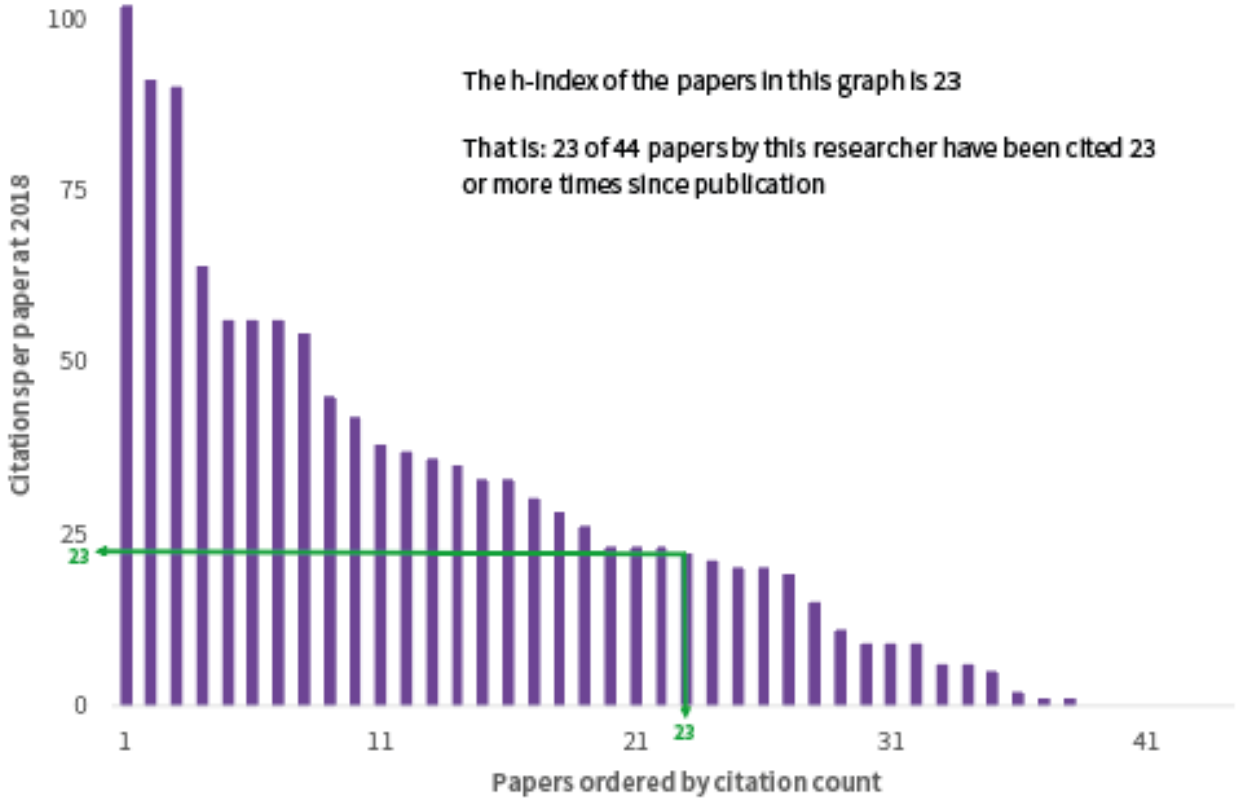
Example 1
The h-index

An h-index = 23 for a researcher who is an author or co-author on 44 citable journal articles over a 15-year period.

In this first example, an h-index = 23 for a researcher who is an author or co-author on 44 citable journal articles over a 15-year period.

Total output included reports and proceedings that cannot be analysed by a single h-index.

Graphing the journal data reveals the spread, skew, and presence of relatively highly-cited items buried under the ‘h’ value. Uncited items disappear.



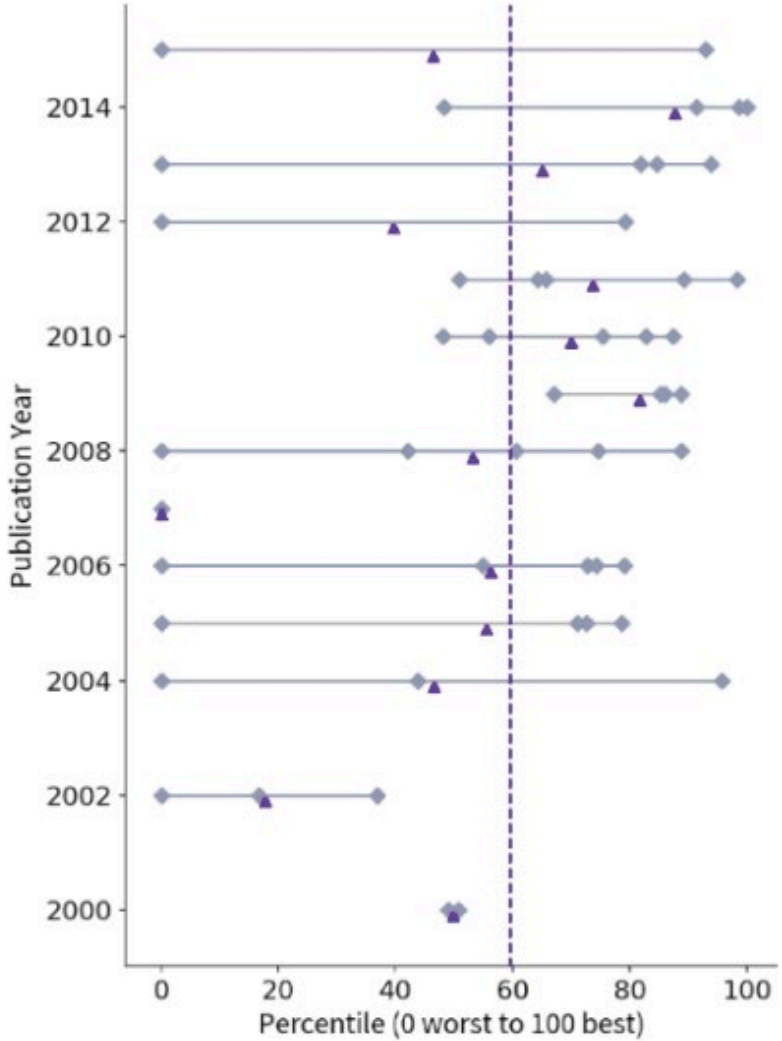
A beam-plot of the same data

Each article is compared to its own reference set by year and category

A percentile is calculated, so 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)

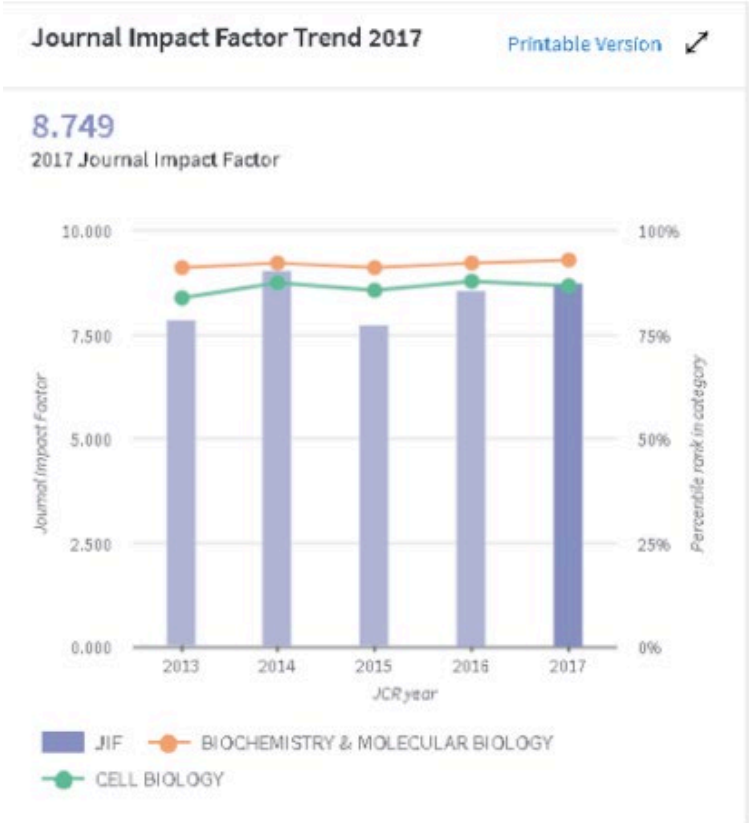
The dotted vertical benchmark is the researcher's overall average: the 59th percentile

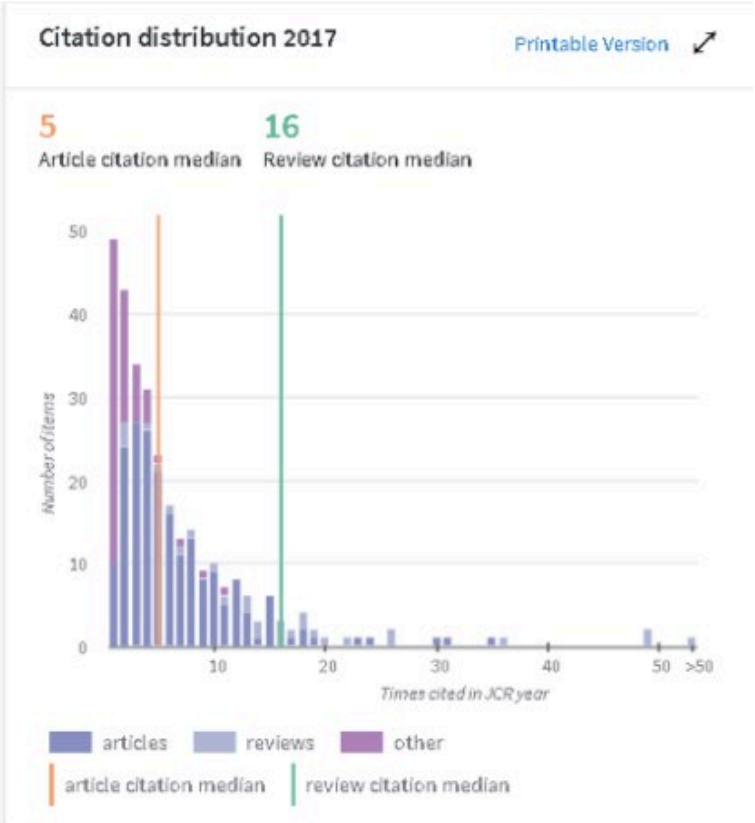


Example 2
Journal Impact Factor (data for EMBO Report**)**

JIF Trend 2017 shows JIF and percentile in category

Citation Distribution 2017 shows medians and overall spread: a more complete background





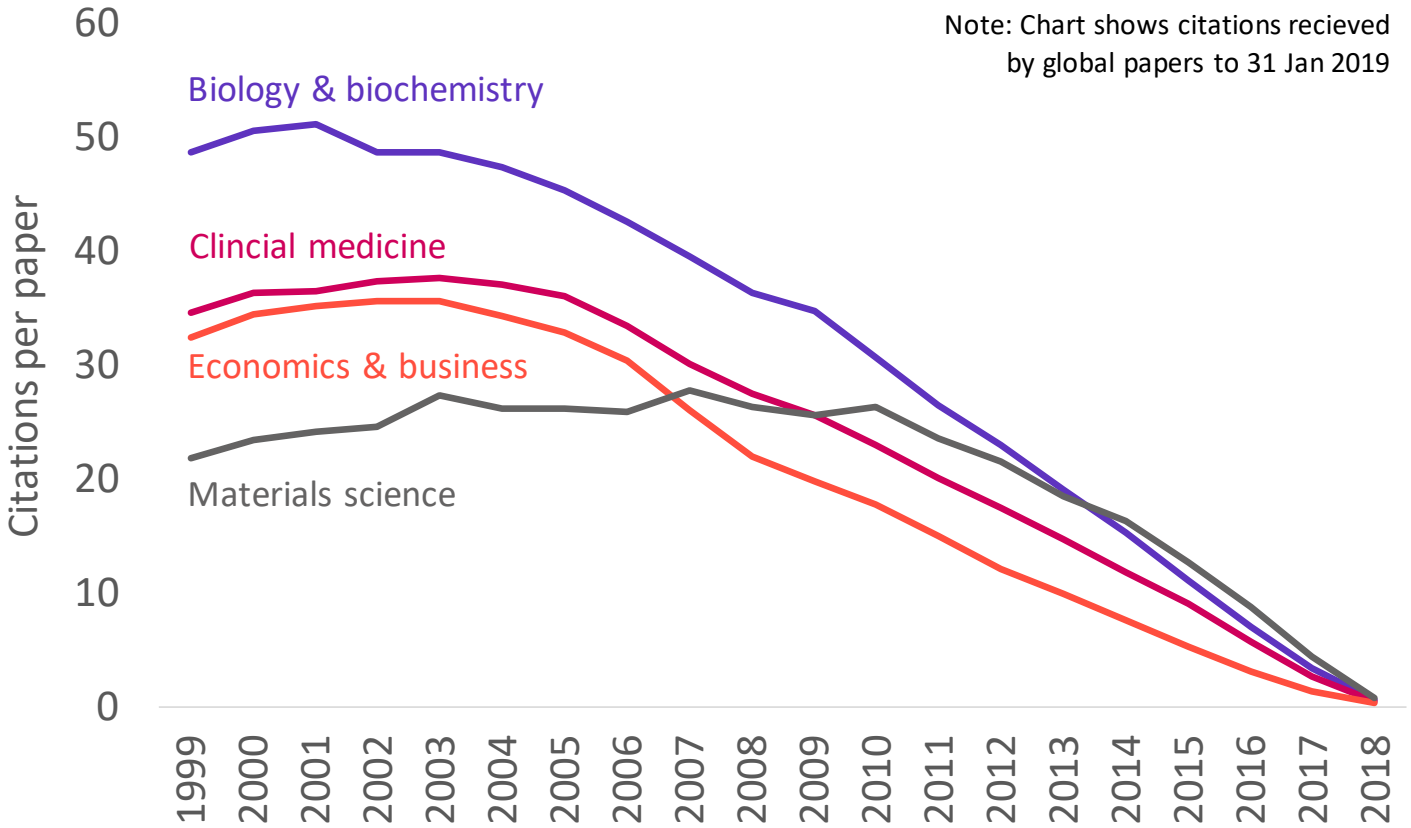
Web of Science Group

Clarivate Analytics

Example 3
Average normalised citation counts

Citation counts rise over time at a rate that is discipline dependent

The citation count for each paper must be 'normalised' before combining data to calculate an 'average' value

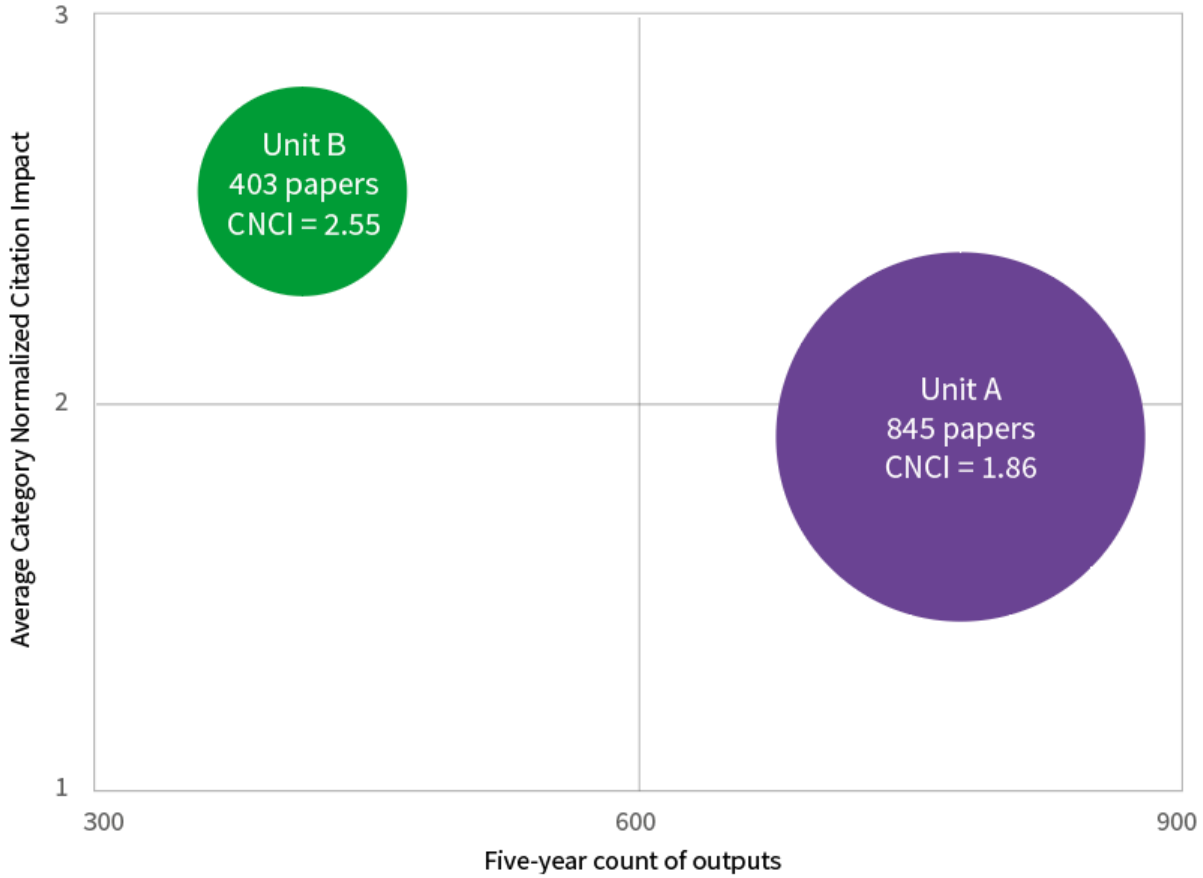


Average CNCI for two UK biomedical research units

The average Category-Normalised Citation Impact (CNCI - 'normalised' by the world average for that publication year and journal category) is shown

Disc size indicates relative five-year volume of output

Unit B has about half the output but a higher average CNCI than Unit A



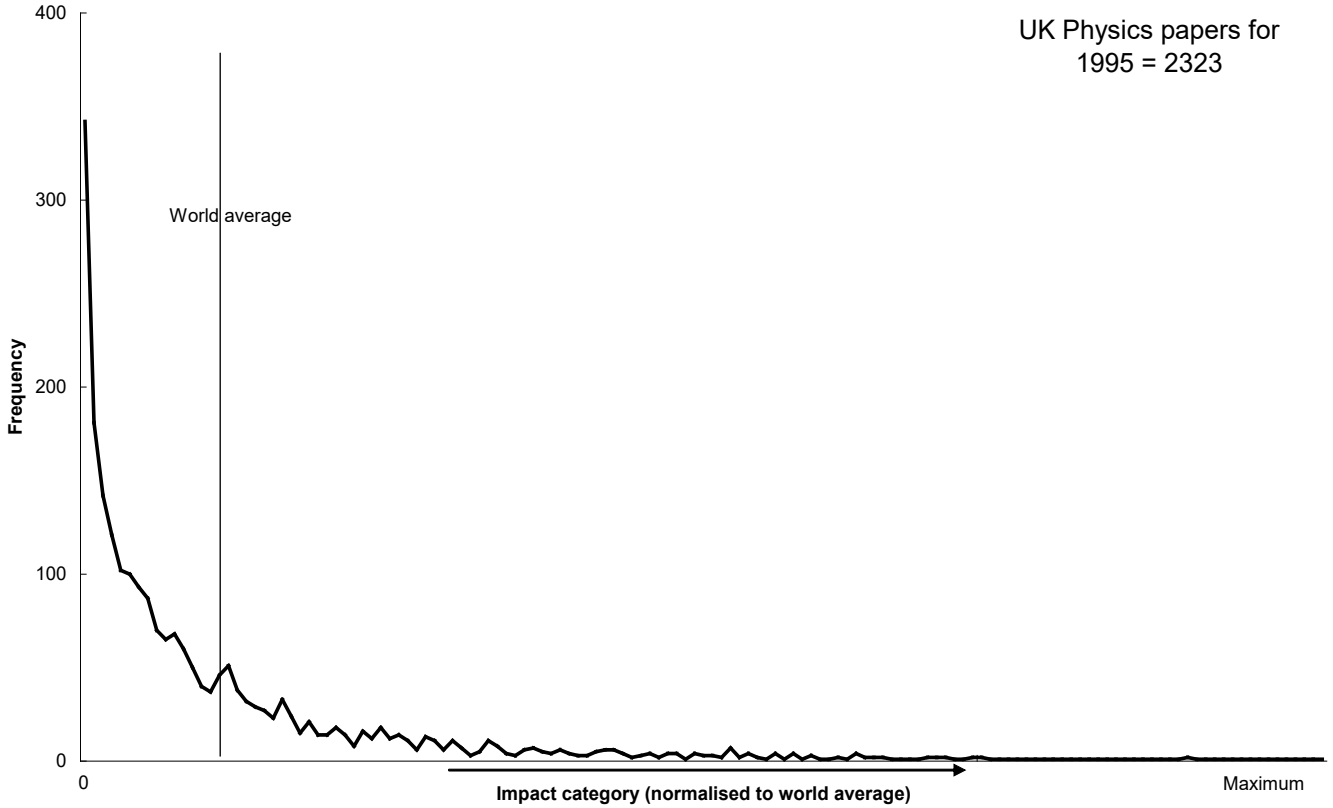
World average CNCI = 1.0

Research activity data are (very) skewed

The average CNCI value hides the underlying distribution

The easy assumption is that an average is a mid-point but in practice the average ‘impact’ will be greater than the median

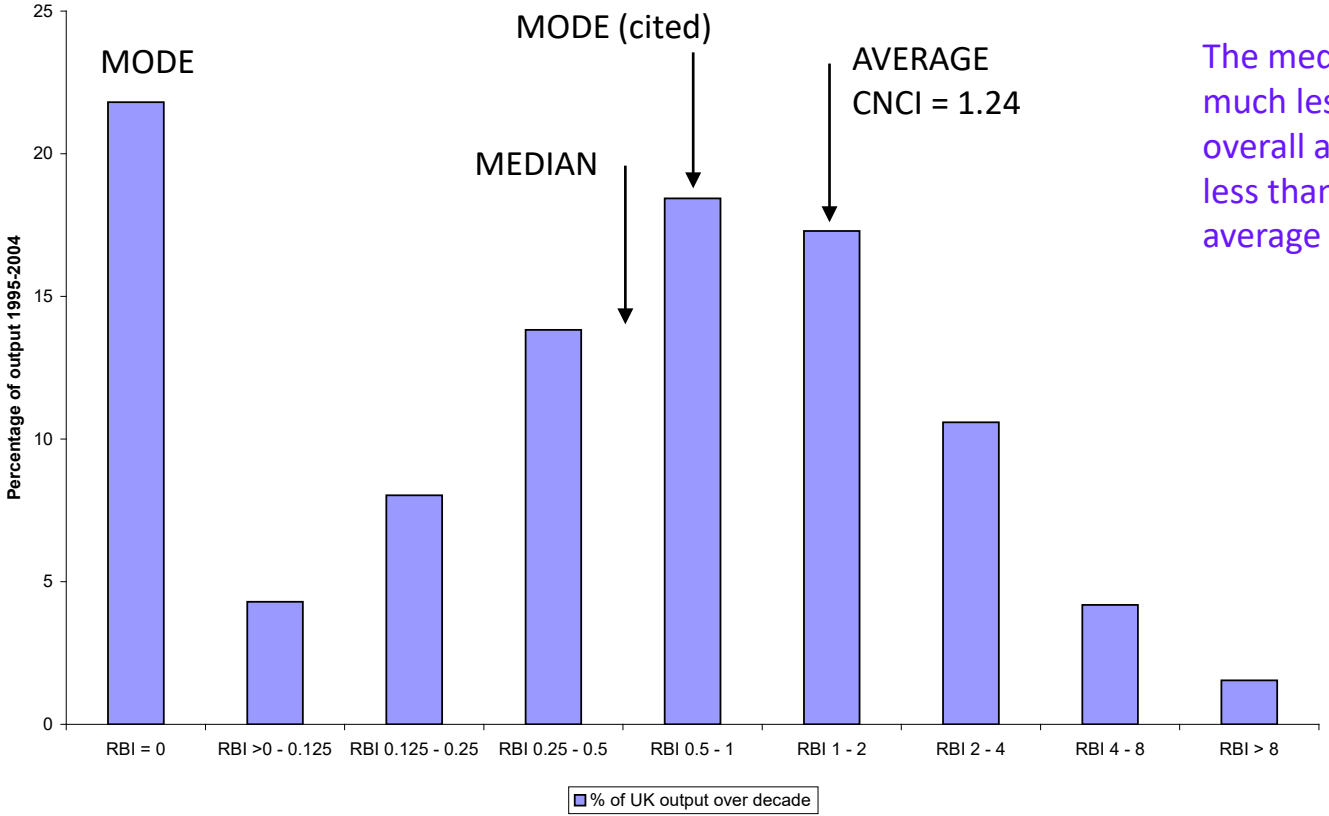
The pattern is also true of funding data, group size, etc



How can we better visualise the distribution of citation impact?

- Scale the data relative to a benchmark, e.g. world average
- Then categorise the values around that benchmark
- All journal articles
  - Uncited articles (to remove zero values)
  - Cited articles
    - Cited less than world average
    - Cited more than world average
      - Cited more than average but less than twice as often
      - Cited more than twice world average

These are UK data for ten years to 2006 (680,000 papers)



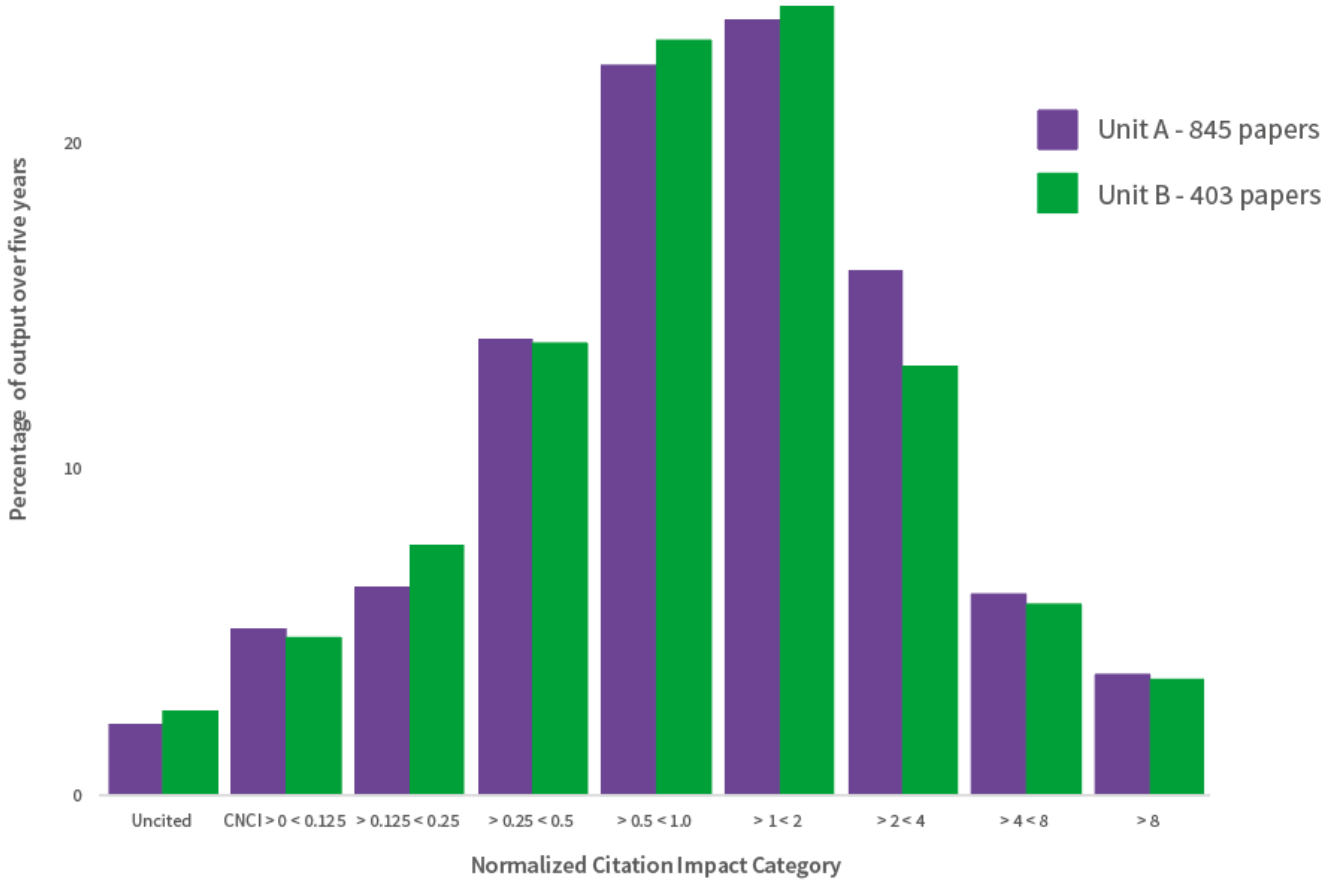
The median is not only much less than the overall average, it is less than world average

Impact Profile (5-year) of the two UK biomedical research units

CNCI of each paper is allocated to a series of bins grouped around the world average (= 1.0; uncited papers grouped to the left)

Counts are shown as percentage output for each unit

The units' Impact Profiles differ much less than their average values



Web of Science Group

Clarivate Analytics

Example 4 Rankings: the global league table position of universities ranked highest in Times Higher Education's World University Rankings (WUR) for 2018.

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

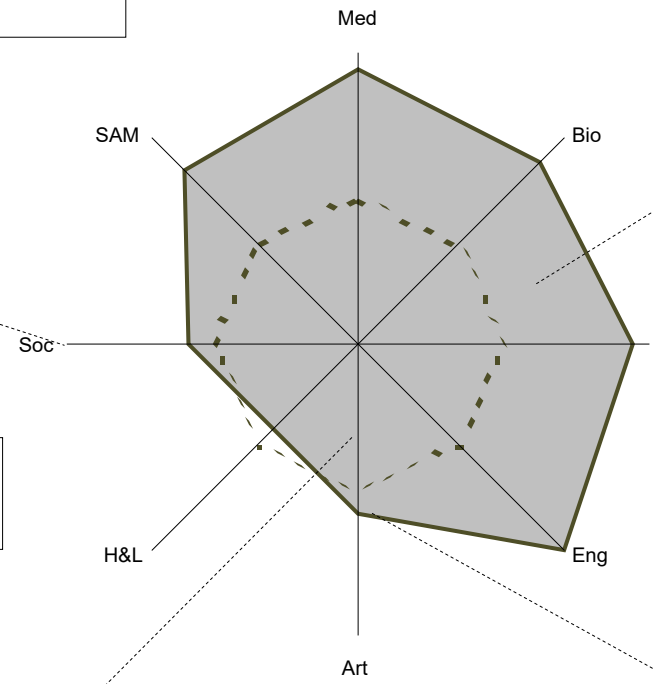
How can we unpack the data in the rankings?

- There are two main spectrums of activity and there are multiple axes for both
    - Discipline: chemistry, economics etc
    - Activity type: money, people, output etc

- A benchmark may also be informative, such as the average for an appropriate comparator group
- We want to display the spread of data for each activity type
- To address this we use Research Footprints: a radar diagram that visualises the institutional ‘footprint’ for a specified dataset on a standardised template

Publication impact

Title identifies research performance indicator covered by this footprint



Dotted line defines footprint of mid-ranked performance in this sector group

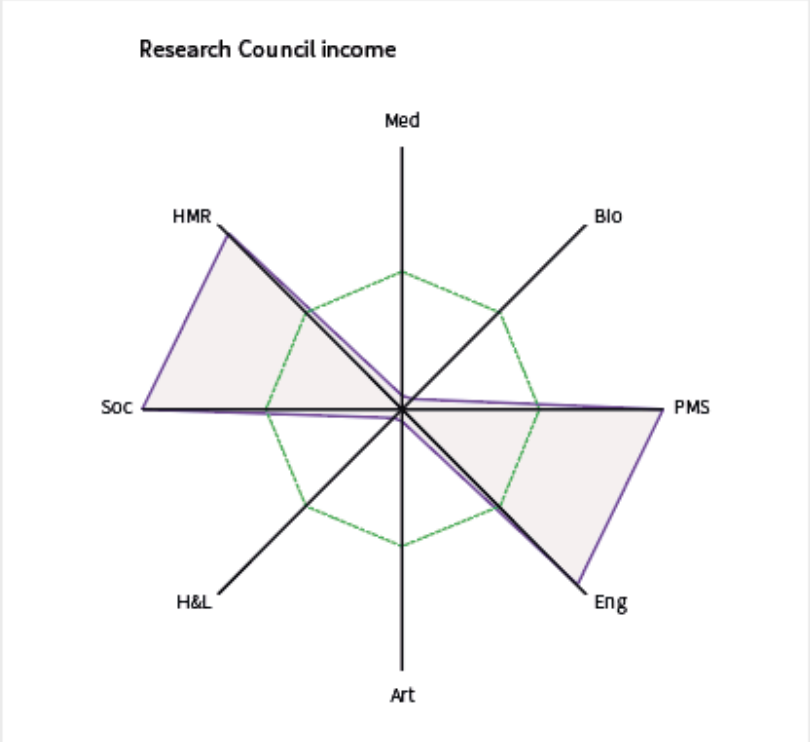
Text identifies subject area on this axis

Axes run from lowest rank to highest (first) within the sector and subject

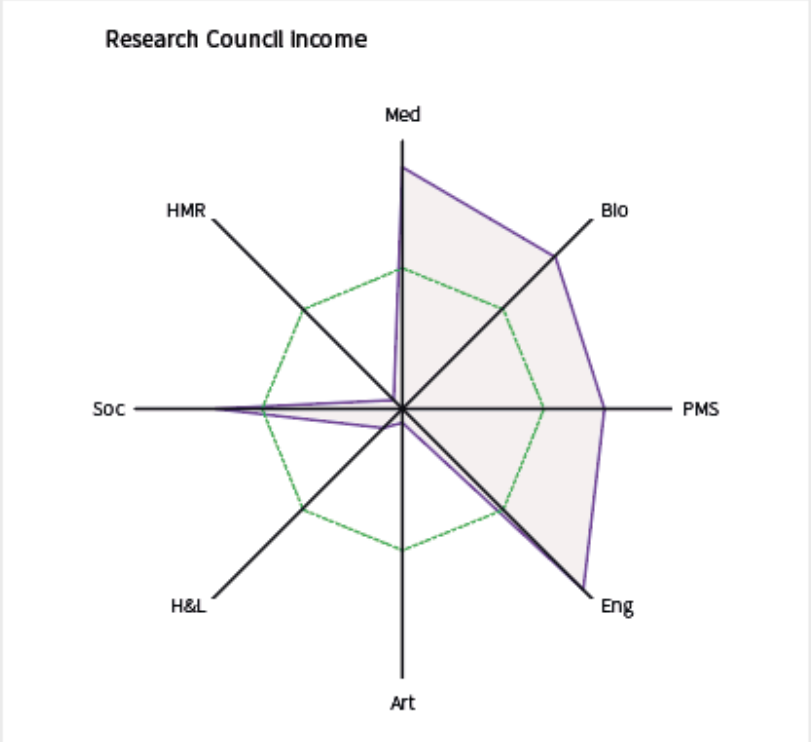
Continuous line defines footprint of institution in this profile

Distance along axis shows strength relative to maximum performance within sector group

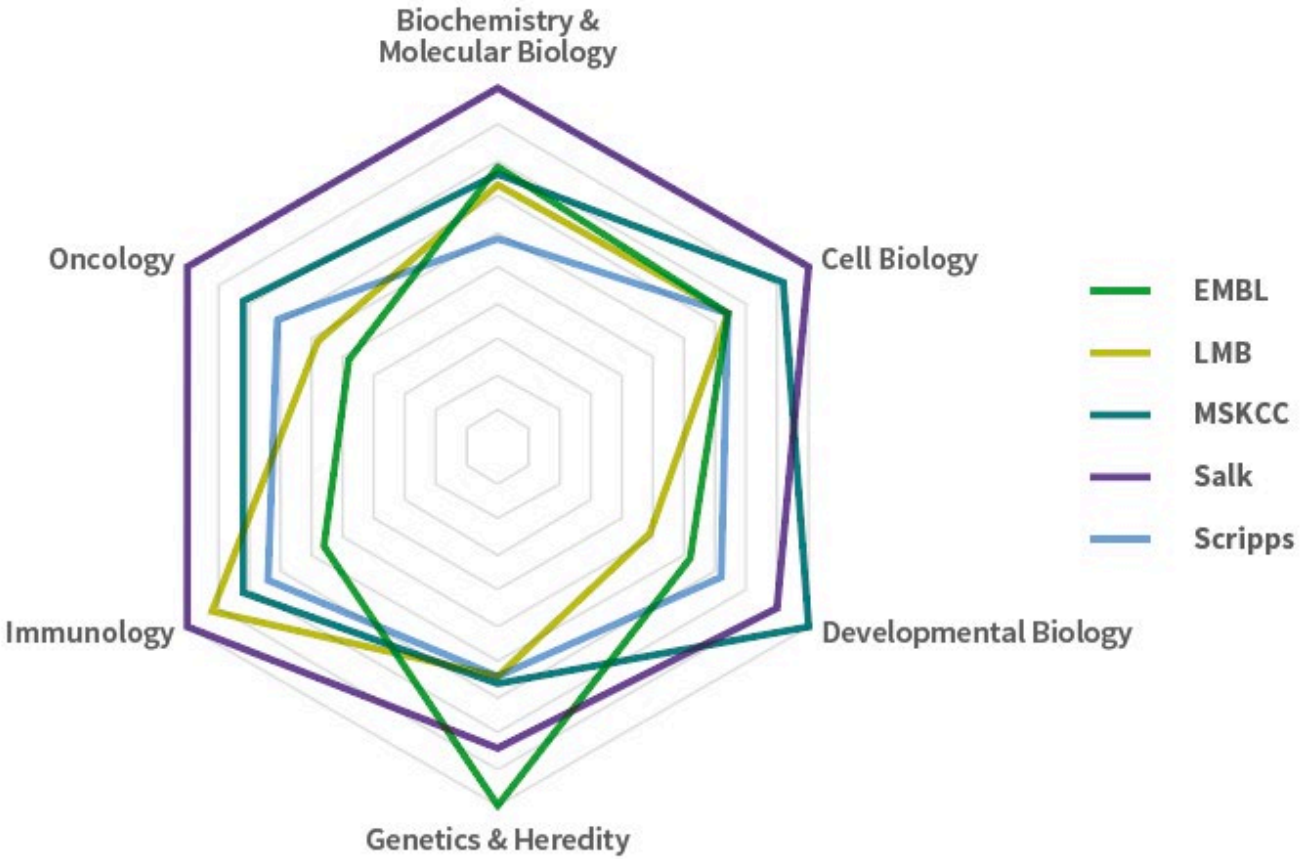
Imperial College, London



London School of Economics



A **Research Footprint** unpacks detail, which in this instance reveals significant differences at Faculty level



A **Research Footprint** can also be used for multiple comparisons

Take home message

- Information is lost when data about researchers and their institutions are squeezed into a simplified metric or league table

- Alternative visualisations can unpack the richer information that lies beneath each ‘headline’ indicator

- These visualisations may initially appear complex but
  - They stimulate additional questioning about the data
  - Which supports more responsible research management
  - And more confident decision making

Profiles, not metrics

QUESTION & ANSWER

Web of Science
Trust the difference

Clarivate
Analytics

Profiles, not metrics

https://clarivate.com/blog/news/institute-for-scientific-information-launches-global-research-report-profiles-not-metrics/