Bibliometrics Introduction

Research impact can be measured in many ways

Quantitatively using:

- Publication counts
- Amount of research income
- Number of PhD students
- Size of research group
- Number of PI projects
- Views and downloads of online outputs
- Number of patents and licenses obtained

Qualitative methods vary; the most important one is various forms of peer-review

Bibliometrics are ways of measuring patterns of authorship, publication, and the use of literature

Caveats / Cautions!

- Use of Bibliometrics and citation analysis is only one of these quantitative indicators
- The ability to apply it and its importance in the overall assessment of research varies between disciplines
- Attempts at quantitative measures can be contrasted with the main alternative assessment approach – qualitative peer-review in various forms
- The balance between use of Bibliometrics and peer-review in assessing academic performance at both the individual and unit levels is currently being played out locally, nationally and internationally

This section provides an introductory overview of the field – later sections look at: the key uses of Bibliometrics for journal ranking and individual assessment; the main metrics available; the main data sources and packaged toolkits available

Attraction of Bibliometrics

- Quantitative nature of the results
- Perceived efficiency advantage, as it is possible to produce a variety of statistics quite quickly in comparison to the resource – intensive nature of peer-review of quality and innovation of intellectual thought

Bibliometrics remain highly controversial as proxy indicators of the impact or quality of published researchers, even in those disciplines where citation analysis “works” in that much research output is fully indexed in the main citation data sources
The two other key areas where Bibliometrics are commonly used are:

- as evidence to support an individual in relation to consideration for promotion, tenure and grant funding
- in deciding where to publish research in to obtain maximum visibility and citation rate by targeting high impact titles

Despite its many shortcomings ranking tables for universities give considerable weighting to bibliometrics in their calculations

“....We publish in books and monographs, and in peer-reviewed journals. However, we have a range of real requirements that include official reporting to state agencies and authorities; public archaeology and communication in regional and local journals and in interdisciplinary publication across several journals, that most bibliometrics are incapable of measuring“ – [An academic]

The building blocks – a data source and a set of metrics

Source datasets

- The main source datasets – databases holding research and citations to it – are those of Thomson Reuters (Web of Science, Journal Citation Reports and other products), Elsevier (Scopus and other products) and Google Scholar plus subject-specialist options in some fields
- Each collects the citation information from the articles in a select range of publications only – the overlap between the content of these sources has been shown to be quite modest in particular studies. So using just one source is providing a partial view of both research and citations to it

Metric tools and techniques applied to the data source

Basic building blocks are a series of techniques such as h-index, Journal Impact Factor (JIF), eigenfactor, SJR, SNIP – these various formulae transform the raw data into quantitative evaluations of both journals and individual researchers.

Publication counts

Publication counts measure productivity but arguably not impact – as an example of this 28% of the 8,077 items of University College Dublin research from 1998-2007 indexed in the ISI Citation Indexes were not cited other than self-citations. Overall as much as 90% of the papers published in scientific journals are never cited. The SCImago web-based free product provides an easy graphical presentation of that non-cited material for each journal title.
Citation analysis

- Most current Bibliometric indicators are based around analysis of citations. The key concept is that the number of times you get cited is meaningful – the more citations the greater the relevance.

- There are three main approaches to citation analysis:
  - **Citation counts** – total number of citations, total number of citations over a period of time, total number of citations per paper
  - **More sophisticated use of citation counts such as** : **number of papers** cited more than x times; number of citations in the x most cited papers
  - **Normalisation and “crown indicators”** Citation counts alone are commonly used but this is really rather meaningless unless normalised by some combination of time, journal of publication, broad or narrow field of research. This benchmarking approach is the most commonly used at present. There are various initiatives to provide metrics sufficiently normalised on a number of criteria such that both journals and individuals can be compared across disciplines. Examples of these newer normalised metrics are SNIP for journals and the universal h-index for individual researchers.

“The terrible legacy of IF [Journal Impact Factor] is that it is being used to evaluate scientists, rather than journals, which has become of increasing concern to many of us. Judgment of individuals is, of course, best done by in-depth analysis by expert scholars in the subject area. But, some bureaucrats want a simple metric. My experience of being on international review committees is that more notice is taken of IF when they do not have the knowledge to evaluate the science independently” – [Alan Fersht “The most influential journals: Impact Factor and Eigenfactor” PNAS April 28, 2009 vol. 106 no.17]

Issues & Limitations

In some fields it is not the tradition to cite extensively the work that your scholarship and research is building upon – yet this is the whole principle of the citation analysis system.

- **Seminal research** is also often taken for granted and not cited.
- Where **citation is common**, the data sources often do not index the publications where research in a field is typically published – local publications, non-English, monographs, conference and working papers are poorly indexed.
- **Negative citations**, critical of a work, are counted as valid.
- **Manipulation of the system** by such means as self-citation, multiple authorship, splitting outputs into many articles and journals favouring highly cited review articles.
- **Defining the degree** of speciality at which to benchmark. An individual’s metric score may be high in relation to the broad discipline, but in fact low in relation to their particular sub-speciality’s citing pattern.
- **Inappropriate use of citation metrics**, such as using the Journal Impact Factor of a journal title to evaluate an individual researcher’s output, or comparing h-index across fields, ignoring the citation pattern variations found.