Annexes to the UCL Bibliometrics Policy

February 2020

These annexes provide guidance to support the use of the UCL bibliometrics policy, and should be read in the context of that document. They are expected to be regularly revised and updated, to allow us to take account of new developments in bibliometrics, new thinking around best practices, and the experience of implementing the policy at UCL.

Please contact bibliometrics@ucl.ac.uk with any suggestions or queries.

Contents

Annex 1 - General principles for the responsible use of bibliometric indicators

Whatever purpose you are using them for, it is always important to consider the inherent limitations in citation indicators - it is easy to produce misleading figures without intending to.

- **Disciplines**: The usual sources of bibliometric data are focused on journal articles. They may therefore be less useful in disciplines that are less reliant on journal publishing, such as the arts, humanities, social sciences, computing science and engineering. The disciplines that do concentrate on journal articles may still have significantly different publication and citation practices. Some indicators can be normalised to allow comparisons across disciplines, but this is only ever approximate, and may be challenging for multidisciplinary work. Different disciplines also have different approaches to collaboration and multiple authorship. But because authorship is the primary means to attribute credit and contribution to journal articles, it may be appropriate to consider discipline specific relationships between authorship and research contribution when assessing research or researchers among disciplines.

  Be alert to the mix of publications in your data; be careful when drawing comparisons between different disciplines; be aware that useful data will not always be available for all disciplines; and use normalised indicators where possible.

- **Publication type:** Review articles will generally get more citations than original research. It is generally recommended to avoid including them in citation-oriented bibliometric analyses. When calculating citation figures for a researcher, try to only consider their research papers, and filter out material such as editorials and book reviews, as these will skew the figures. For some research outputs, such as practice-based material or policy documents, it is unlikely that any citation analyses would be productive.

- **Different meanings of citation:** Citations do not always signify the same thing. A paper may be cited as the fundamental basis of a later work, or as a cursory reference in a long list of loosely related prior research; either of these will show up as a single citation in a database and be counted in the same way with the same perceived value. Negative citations also exist – a particularly contentious paper may be frequently cited in order to dismiss its findings, or merely to remark upon the controversy.

  It is not usually possible to avoid counting negative citations, and in some fields they may be more frequent than others. Be aware of their existence, and look critically at any strange outliers.

- **Number of authors:** Where a document has multiple authors, most databases and indexes will attribute a publication and associated citation data to all authors. This means a single publication will count once for each author regardless of whether they were the sole author or somewhere in a long list of authors; and once for each author's institution regardless of whether the institution contributed one or more authors. Many existing metrics do not take account of shared contributions, and this can give misleading results particularly in highly collaborative fields.

  Methods exist to weight contributions, such as fractional weighting or author ordering. However, weighting relies on a clear sense of what any individual has contributed, and this information is rarely available. Without a fixed standard, conventions vary substantially between fields or journals, and weighting for shared authorship must be interpreted cautiously.

- **Time since publication**: Citations continually accrue over time, so older papers will tend to have more citations than newer ones. Similarly, older researchers usually have more papers and have had more time to accrue citations to them, so an indicator like the h-index will be misleading for researchers at different stages of their career.

  Where possible, compare papers against those of a comparable age, use a fixed time period after publication, or estimate of citations per year.

- **Basic principles of good data and statistical analysis and interpretation:** When using and interpreting metrics, particularly those that result from calculations or aggregations, ensure you have sufficient knowledge of fundamental data analysis practice and basic statistics principles to undertake the desired work; and do not ask those that are not skilled in these areas to make decisions based on such metrics and statistics without clear interpretation guidance. Common issues include the following.

  Do not apply aggregate level metrics to individual subjects, and vice versa; such as incorrectly assess the quality of individual papers based on the impact factor of the publishing journal.

  Understand composite metrics expressed as single values (i.e. single value metrics based on calculation among other metrics), fail to capture the complexities, nuances, and multidimensional attributes of the underlying data. And still, they often miss or cannot encompass or account for all of the mediating factors for the characteristic being estimated, particularly at the individual level. This is why UCL recognizes the h-index is not valid - and metrics in general should be avoided - when assessing individual researchers, as they do not account for diversity in research (sub)disciplines and individually variable attributes (e.g., niche research questions, career stage, personal circumstances, and teaching/administrative commitments; see below for more on Author Background).

  Aggregated metrics, such as calculated means, are sensitive to sample sizes and give no information on variability. Hence, these can be unreliable if based on too few observations, as it's hard to distinguish whether the mean or aggregated metric represents a genuine indication or trend, or is the result of chance or subject to outliers with high leverage. At a minimum, the number of observations used to calculate an aggregate metric should accompany written or graphical representations of the metrics, as well as at least one estimate of dispersion, variation or distribution (e.g., standard deviation, error bars) for means.

- **Author background**: The personal background of the author can affect both their publication rates and their citation rates. Full-time researchers will tend to produce more papers than those on part-time contracts, those who have recently had career breaks (for example, through illness, or family or care commitments), or those with substantial non-research responsibilities (teaching, administration, etc).

  Individual factors (such as gender, ethnicity, age, and language) may also lead to existing large-scale systemic biases being reflected in (or even reinforced by) the citation indicators. For example, gender biases, where male researchers are cited more often than female (known as the Matilda Effect), are well documented. Further, a known "Matthew effect" – the rich get richer – demonstrates the more citations tend to be accrued by more prominent, well-established, researchers.

  These are hard to normalise for, and as such indicators based on individuals should be treated with caution.

- **Data sources**: Different citation databases will give different results, because they index a different range of material in different ways. Some have better coverage of monographs, reports, and conference proceedings than others; some omit specific journals. Some have better coverage of non-English language sources than others.

  Comparisons should always be based on the same data source, and where possible using data gathered at the same time. Always be cautious of benchmarking against citation indicators from an unclear source.

- **New and alternative metrics**: There has been a sharp growth in recent years in various commercial "altmetrics" services. These often use similar source data (eg number of tweets or download figures) but interpreted and presented in different ways. Depending on what indicators are used, they can show scholarly interest (eg Mendeley bookmarking), media interest (eg news stories), or public interest (eg social media activity). They can also be used to identify the use of research in policy documents or other official publications which may not appear in the conventional citation databases. Spikes in activity may come if a piece of work is particularly contentious, timely, or simply on a topic that catches the public

imagination. It is harder to gather standardised and comprehensive new metrics than it is traditional citation data.

In general, it is best to treat figures from these metrics as broad indicators – high activity tells us that there is something interesting there, but the details should be examined before drawing conclusions. They should never be used to quote a single numeric “score” for ranking a paper or author.

Annex 2 - Recommended sources for bibliometric data

UCL has access to a number of tools for providing bibliometric and citation data. They all have their own strengths and weaknesses. We would recommend that:

- For publication lists, **use RPS/IRIS if practical**
- For simple citation counts, **use data from Scopus, Web of Science, or InCites**
- For more complex indicators, **use data from InCites**

Lists of publications

The most authoritative list for a researcher's publications is one that they control and maintain themselves. The majority of UCL researchers have an actively maintained profile in RPS for publications from at least 2016 onwards. As such, we would recommend that unless you know it to be out of date, use the list from **RPS/IRIS**.

RPS data is fed into InCites through the MyOrganisation tool, so the lists in **MyOrganisation** for a specific person (or department) should be equally reliable, though with some additional limits.[1]

If this is not possible, **Scopus** and **Web of Science** both offer ways to search for papers by a specific author.[2] This is generally not quite as reliable, especially for authors with common names or working in multidisciplinary fields, so take care with the results and look out for omissions or false positives.

Citation counts

The two standard databases are **Web of Science and Scopus**. These will give reasonably consistent numbers of citations – on average Scopus tends to be slightly higher as it includes a slightly broader range of publications and better coverage of non-English material, but for pre-2000 citations, Web of Science has better coverage.

Citation counts in **InCites** are drawn from Web of Science and will be consistent with that source. However, it is a little awkward to find the number of citations for a single paper in InCites, as it is mainly intended for aggregate data, so it will usually be simpler just to use Web of Science directly.

A third commonly-used source is **Google Scholar**, which tends to return much higher citation counts than either Web of Science or Scopus. We would not recommend using it if the others are available, as these numbers can be misleading. Google Scholar indexes a large amount of non-scholarly material, and often contains duplicate records, giving inflated citation counts.

It is always important with citation counts to mention the source you recovered them from, as all sources have slightly different numbers. Try to avoid combining citation counts from different sources as this will inevitably be confusing for the reader.

Complex indicators

Web of Science, Scopus, and Google Scholar do not present any normalised or aggregated citation indicators beyond an h-index and total citation count per year, neither of which are very useful in most circumstances, and we would not recommend using these.

For all normalised metrics, use InCites. The key indicators we would recommend looking at for groups of papers are the category normalised citation index, share of papers in the top 10%/1%, and the average percentile, all of which are available through InCites but not other sources.

---

[1] At the time of writing, Dec 2019, MyOrganisation contains all RPS papers since January 2013, with the exception of those that are not indexed by Web of Science.

[2] For Scopus, see https://library-guides.ucl.ac.uk/scopus/author-search ; for Web of Science, see https://library-guides.ucl.ac.uk/web-of-science/author-search

Annex 3 - Common bibliometric indicators

Recommended metrics

While no bibliometric indicator is perfect, some are better and more robust than others. Where possible, we always recommend using normalised indicators – these take account of the differences in citation due to the subject area, the age of the paper, and the type of publication, to ensure that citations can be more meaningfully compared. This is a much better approach than simply counting total citations.

Three widely accepted normalised indicators are **category-normalised citations**, the **percentile rank** of papers, and the **share of papers in the top 1%/10% of citations**. All three can be obtained from InCites, and we would recommend using a mix of them to get an overall picture of citation impact – they all measure slightly different things.

If it is not practical to get any of these indicators, and you are confident that all publications are in the same subject area and otherwise comparable, then a quick method of normalisation is "citations per year" – but do be careful that the papers are all directly comparable, and that you are not, for example, comparing research articles to reviews or conference papers.

Category-normalised citations

Category-normalised citation impact (CNCI) calculates the average number of citations to all papers of the same type (articles, reviews, conference papers, etc), in the same publication year, and in the same subject area ("category"). It then divides the number of citations to the target article by this average.

Papers that are multi-disciplinary and assigned to more than one subject area are given a single value, which is an average of their CNCI for each category.

The key weakness of the CNCI is that the "category" is calculated by assuming all papers in a given journal belong to a specific subject area (or multiple different subject areas). This means that it may give confusing results for papers in journals that publish a wide range of topics; in multidisciplinary megajournals, such as *PLOS One* or *Scientific Reports*, it may not be meaningful at all.[3]

The CNCI is scaled so that the world average is "1", and a CNCI of "1.5" would thus mean that this paper had 1.5x the global average citation for comparable papers. You may find it more helpful to benchmark against the UK average (across all subject areas, approximately 1.4) or the UCL average (across all subject areas, approximately 1.9).[4]

Percentile rank

The percentile rank calculates where a paper sits, on the basis of its absolute citation count, compared to all others of the same type (articles, reviews, conference papers, etc), in the same publication year, and in the same subject area ("category"). A paper with zero citations would be in the hundredth percentile; an average paper would be in the fiftieth percentile, and so on.

As with CNCI, "category" is determined by the journal it was published in, and may give odd results for multidisciplinary research, or papers in multidisciplinary journals. Papers that are assigned to two or more categories are given the best (lowest) percentile rank rather than averaging them.

This is expressed as a whole number, so a value of 1 indicates that a paper is in the first percentile, or in the top 1%. A value of 1.1 indicates that it is just outside the top 1%, and a value of 0.9 indicates that it is just inside it. 0.1 indicates the top 0.1%, and so on.

---

[3] InCites attempts to reassign papers in some multidisciplinary journals, such as *Nature* and *Science*, to more appropriate categories based on the other papers they cite. However, this process is not perfect and does not work for all papers.

[4] All benchmark figures are calculated for research articles only, published 2013-2018, affiliated to any UK institution or just to UCL, using the November 2019 InCites data.

A value of 100 usually indicates no citations at all, while (due to rounding) an incredibly highly ranked paper may be shown with a value of zero.

The percentile rank applies to single papers; for a group of papers, this is shown as an "average percentile", calculated using a weighted average of all items.[5] Again, the lower the value the better. The way this average is calculated tends to dilute the effect of outliers, making it useful to see the overall impact of a set of papers.

As a benchmark, the world average percentile for all papers is around 55 (rather than 50, in part due to cross-disciplinary effects); the UK average is around 48, and the average for UCL is around 41.

Share of papers in top 1%/10%

For large groups of papers, a useful addition to the average percentile rank or CNCI is to identify the share of papers which are among the most highly cited in their field.

These measures can show the amount of work which is very highly rated more effectively than just using an overall average value. For example, the CNCI would look the same for a group of papers which are all quite well cited, or for a group which has many average items and a few really standout ones. Having (for example) 20% of papers in the top 10% band would indicate that papers reach this high citation level about twice as often as the world average.

InCites offers pre-calculated data for papers in the top 10% or top 1% of all comparable papers (again, normalised by category, publication year, and publication type). Where a paper is classed into more than one subject category, it is given the best percentile value available – so a paper which is top 1% in cell biology but outside the top 1% in oncology would be counted as in the top 1% overall.

The global average across all disciplines is (of course) about 10% of papers in the top 10%, and 1% in the top 1%, though the effects of averaging multi-disciplinary papers can mean the exact figure varies a bit by subject area. The UK average is around 16% of papers in the top 10%, and 2.3% of papers in the top 1%, while for UCL the average is around 21% in the top 10%, and 3.9% of papers in the top 1%.

Metrics to avoid

Some very commonly used metrics are potentially very misleading, and we do not recommend using them in most circumstances. The two most common are the h-index and the impact factor.

The h-index

All things being equal, people evaluating researchers and research groups and organisations tend to favour the production of a higher number of research outputs, because funders and employers get more research for their investment. But, of course, things aren't equal, and research outputs vary in their quality (at least in principle – determining that variation is a different challenge addressed elsewhere).

The h index is a simple metric which tries to balance the quantity of papers produced by an individual researcher with the quality of those papers. Notwithstanding the issues with using citations as a proxy for quality, this metric was designed to allow a meaningful comparison between a researcher who has written a small number of highly-cited papers, and one who has written a large number of lightly-cited papers.

The index is defined so that someone with an h-index of h has at least h papers, each of which has h citations. For example, someone with 50 papers, of which 20 have been cited 20 or more times, would have an h-index of 20. The calculation is unaffected by how many times those twenty papers were cited – they could be cited thousands of times and the index would remain the same.

[5] It uses the harmonic mean, a form of averaging which gives greater weight to lower values

The h-index is a popular measure, with many researchers knowing their own h-index (and how it compares to their colleagues!). It is easily calculated, intuitively comprehensible, and as a result it is often used to compare researchers, or as a benchmark of career stage.

However, while well meaning and an intuitive counter to producing copious amounts of suboptimal research, the h-index has its own deficiencies that make it an unreliable indicator. There are three significant problems with the h-index, none of which are easy to correct for, and all reasons why UCL's bibliometrics principles is firm against its use.

- It will never decrease unless papers are retracted. Because older papers continue to gain new citations, the h-index can potentially increase indefinitely, even after the researcher has stopped actively publishing. A high h-index tends to reflect a long career as much as it reflects quality of research.

- The h-index is predominantly driven by the oldest papers, which have had more time to accrue citations. Newer papers – which may be more relevant or more significant – have little to no effect on the totals. This also means that a researcher's past career can have a disproportionate effect on their current h-index – two equally productive researchers, one of whom took a career break ten years ago and one of whom did not, are likely to have quite different h-indices today as a result of the long-term effects of that break.

- Finally, because the average number of citations accrued by papers, as well as the number of papers published over a career, varies so dramatically between disciplines, it is not possible to use the h-index to compare researchers in different areas. There is no absolute way to calculate a "good" h-index.

Bearing in mind these issues, **the h-index, or any similar metric, should not be used to directly compare researchers**, and we recommend that it is used with great caution in other circumstances. There are simply too many complicating factors to be able to produce a single number to meaningfully describe someone's overall publication record.

The impact factor

What's wrong with the journal impact factor?
UCL's bibliometric policy places very few specific restrictions on metric use; instead it defers to the learned community's judgement under a general framework that guides flexible metric use among disciplines and purposes. However, one key aspect is eliminating the use of the journal impact factor (JIF) as an indicator for article quality.

To some researchers, this is initially controversial and challenging because the JIF has become engrained into many research cultures. But the JIF is an extraordinarily poorly used metric, which has led some commentators to suggest it may be the cornerstone of an unhealthy research culture with the potential to distort the scientific process. Misuse of the impact factor is perhaps the single largest force behind the drive for responsible metrics.

Because JIF is so commonly and fundamentally misused, and its misuse may damage the integrity of the research system, it is worth explaining why it is singled out in UCL's bibliometrics policy, and why discouraging its use is also one of the few shared principles among most external responsible metrics initiatives, such as DORA and the REF guidance.

So what's so wrong with one of the most ubiquitously used metrics?

1) **Aggregate does not equal individual.**
   JIF is an aggregate value calculated based on citations of the individual articles in a journal. But because citation counts are highly variable among articles, JIF cannot tell us anything about the quality of any given individual article.

2) **Skewed distribution**
   Some might argue that it is acceptable to use as a group mean, like JIF, as a rough indicator for citations of individual papers, because the citation counts of most papers will be around the mean, and on average variability will balance out. However, citation data of papers within a journal are almost always heavily skewed: most articles receive very few citations and a few articles are very highly cited. The mode, median, and mean for citation counts within a journal are usually significantly different values, and using the mean is unlikely to tell you anything about an individual paper. The JIF of most journals is dragged up by a few highly cited papers, and most articles receive relatively few citations regardless of where they are published.

3) **It's not calculated the way you think it is**
   The average number of citations per paper in a journal over two years sounds straightforward – except that's not *quite* how the JIF is calculated.

   Averages are usually calculated by dividing the sum of the values for a sample of observations (numerator) by the number of those observations (denominator). Note simple averages are usually symmetrical – which means the sum of the sample is based only on the observations counted in the denominator. But the JIF isn't calculated in this way.

   Instead, the numerator – sum of citations – is based on all of the citations received by items in a given journal. This includes articles and reviews, but also letters to the editor, comments, and other front matter that aren't primary research articles – even news and obituaries; while the denominator is based not on the number of cited documents, but only articles and reviews. Hence, a journal's impact factor is driven not only by their research articles, but inflated by the other accompanying material in the journal. While this doesn't always get very heavily cited, it does usually add some extra citations.

4) **JIF is not even objective.**
   All quantitative and logic issues aside, at least the numbers behind JIF are objective? Except again – they're not. What constitutes citable material - whether something is primary research, front matter or another category – is negotiated by journals and the indexer of journals. These definitions are subjective and open to manipulation.

Skewed citation distributions and subjective calculation are enough to abandon JIF for inferring article quality. But if you're still not convinced, here's more problems which make it inappropriate.

1) **Short temporal window**
   The period of calculation is citations over two years. Such a small temporal window can lead to sensitive and highly variable metrics. For example, in a crystallography journal, a single massively-cited paper caused the journal's impact factor to leap from around 2 to around 50, and then drop back again two years later.

   Short temporal window is also inappropriate for most research fields because it takes much longer for a piece of work to be read, synthesized, appreciated and the built upon by the research community. Even in fast paced fields. For example, it takes 8 years for a paper to reach half of its eventual citations in the biomedical sciences.

   A short window also favours quick comment or turnaround of subsequent studies which may not be desirable in situations where it takes many years to research and design new work to examine the findings of a study to cite.

2) **Poor predictive tool**
   Building on the skewed distributions point, JIF is a poor predictive tool for future citations of an individual article. If you publish in a high impact factor journal, chances are you will get about as many citations as you would by publishing anywhere else. Indeed the most likely outcome remains that your paper will receive much fewer citations than the JIF – because the JIF is much higher than the median article citations.

3) **Matthew effect**
   Even though most papers are poorly cited, following journals because they've got a high JIF could lead to Matthew effects – the rich get richer – even though the quality of research is unchanged.

4) **Not transparent, nor reproducible.**
   The negotiations for what defines the citable material (and thus used to calculate JIF) is not public knowledge, and you can't reconstruct JIF from the data available. So we really don't know what exactly goes into any given JIF, or how reliable the calculation is.

5) **Other issues of metrics misuse present or exacerbated in JIF**
   As with many other metrics, it is important to remember that they only measure one thing, citations, and that those are a loose proxy of the abstract characteristic we're usually interested in – i.e., quality. There is strong variation in citation rates (and thus impact factors) among different fields and document types, which would mean that review journals or those in biomedical fields will tend to have higher impact factors. And when the metrics are incentivised and become targets in their own right, they lose meaning and is vulnerable to gaming lead to gaming

How did we get here? And where do we go?

In the pre-digital era, the unit of distribution for science was the physical journal volume. Libraries needed to make decisions on which journals to purchase and retain, and so the JIF was developed with no intention of reflecting research quality – but rather research readership and use. A journal with a high impact factor likely had a large number of potential readers, and the journal was likely to be heavily used. This explains the numerator/denominator issue, because all citations in a journal is a better estimate of readership and journal use than only research articles or reviews. Further, bibliographic tools like Web of Science and Scopus did not exist, and citation counts per article were not accessible to most researchers.

In the absence of any more detailed information, an average number of citations over a defined period was a reasonable metric to determine the use of journals. However, the digital era has effectively unbundled the unit of research consumption – researchers can read and use a single article without any awareness of the articles within a journal.

Because the impact factor was for so long the *only* citation-based metric readily available, it became popular as a metric of quality – despite all the issues discussed above. But metrics are now easily attributed directly to the individual articles – we can count how many people are reading, downloading, and citing a journal article. This means that we no longer need to estimate the impact of papers when we can get that data directly, more informatively, and more accurately.

Further reading
https://www.natureindex.com/news-blog/whats-wrong-with-the-jif-in-five-graphs
https://arxiv.org/ftp/arxiv/papers/1801/1801.08992.pdf
https://www.sciencemag.org/news/2016/07/hate-journal-impact-factors-new-study-gives-you-one-more-reason
https://im2punt0.wordpress.com/2013/11/03/nine-reasons-why-impact-factors-fail-and-using-them-may-harm-science/
https://blogs.lse.ac.uk/impactofsocialsciences/2019/04/26/the-impact-of-the-journal-impact-factor-in-the-review-tenure-and-promotion-process/

List of frequently encountered indicators

Indicator	Advantages	Disadvantages	Further Information
Bibliometric measures			
Measures for individual research outputs			
Times cited: the number of citations received by a particular research output.	Indication of the scholarly impact of a piece of research on subsequent research	Recent works will not yet have received citations. Figures vary depending on source. Individual citations may be positive, negative, or ephemeral – a count does not consider context.	
Measures for groups of papers, e.g. authors, institutions, countries			
Average-times cited: The average (mean) number of citations to a particular author's work.	Indication of the impact of an author's work on subsequent research based on the average times their work is cited.	Figures may be skewed by one or a few particularly highly cited papers. It does not reflect the volume of an author's work. Authors with recent papers are at a disadvantage as new papers will not yet have received citations.	Information on average citations from InCites
h-index: The number of publications for an author that have at least that many citations. E.g. an h-index of 20 means that 20 papers by that author have 20 citations or more	Indication of an author's impact based on the citation rates of their own outputs. The figure is not skewed by a paper with a particularly high or low citation count.	Researchers with many older papers have the potential to achieve a higher h-index that a researcher who is currently more active and influential.	Hirsch, 2005 Colledge, 2017, p.97 Information on h-index from InCites
5 year h-index: h-index calculation	Gives an indication of current	Researchers in fields that are	

based only on citations and publications from the last 5 years.	academic impact of a researcher.	slower to generate citations are at a disadvantage.	
Category Normalised Citation Impact (CNCI) / Field Weighted Citation Impact (FWCI): a measure of the number of citations in relation to the expected citation rate for that subject category.	Unbiased in respect to age or subject focus. The mean value is 1.00, so the relative influence of the article(s) is clear without looking at comparative examples. Good for benchmarking at institutional levels.	Very highly cited papers can have an inflated influence on the CNCI / FWCI. Baseline values for the current year can fluctuate more than expected.	Information on CNCI from InCites Information on Field Weighted Citation Impact from Scopus. Colledge, 2017, p.101
Percentage of highly cited papers	The proportion of an author/group's papers which are among the top 5% by citation rate in the field for that year. (Top 1%, 10% sometimes also used)	The calculation can be complex and requires a clear way of defining "the field" to which they are to be compared. Can break down for small numbers of papers.	Todeschini & Baccini 2016, pp.300-7
Measures for journals			
Journal Impact Factor (JIF): A citation impact of a journal based on the number of citations to articles from the previous 2 years received in the JIF year in Web of Science Core Collection.	Gives an indication of the academic impact of a journal. Useful for researchers deciding to which journals to submit their publications.	JIF has been mis-used to represent the impact of an author or institution, and is explicitly discouraged by DORA. Journals cannot have a JIF unless they have been indexed in Web of Science Core Collection for at least 3 years.	Information on Journal Impact Factor from InCites Declaration on Research Assessment (DORA)
Journal immediacy index: The average number of times an article is cited in the year it is published	Indicates how quickly articles in a journal are cited. As it is a per-article average, larger journals do not have an advantage over smaller ones.	Frequently issued journals may be at an advantage because they will have more articles published early in the year. Based on Web of Science citation data only.	Information on Journal Immediacy Index from InCites
Eigenfactor Score: Based on the number of times articles from the journal published in the past five years have been cited in the JCR	Citations in highly cited journals are given greater value so this is a measure of the prestige citation impact. Self-	Based on Web of Science citation data only.	Information on Eigenfactor Score from InCites

year but incorporating weighting based on the value of the citation.	citations are also removed.		
Article Influence Score: a measure of the average influence of a journal's articles over the first five years after publication.	The mean Article Influence Score is 1.00, so the relative influence of journal is clear without looking comparative examples.	Based on Web of Science citation data only.	Information on Article Influence Score from InCites
SCImago Journal Rank (SJR): A prestige citation impact of a journal	A metric based on prestige, i.e. citations are weighted depending on the prestige of their source.	Based on Scopus citation data only.	Information on SJR from Scopus González-Pereira, Guerrero-Bote, & Moya-Anegón, 2010
Source Normalized Impact per Paper (SNIP): The citation impact of a scientific journal using a source normalized approach	Allows for differences in citation rates in different fields.	Based only on citations received in Scopus.	Information on SNIP from Scopus Moed, 2010 Waltman, van Eck, van Leeuwen, & Visser, 2013
CiteScore: Based on the calculation for the JIF but using citations to articles in the previous 3 years in Scopus.	Scopus includes more journals than Web of Science Core Collection so some journals may have a CiteScore where they do not have a JIF.	Journals cannot have a CiteScore unless they have been indexed in Scopus for at least 4 years.	Information on CiteScore metrics from Scopus
New and alternative metrics			
Includes measures of: Usage: Views, downloads Captures: added to citation management packages, saved within database profiles Attention / social media: Facebook, twitter, blogs Mentions / dissemination / engagement: blog	Provides evidence of impact beyond citation / academic impact. Metrics are timely, appearing rapidly and before traditional citation metrics. They apply to more than just journal publications and books. Indicators can be obtained for data sets, software, presentations and	These metrics are relatively new, more research into their use is needed. They can only be used as a measure for relatively recent research outputs.	Information from altmetric.com Information from Plum Analytics Wilsdon et al., 2017

posts, news, Wikipedia entries Influence / impact: referenced in public policy documents, clinical guidance, commentary from experts or practitioners	other scholarly outputs.		

Bibliography

An overview of all of these metrics, and many more specialised ones, can be found in Todeschini & Baccini (2016).

Colledge, L. (2017). Snowball metrics: recipe book. [3rd edition] Retrieved from: https://www.snowballmetrics.com/wp-content/uploads/0211-Snowball-Metrics-Recipe-Book-v7-LO-1.pdf

González-Pereira, B., Guerrero-Bote, V. P., & Moya-Anegón, F. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. Journal of Informetrics, 4(3), 379-391. doi:10.1016/j.joi.2010.03.002

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102(46), 16569-16572. doi:10.1073/pnas.0507655102

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. Journal of Informetrics, 4(3), 265-277. doi:10.1016/j.joi.2010.01.002

Todeschini, R., & Baccini, A. (2016). Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research. doi:10.1002/9783527681969

Waltman, L., van Eck, N. J., van Leeuwen, T. N., & Visser, M. S. (2013). Some modifications to the SNIP journal impact indicator. Journal of Informetrics, 7(2), 272-285. doi:10.1016/j.joi.2012.11.011

Wilsdon, J., Bar-Ilan, J., Frodeman, R., Lex, E., Peters, I., & Wouters, P. (2017). Next-generation metrics: Responsible metrics and evaluation for open science. Retrieved from Luxembourg: https://ec.europa.eu/research/openscience/pdf/report.pdf

Annex 4 - Case studies of when and how to use bibliometric indicators

Use basic principles of good data and statistical analysis and interpretation

Expressing a desired property as a single value has a number of benefits, including: quick and easy readability and comparisons; straightforward subsequent calculations and processing (such as totals and averages); and are easy to plot and understand in graphs and figures. For raw data and metrics, such single value metrics are in themselves not inappropriate because they do give the exact values for the defined metric in use; although, these must still be used and interpreted responsibly because they are often used to represent abstract concepts or have different meanings in different contexts (see points 1, 3 and 4 of the policy). However, single value metrics can also commonly be the result of calculating mathematical equations or functions, and assumptions made transitioning among raw, calculated, or aggregated metrics can lead to unintended misuse and misinterpretation. However, in transitioning from many to one number, we lose important information contained in all of the values, and so limits our capacity to understand the nuances and completeness of the metrics. Hence, when responsibly using metrics that result from calculations, we must scrutinise the processes behind their calculation: what data was used, what data is missing; and interpret the metrics under these conditions.

Some very common aggregated metrics include familiar statistics, such as the mean or average. For example, the mean is comprised of all the values in a list of observations divided by the sum of the number of observations. Here, the mean is given as a single value, but in fact represents several values. Again, calculated single value metrics such as the mean have their own meaning and advantages for simplicity and ease of use, and are not inherently problematic. Depending on the question such metrics are entirely appropriate, yet they are still vulnerable to irresponsible use and interpretation. For example, means that are based on small number of observations (i.e., sample size) tend to be more variable and interpretations less reliable. Therefore, it is important to provide or request samples sizes, so one can make an informed decision about whether a mean, for example, is representative of a consistent pattern or the result of a few, possibly by chance, outliers.

Further, when interpreting such aggregated metrics, we must ensure that metrics are applied and interpreted at the correct scale of the subject of investigation. That is, do not apply aggregate level metrics to individual subjects, and vice versa; this is a fundamental logic error termed the ecological fallacy or fallacy of division. In bibliometrics and research evaluations, the most prominent example of this is incorrectly assessing the quality of individual papers based on the impact factor of the journal in which they were published. Here, variation in the number of citations per article within a journal means that assuming the average represents an individual paper would usually misestimate the actual citations of an article. Overestimates are particularly likely because the distribution of citations per article within a journal are usually highly skewed: a few papers extremely well cited, but most with few or no citations. This is why UCL is committed to not using the journal impact factor to assess individual articles.

Often many single value metrics are developed as attempts to represent complex and multidimensional attributes of research or researchers. However, such single metrics tend to fail to represent the desired property because they cannot capture the context-dependencies and multidimensional characteristics of the underlying data, and do not account for variation at the individual level or the circumstances that mediate the value. For example, the h-index attempts to combine publication number with citations to generate a single value indicator of researcher productivity and impact. But (issues of what constitutes “productivity” and “impact” notwithstanding; see policy point 1) the h-index does not provide independent measures of either original metric which themselves are more appropriate in contexts where a more complete understanding of both is required.

Further, the h-index - and metrics in general - should not be used to assess individuals, because they do not account for individual variation due to diversity in research (sub) disciplines and individually variable attributes (e.g., niche research questions, career stage, personal circumstances, and teaching/administrative commitments; see below for more on Author Background). Hence, when assessing researchers for promotion, for example, it is invalid to judge individual researchers with metrics such as the h-index.

At times, considered and appropriately chosen suites of metrics and indicators may better reflect complexity, diversity and comprehensiveness of a subject of investigation. For example, whenever a mean is calculated and quoted, it should be accompanied by at least the number of observations used to calculate it and an estimate of dispersion, variation or distribution of the observations around the mean (such as standard error). In a bibliometric example, total number of citations for a person or organisation is meaningless in itself in most contexts but adding how many publications generated those citations and how many researchers were involved provides a set of metrics with the key information needed to interpret the original citation metric.

Using bibliometrics in research evaluations

Evaluating and comparing research is an implicit and important component of the contemporary research landscape needed to inform decisions for publication, planning, resource allocation, funding, hiring, and awards. Evaluations can occur both within UCL and as part of external exercises, and at various scales from individual researchers to the research group, department, institution and national levels. Metrics can play a role in evaluating research in each of these contexts. Although, bibliometrics analyses are usually indicative and incomplete proxies for research quality, and considered subordinate to expert judgement in assessments.

Metrics in evaluations can be grouped as: 1) Formal - metrics used by evaluators as part of official assessment criteria or supplementary information; 2) Exploratory - metrics used by those being evaluated as part of their own internal process to find research outputs. 3) Support - metrics to provide quantitative evidence in support of outcome statements.

Formal. Research evaluation exercises (such as REF or other formal research assessments) typically have set criteria that will be used by both those being evaluated and as instructions to evaluators for how to judge research. Depending on the scheme, this can include whether metrics will be part of the criteria or provided as supplementary data to inform formal criteria, what metrics will be used (or explicitly excluded), and any weighting/calculations towards overall assessment.

When UCL personnel are investigating research for their own purposes, as evaluators, they should carefully consider whether metrics are useful for their specific evaluation purpose; scrutinise any particular metric; ensure no inherent biases disadvantage any subset of research or researchers being assessed; and provide transparent guidance to those being evaluated and reviewers on how bibliometrics will be used and interpreted in the assessment process. These must be considered under the guidance provided in Sections 1-11 of the Bibliometrics policy proper; and details provided in the cases studies of this Annex.

When UCL researchers and research bodies are being assessed, they should carefully read and follow the criteria and guidance on any metrics use set by evaluators and be aware these may not match UCL's bibliometrics policy when assessments are conducted externally. Depending on the scheme, inappropriate use of metrics may be disregarded or make a submission ineligible. Always seek the most up to date rules and guidance provided by evaluators, and support for interpreting such rules can be sought through UCL Library Services or OVPR Research Evaluations and Metrics teams.

Exploratory: Some metrics are known to correlate with qualitative assessment results. Hence, in addition to any metrics provided in formal contexts, bibliometric analyses may be useful in some evaluation contexts to help identify potential research outputs for submissions or to be highlighted in a research portfolio, particularly when faced with a large number of outputs where qualitative analysis of individual papers is inefficient or unfeasible. However, the presence and strength of relationships between metrics and quality vary among discipline and metric. For example, an HEFCE analysis of the relationships between 15 metrics and reviewer scores for individual outputs in REF2014 showed few correlations between metrics and REF scores in Panel D; but stronger relationships were more common in Panels A and B.[6] Although, even the strongest correlation did not explain all of the variance in research scores; and so, it is important to reiterate that metrics remain indirect and incomplete proxies of quality.

Both Scopus and Web of Science can be used to identify papers with high bibliometric scores (such as citations), and can be filtered to only show UCL authors, with papers in a particular discipline, and from a particular period. Publications listed by any indicator should be treated judiciously. For example, the expected level of citation varies with the age and subject of a paper. If the indicator used is not weighted by age, then care should be taken to only compare it to those of the same age. For very recent papers, those published in the current or previous calendar year, even weighted metrics can potentially be misleading – the small number of citations involved over a short period can make indicators very noisy and volatile. The indicators used should reflect the quality of the paper; in contrast to those generated at the journal level, such as the Journal Impact Factor, that should not be applied to individual papers or used to assess their quality.

Such methods have limitations and may miss high quality research that performs poorly in indicators (such as multidisciplinary work, or that not published in English). Hence, bibliometrics should not be the only avenue employed when investigating research qualities. Likewise, a work which scores poorly on a particular metric should not be automatically rejected on that basis alone, as there are likely other reasons to consider its value. Alternatives methods can include giving researchers the opportunity to provide what is, in their opinion, their best papers; and opportunities to provide work that may not be captured by bibliometrics providers.

Last, indicators do not always tell us why a paper was highly or poorly cited. Some papers might perform well on an indicator for reasons other than quality (e.g., a specific paper may be repeatedly referenced because it is profoundly erroneous and misleading). Hence, any list of publications generated by bibliometrics for evaluations must be further investigated qualitatively.

Support. For some evaluations**,** it is useful to quantify other elements to justify the quality, engagement, or impact of a piece of research. For example, the number of visitors to an exhibition informed by UCL research; or the number of PhD students to complete their degree within a department or lab group.

Some of these metrics (particularly those connected with engagement and dissemination) fall under the altmetrics group of indicators. Broader guidance on using these is available in Annex 3 of the policy, and evaluations schemes may have specific guidance on what types of altmetrics can be used; for example, REF2021 has draft guidance (currently out for consultation) for impact indicators.[7] It is important to consider quantifying these with the same diligence as more traditional bibliometrics, such as checking the rules on which indicators provide supplementary information.

---

[6] *HEFCE (2015). The Metric Tide: Correlation analysis of REF2014 scores and metrics (Supplementary Report II to the Independent Review of the Role of Metrics in Research Assessment and Management). HEFCE. DOI: 10.13140/RG.2.1.3362.4162*

[7] *http://www.ref.ac.uk/media/ref,2021/downloads/Consultation%20on%20the%20draft%20panel%20criteria%20and%20working%20methods%20REF%202018_02.pdf*

For evaluators developing guidance for the use of these metrics in their own assessment processes, the FFRM has developed a list of useful quantifiers.[8] While this advice was specially designed for REF2021, the principles and metrics discussed by the FFRM could usefully be applied to other forms of evaluations.

Guidance for REF2021

In REF2021, citation metrics will be provided alongside research outputs to reviewers in some Units of Assessment to provide supplementary information on the originality, significance, and rigour of outputs to support expert review. While the other UOAs will not use citation metrics during assessment, there are some circumstances where metrics may help support UCL's internal REF processes across these other UOA's.

With this in mind, the there are some complexities regarding the REF rules, the utility of research metrics under REF is also variable, and UCL has commitments to metric use set out in our REF code of practice. But specific REF2021 guidance is beyond the scope of these general bibliometric guidance annexes.

And so, for those that wish to use metrics to help inform their decisions, UCL's bibliometrics and REF personnel have put together some guidance on research metrics for UCL's internal submission processes. Please contact the REF team for further information.

Additional material:

UCL REF Code of Practice

https://www.ucl.ac.uk/research/evaluation/research-excellence-framework/ref-2021/governance

Metrics Tide including analysis of REF2014 citation data

Final Metrics Tide report (contains more information on citations, metrics and their use in research evaluations)

https://re.ukri.org/documents/hefce-documents/metric-tide-2015-pdf/

Supplement I (literature review information final report)

https://re.ukri.org/documents/hefce-documents/metric-tide-lit-review-1/

Supplement II (contains analyses of relationships between REF2014 scores and some metrics)

https://re.ukri.org/documents/hefce-documents/metric-tide-lit-review-2/

REF Guidance and Criteria

Citation and metrics mentioned throughout

https://www.ref.ac.uk/media/1084/ref-2019_02-panel-criteria-and-working-methods.pdf

Research England and Clarivate pages on interpretation of citation data

Research England and Clarivate have made available a set of reference tables that provide context around citation data. A full list of the guidance to the contextual data can be found here:

https://www.ref.ac.uk/guidance/citation-and-contextual-data-guidance/

[8] *https://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/The%20Forum%20for%20Responsible%20Research%20Metrics/FFRRM%20advice%20to%20REF2021%20panels%20on%20impact%20and%20environment%20indicators%2026July2018.pdf*

Using metrics in recruitment, assessment, and promotion of individuals

When using metrics in the contexts of recruitment, assessment, or promotion, great care should be taken to ensure that any metrics used are appropriate – in many cases, it will not be appropriate to use any. In *some* circumstances, bibliometric indicators can be useful to inform aspects of the academic significance and contribution a researcher has made to a field (relative to career stage and opportunity), as one of many contributing factors to the overall profile of that researcher. Bibliometrics should never be used apart from qualitative assessment, and should be avoided for fields where they are known to be inappropriate (e.g., many fields in the arts and humanities). If in doubt, err on the side of caution, and avoid indicators that are not widely used and understood. Any use of quantitative indicators should be done in accordance with existing UCL policies and guidelines, such as the Academic Careers Framework.[9]

The most commonly used metric for looking at the productivity of a single researcher is the **h-index**. This looks at the publications and citations they have achieved across their entire career, aiming to balance productivity and impact. It is perhaps the most widely cited metric after the impact factor. However, it has substantial limitations and should not be used for the assessment or ranking of individuals. As a total-career metric, it strongly favours established researchers; indeed, it is possible for the h-index to continue to increase even after retirement. The use of non-normalised citations makes it strongly dependent on the field. It is also likely to emphasise existing career-opportunity biases, penalising (for example) researchers who have taken career breaks.

A number of similar metrics are available, such as the **5-year h-index**, which only considers papers published in the last five years. These avoid some of the specific problems with the h-index, but should still be avoided in most cases.

Journal-level metrics (eg the **impact factor** or publisher) should never be used as a proxy for the quality of individual papers in those journals, or by extension the impact of their authors.

Metrics for individual documents can be combined to generate indicators that summarise a researcher's entire output, which may be useful to compare researchers and could be included on a CV. When assessing researchers, such metrics are probably only useful when comparing like-for-like, and attempts should be made to take into consideration career stage and other factors that may affect the metrics. One approach is to consider the proportion of their papers which are among the most highly-cited papers in their field – usually the benchmark is the top **5% or 10% of papers published in the field** that year. This figure will (to some degree) be comparable across different disciplines. It is also possible to consider the average **category normalised citation impact** of the papers, which is normalised by field.

However, for any analysis based on small groups of papers, remember that the small number of publications involved can make statistics over a short period (1-2 years) unreliable, and that apparent differences between groups can often come down to the effects of counting a single paper.

For considering the impact of a small number of recent papers, which have not yet accrued citations, it can be useful to use an **alternative metrics** approach (this data is available on a per-paper basis through RPS, and can also be found in Scopus). Mendeley reader counts usually have a strong correlation to later citations, though this is usually only indicative information and cannot be easily quantified. It may also be valuable for fields where a strong level of policy/practitioner impact is expected, such as clinical medicine.

The use of metrics can be complemented and strengthened by using a short **career narrative**. Context and content are vitally important, and indicators should be qualified with text. For example for hiring and promotions, a possible approach could be to ask researchers to provide such statements and perhaps name 3-5 research outputs and why those represent their best work. For researchers, it may be beneficial to provide such statements even when not prompted. This gives applicants the opportunity to demonstrate why their research is uniquely important, beyond paper-level metrics.

[9] https://www.ucl.ac.uk/human-resources/policies/2018/jun/academic-career-framework

The same general principles apply when advertising for a post – consider carefully what metrics you are asking for, if any. Do not, for example, require "papers in high-impact journals" as a condition for shortlisting, or ask applicants to state an h-index.

Using metrics to assess research groups or departments

On a small scale, metrics for research teams or ad-hoc groups are best assembled by aggregating the papers produced by their members. On a larger scale, such as departments or faculties, the new InCites “My Organisation” functionality (available in early 2019) may be able to produce pre-aggregated data for analysis.

Many of the caveats discussed above for individuals still apply – for example, the use of impact factors, or counting the number of papers from a team in “prestige journals”, is not appropriate and should be avoided.

As groups always vary in size and productivity, normalised approaches are important here to give a fair comparison. The **category normalised citation impact** (CNCI, Web of Science) gives a normalised measure, as does the **percentage of highly-cited papers**. As these metrics are normalised by field, they can also be used (to some degree) to compare departments or groups in different disciplines. They are also normalised by year, making it practical to use them to consider changes in citation rates over time.

When comparing groups, remember that field-normalised indicators do not explicitly take account of many practical differences between groups. For example, they do not consider the effects of shared authorship, which may be substantial in one field and rare in another. The relative mix of staff – junior versus senior researchers, full-time versus part-time – may also need to be considered, as does any difference in focus on teaching versus research. If either department is in the arts, humanities, or social sciences, bear in mind that citation analysis approaches will be less appropriate and less reliable for these fields.

Comparing UCL to other institutions

The best way to obtain metrics to compare UCL with other institutions is through InCites, which provides a large range of data prepared on an institutional basis, and can be filtered by region, year, or discipline. It also provides some non-bibliometric figures, such as the papers or research income per researcher, which are useful for contextualising the data. It should always be remembered that different institutions have different specialities, which can affect many citation metrics; UCL, for example, has a much higher share of biomedical research than some comparable universities.

Good measures for comparison include the **percentage of documents in the top 1%** (or 10%), and the mean **category normalised citation impact** of its papers. These are both normalised by field and year.

The **percentage of international collaborations** can be useful, but the expected level of collaboration varies dramatically by field, and so care should be taken using it at an institutional level. It is most meaningful when used to show developments over time, or when activity in one particular field is being considered.

The simple **number of publications per year** can be useful in some cases, but it is not normalised by field, and is directly linked to the size of the institution so may not be meaningful for direct comparisons. The **h-index of an institution** should always be avoided, as it is not possible to scale this by the field of research or the underlying size of the institution.

The **mean citation impact** (and **impact relative to world**) should be treated with caution as it is not normalised by field or year. It may be appropriate if the output for a particular discipline is being considered but should not be used institution-wide or over a multi-year period. The **percentage of documents cited** is likewise field-dependent and year-dependent.

Reporting on change in publication and citation activity over time

The simplest approach is to sum the **number of publications** in each year. If there have been substantial changes in the size of the group being examined, it may be worth normalising this by the number of active researchers to avoid a misleading conclusion.

To look at citation impact, **mean citations per paper for each year** can be used in a comparable fashion. The most appropriate way is to count all citations received by papers in one year over the next two or three years. This means that it can be difficult to assess the change in impact for recent years; in many fields, citation data for papers in their first year or two is often quite limited and should be treated with caution.

The normalised citation impact metrics discussed above are also useful here – it is possible to look at the yearly change in the **category normalised citation impact**, or the **percentage of highly-cited papers** produced by the group in each year. Note, however, that these can give spurious results for small numbers of papers, or if based on particularly small research fields, where one or two outliers can account for a very large share of the output.

Measuring outputs in the arts, humanities, and social sciences

These fields are distinguished by their heavy use of non-journal publications, primarily book chapters and monographs. These publications are not well covered by the standard bibliographic databases (Scopus and Web of Science) – many books are not covered, while others have incomplete metadata or citation records – and so it is difficult to carry out meaningful analysis on them. These fields may also have a higher proportion of their journal articles in regional or non-English publications, which are also less well covered in citation databases.

Coverage is better in Google Scholar, but this database has other problems and inaccuracies which make it unsuited for bibliometric analysis. The situation has improved somewhat in recent years, with coverage of monographs increasing, but at the moment any citation analysis for these fields should be done with a great deal of caution. Any conclusions that can be drawn are likely to only represent a sample of the relevant publications and may be very misleading. There is, as yet, no widely accepted good solution here. In the absence of comprehensive publication databases, narrative based methods may be employed to allow scholars to present their research.

Nonetheless, some limited metrics are practical. For example, the measurement of **publication numbers** can be done using data from RPS, which is hopefully more comprehensive than the bibliographic databases as it will include a larger set of material, and this can be used to consider year-on-year changes.

In some fields, such as education, policy or practitioner impacts may be particularly significant. In these cases, using **alternative metric** approaches to identify work with substantial impact can be useful and appropriate, though these should only be used to highlight material for further investigation, and not used as quantitative metrics in their own right.

Assessing the quality of a group of journals

It is sometimes useful to compare a group of journals against each other, to help select a desirable publication venue. The traditional metric for this is the **journal impact factor (JIF)**, which can be found through Web of Science/InCites. For 2017 onwards, this service also provides graphs of the citation distribution, showing how the impact factor is calculated and whether it reflects a meaningful and representative picture of the journal's output. Impact factors are not comparable between different disciplines, but for comparing a group of similar titles, they can still be useful. **JIF quartiles** are also sometimes used, giving the journal’s approximate position in its discipline.

A group of conceptually similar but more complex metrics are also available. InCites provides the **Eigenfactor**, with weighting based on the citing journal, and the field-normalised **Article Influence Score**. Scopus gives the **CiteScore**, similar to the JIF but with a slightly different calculation, the **SCImago Journal Rank** (SJR), where citations are weighted by the citing journal, and the **Source Normalized Impact per Paper** (SNIP), normalised by field. Be cautious when using these indicators – some can be easily skewed by journals that only publish a few dozen articles per year, or focus heavily on review papers.

Note that other factors – such as turnaround time and acceptance rate – may be equally or even more important than the expected number of citations. These are rarely indexed and in some cases may not be easily discoverable. The **journal immediacy index** (InCites), which shows the average number of citations in the year of publication, can sometimes be informative here.

As always, it is important to reiterate that journal metrics do not necessarily reflect the individual papers in the journal, and should *never* be used as proxies to assess individual publications.

Assigning credit for multi-authored papers

In many fields, the majority of research is now authored by more than one person. In some fields, it is normal to have five to ten authors (hundreds or thousands in some circumstances) from a number of institutions. However, most bibliometric analyses tend to assume equal contribution from all authors, when authors may have contributed different amounts of intellectual and technical/experimental work. It should be noted that "gift authorship", where someone is credited despite not having contributed in any way to the research, is generally considered unethical and is against UCL's own research Code of Conduct.[10] From a bibliometric perspective, the practice complicates fractional weighting, because it diminishes the relative contribution of legitimate authors; there is no practical way to correct for it, but it may be important in some contexts to remember that it can be present.

There are a number of common approaches to assign credit for multi-authored papers. The simplest is to give **full credit** for the paper to each author. This means that someone who co-authored four papers will be considered to have the same level of output as someone who produced four on their own. This is not an inherently wrong approach, but it can be misleading, especially if comparing work in two disciplines with very different standards for co-authorship.

Where a contribution estimate is desired, **fractional counting** can be used. Here, each author receives credit in proportion to the total number of authors on the document; e.g., on a ten author paper, each author is credited with one-tenth of a paper. If multiple authors are from the same institution, these fractions can be summed to estimate an institution's contribution; alternatively, each institution may receive fractional credit proportional to the total number of institutions on a paper, independent of author number. Fractional counting gives a fairer sense of the productivity of prolific co-authors compared to less prolific solo authors, but it does not take account of any relative difference in contribution.

A common way to weight relative contribution is some form of **harmonic counting**, where the first author receives a large share of the credit, the second a smaller share, and so on. Again, this approach has limitations as it relies on a model where the first author is the primary contributor and the list is in descending order after that. However, in many fields, the convention is for the last author to have also been a substantial contributor; in others, it is conventional simply to list authors alphabetically. In cases like this, any attempt at weighting based on the author list will actively produce misleading results. It should not be carried out unless you are confident that all papers share a similar approach to author listing.

The most reliable way to account for relative contribution is to examine the papers and look for **statements of contribution**. These are becoming more common in many journals, and establish the contributions of individual authors and estimate their level of involvement. However, they are not indexed, and are rarely quantitative, meaning that they must be individually assessed.

[10] http://www.ucl.ac.uk/srs/governance-and-committees/resgov/code-of-conduct-research

In cases where there are dozens or hundreds of authors ("hyperauthorship"), it is unlikely that any of these approaches will give meaningful results. Author position in such papers is often purely alphabetical, statements of contribution are rare, and any fractional count would give only tiny values. Depending on the circumstance, it may be useful to give all authors a notional fixed share of the credit (say 0.05), or simply to avoid using these approaches for such papers.

Annex 5 - Responsible use of new and alternative metrics

In 2016 the National Information Standards Organization[11] identified three main use cases for new and alternative metrics ("altmetrics")

- Showcase achievements: Indicates stakeholder interest in highlighting the positive achievements garnered by one or more scholarly outputs.
- Research evaluation: Indicates stakeholder interest in assessing the impact or reach of research.
- Discovery: Indicates stakeholder interest in discovering or increasing the discoverability of scholarly outputs and/or researchers.

The table below shows UCLs own good practice interpretations of these use cases.

Use	Examples	Good practice and pitfalls
Showcase achievements	Researcher: assess the reach, engagement, and influence of my own research outputs, by, for example, incorporating altmetrics in my portfolio to complement my other accomplishments	✓ Formulate specific attention, engagement and impact goals *before* you start looking at your actual altmetrics[12] – then identify and use metrics as just one way of assessing your success and to provide insight. ✓ Use altmetrics to identify channels through which you can achieve your goals, not goals in themselves ✓ Assess a mix of metrics ✓ Look at the *content* of the attention you are receiving – what is the sentiment of the attention? High numbers of mentions might follow critical as well as favourable reaction. ✓ Look beyond *how many* to understand *who* is talking about your research – particularly for social media. ✓ Twitter isn't *all* cats and celebs – get to know how serious researchers are using it to hold global conversations around your research area[13]. × Passively 'tracking' numbers that happen to be available × Using metrics in the absence of a strategy or goal × Passively tracking numbers just because they happen to be available × Mistaking metrics that signify a process or pathway (such as blogs or tweets) as eventual societal impact
	Administrator: showcase the achievements of my organisation to other stakeholders. E.g. to demonstrate the achievements of my institution's researchers to potential hires, students, collaborators, and other researchers.	✓ Find out what metrics work best for each audience – ask them or consult their organisation strategy and stated goals. ✓ Re-use existing data – for example ResearchFish returns[14] ✓ Initiate a unit-wide conversation to develop overall goals and strategy around engaging and impact ✓ Use metrics as pointers to help you initially identify specific tangible and more narrative examples of impact. Follow these up with the researchers to check your understanding and what story UCL can tell. × Use numbers without context – especially if you aren't really sure where they come from or what they really represent

[11] Outputs of the NISO Alternative Assessment Metrics Project - A Recommended Practice of the National Information Standards Organization
[12] E.g Social media strategy for researchers - http://www.fasttrackimpact.com/single-post/2015/10/27/Create-a-social-media-strategy-for-your-research-that-delivers-real-impact
[13] E.g http://blogs.lse.ac.uk/impactofsocialsciences/2011/09/30/academic-tweeting-your-suggestions-and-tips-collected/ https://www.slideshare.net/UniofYorkLibrary/twitter-for-researchers-22963915
[14] Access / log in details can be requested from Jacob Sweiry, Head of Research Metrics.

		✗ Drawing spurious comparisons, between institutions or over time, is tricky due to changes in and partial nature of coverage of the tools over time, the underlying growth of social media / blogs etc. Without understanding this vital context quantified comparison is likely to be flawed at best.
	Administrator: support researchers in applying for competitive funding by effectively showcasing positive achievements of their research outputs.	✓Look at funders annual reports etc. to see what metrics they choose, if any ✓Search for existing more narrative impact stories from your area[15] ✓Talk to your Strategic Research Facilitation team[16] ✓ Look at case studies of how researchers can responsibly incorporate altmetrics in funding bids[17][18] ✓ Provide context and examples – for example, for public policy citations, what is the spread of countries? If your research were cited as influential in a public policy report, how did that actually come about? For example, do you have formal links with the government body such as advisors – (or was it just luck?) ✓ If your funder cares about demonstrating impact, show that you care about it too – hold training sessions and be a champion ✓ Combine metrics with qualitative information – for example, "Over 100 schools have downloaded the guide – unsolicited feedback from one said it was "one of the best guides to raise teachers capability they had seen". ✗ Citing numbers blindly, assuming they are 'good' when in fact they show little impact
Research evaluation	Librarian: Support the institution in their promotion / recognition exercises by offering impact-report services.	✓ Position altmetrics a starting point for discussion / further investigation ✓ Use altmetrics as a way to identify potentially interesting but unrecognised impact – but that expert-led and discipline-specific discussion is needed to interpret the data ✓ Recognise that researchers have different goals; E.g some excel in mass media and some excel through smaller highly specialist research-focussed groups. ✓ Use metrics as a learning opportunity to understand audiences and set direction ✓ Use caution and guard against high stakes decisions being unduly influenced by 'spammable' metrics such as downloads or retweets. ✗ Positioning metrics as 'scores' that can be used to rank performance between researchers or passively measure success ✗ Inadvertently set perverse incentives, such as to 'self-spam' downloads or retweets, by making the numbers alone seem an important part of promotion.

[15] http://iris.ucl.ac.uk/rics/search/impact-record (UCL log in required)
[16] http://www.ucl.ac.uk/research/contact
[17] https://www.altmetric.com/case-studies/?audience=researchers
[18] https://www.altmetric.com/blog/metrics-grant-application/

	Researcher: Gauge the performance and achievements of my / my institution's scholarly outputs	✓ Know what you want to achieve as a researcher and set this out *before* you set about assessing your current altmetrics. By all means then iterate your goals, but do be clear on those *first*. ✓ Ensure you are clear in your own mind what altmetrics do and do not measure. For example, the coverage of the tools. ✓ Always examine the actual content (who, when, where, what are they saying?) that lies behind the numbers × Conflate altmetrics (or indeed other bibliometrics) with research quality
Discovery	Researcher: Identify potential collaborators at other institutions with whom to partner on grant applications and other projects.	✓ Usually a straightforward task – see which countries, institutions and researchers are reading or accessing your work. ✓ Don't limit your discovery to researchers – stakeholders, users, audiences and beneficiaries can also be discovered; forming relationships with them often key to maximise the societal value of research
	Researcher: Discover influential research and associated debate that is important / interesting in my field	✓ Usually a straightforward task – e.g use Mendeley downloads to spot the most read papers in your area of interest long before citation metrics become available ✓ Consider using altmetrics to refresh readings lists × As with citations, high attention can be a marker of controversy or negative reaction as much as good research, so read the paper / views.
	Press officer: Identify popular / newsworthy papers or topics	✓ Discuss a press / media strategy before you start – who do you want to target, in which geographies, with which research topics, through which channels, and to what ends? ✓ Make use of geographical information to identify which countries UCL research has made the news – are any countries you would want to target missing? × Generating and measuring 'news for the sake of news' × Confusing newsworthiness with research quality or research worth
	Administrator: identify strong candidates or supporting evidence for a REF Impact Case Study	✓ Think through all of the uses you and your unit might have for a case study – not just to meet the needs of a REF Impact Case Study. E.g. sharing best practice and lessons with colleagues, telling your story to funders and the public. ✓ Only progress an Impact Case Study if you have been formally asked to do so for UCL – very few are required. If you have been asked to do so, the Research Impact Curation and Support team will provide 1:1 assistance and guidance. × Thinking of impact only in terms of REF (a very small sub-set)