

Guidance on metrics for papers with many authors

August 2020

bibliometrics@ucl.ac.uk

Overview

The [UCL bibliometrics policy](#) sets out some principles for the use of citation metrics in research assessment at UCL. As part of [the overall guidance](#), this paper sets out some general advice on interpreting metrics for papers with many authors.

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 large number of institutions, and single-authored papers are a rare exception. 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.

This can result in metrics giving an overestimate of the productivity of researchers in high-collaboration fields, especially when compared to researchers in lower-collaboration fields. These issues are difficult to correct for, and there are no easy solutions at the moment. However, there are some ways to avoid including papers with very large numbers of authors in aggregate metrics, and some approaches being developed which may help us better analyse individual contributions in future.

Possible approaches for interpreting multiple authorship

There are a number of common approaches to assign credit for multi-authored papers when calculating bibliometrics. 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. Conventions vary dramatically; what constitutes a sufficient contribution to be a named author in one field may be simply an acknowledgement in another, and these decisions are often informal and highly subjective.

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. There is a [growing movement](#) towards using approaches such as the [CREDIT taxonomy](#) to structure this information; if adopted, these may lead to more accurate accounts of contribution to be used in bibliometric analyses in the future. However, at the moment, they are not routinely indexed by databases, and analysing them individually may only be practical for small numbers of papers.

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. Absent clear statements of contribution, the only way to infer this is from author ordering.

To incorporate different levels of contribution, some form of **harmonic counting** can be used, 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, or you are able to assess each paper individually and choose an appropriate weighting method.

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 bibliometric approaches for such papers. This situation is most common in certain areas of high-energy physics, but may also be found in areas of biomedicine, or on large-scale review papers.

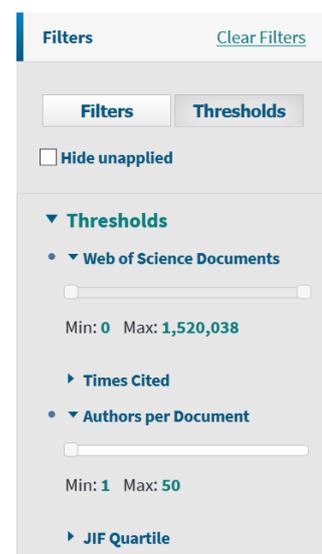
Finally, it should be noted that the subjective nature of authorship can make it open to abuse - "gift authorship", where someone is credited despite not having contributed in any way to the research, is a known problem. It is generally considered unethical and is against UCL's own research Code of Conduct.¹ 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.

Where can we find appropriate metrics?

Unfortunately, at this point in time, fractional or harmonic metrics are not routinely available from the citation databases. It is not possible to acquire these from Web of Science, Scopus, or InCites.

However, InCites does have some functionality to filter out papers with a large number of co-authors. This can be used to ensure that metrics are only calculated on the basis of papers where it is reasonable to assume each individual author made a substantial contribution.

To apply this, use the "thresholds" tab on the left-hand column in InCites, and select an appropriate maximum number of authors. For many fields, 50 is a reasonable threshold to use, but this will vary substantially depending on the discipline. It is not possible to look for the number of different *institutions* represented on a paper, only named authors. By changing the minimum number of authors instead, it is also possible to only analyse multi-authored papers, or only hyper-authored papers.



¹ <http://www.ucl.ac.uk/srs/governance-and-committees/resgov/code-of-conduct-research>

If you would prefer to download data from InCites as usual, and look at author numbers separately, column F contains a list of authors. It is possible to count this in Excel using a formula such as `=SUMPRODUCT (LEN (F2) -LEN (SUBSTITUTE (F2, ";", "")))` – however, note that the author list in the downloaded file is truncated, and will only include a maximum of 200 names. It will thus not display accurate counts for larger author lists.

InCites can be accessed through a link in the top bar of Web of Science, or else by going to <https://incites.clarivate.com>. It requires registration when you first use it – we recommend you use your UCL address here. You will need to be on the UCL network when you first register, so be sure to use a remote desktop or VPN if you are off-site.