Independent review of the role of metrics in research assessment: Call for evidence

Summary of areas for advice

1. This template provides a summary of the areas on which the steering group is seeking advice. You may use this template to respond with your advice if you would find it helpful.

2. Please send responses to **metrics@hefce.ac.uk** by **noon on 30 June 2014**.

Independent review of the role of metrics in research assessment: Call for evidence	
Name: Professor Ken Brown	**Job title: Vice-President**
Organisation: London Mathematical Society	
E-mail: lmspolicy@lms.ac.uk	**Phone: 020 7927 0801**

Identifying useful metrics for research assessment:

- What empirical evidence (qualitative or quantitative) is needed for the evaluation of research, research outputs and career decisions?
- What metric indicators are currently useful for the assessment of research outputs, research impacts and research environments?
- What new metrics, not readily available currently, might be useful in the future?
- Are there aspects of metrics that could be applied to research from different disciplines?
- What are the implications of the disciplinary differences in practices and norms of research culture for the use of metrics?
- What are the best sources for bibliometric data? What evidence supports the reliability of these sources?
- What evidence supports the use of metrics as good indicators of research quality?
- Is there evidence for the move to more open access to the research literature to enable new metrics to be used or enhance the usefulness of existing metrics?

In our answers to this and subsequent questions we shall restrict our answers to outputs, research activity and assessment in mathematical sciences.

- **What empirical evidence (qualitative or quantitative) is needed for the evaluation of research, research outputs and career decisions?**
  This is really several questions in one, the assessment of research outputs as in the REF being a completely different matter from the assessment of a grant application or of the suitability of an individual for appointment or promotion. Given the current context, we'll focus here solely on evaluation of outputs as part of a large-scale REF-style operation. Our fundamental position here remains that expert human judgement is the only assessment method which commands a reasonable degree of confidence from the mathematical science research community. Nevertheless, for future REF exercises it seems clear that the number of outputs to be assessed and the scale of the available resources for assessment of them (ie the number of person-hours available) mean that some use of metrics as an aid to assessment is inevitable. We will discuss in later sections ways in which the use of metrics could provide a useful adjunct to the assessment process.
- **What metric indicators are currently useful for the assessment of research outputs, research impacts and research environments?**
  Two particular types of metric seem relevant here: a ranking of journals; and citation indices. We shall discuss the use of each of these below.
- **What are the implications of the disciplinary differences in practices and norms of research culture for the use of metrics?**
  Aspects of the research and publishing culture in the mathematical sciences negatively affecting the value of metrics in research evaluation include the following:

  Use of journal rankings: The existence of very many high quality international subject-specific journals within the mathematical sciences (so that a fine-grained ordering is impossible); loyalty of authors to a particular editor or geographical area; publishing in a journal as a response to an earlier publication in the same journal**.**

  Use of citation indices: Worldwide numbers (of researchers, of outputs, of citations) are comparatively low in the mathematical sciences, so that random effects have a disproportionately large effect; the half-life of papers in the mathematical sciences is very long, so that the true impact of a paper only becomes apparent over a long time-scale; the gap between submission and publication of outputs in the mathematical sciences is long (more than 2 years is not unusual), so that many papers submitted in the REF do not have time to garner any citations; the risk of gaming, especially in fields where the number of researchers is small.
- **What are the best sources for bibliometric data? What evidence supports the reliability of these sources?**

(a) **Sources of citation data**: This is a complex question for the mathematical sciences, and is in itself another example of the problems in using citation data for assessment. For many areas of mathematics, in particular in particular those fields traditional described as “pure mathematics”, the MathSciNet database of the American Mathematical Society is by far the most important source of bibliometric data used within the community, (see http://www.ams.org/publications/math-reviews/math-reviews). This is because of MathSciNet’s great accuracy and convenience, and also because its value is hugely enhanced by dense hyperlinking to reviews of cited papers and to the actual journal literature. But MathSciNet is heavily skewed in its coverage towards pure mathematics. This greatly reduces the scope for its use across a wide-ranging assessment exercise, and indeed there are substantial areas of the mathematical sciences (including more applied fields, some parts of mathematical physics, theoretical computing science, and statistics), where MathSciNet is not used, because it does not cover outputs outwith mathematics, does not include the arXiv, and/or has patchy coverage of some areas of the subject. Many researchers in applied mathematics and statistics use Web of Science and Google Scholar, with some use also of more specialist databases (for example, http://inspirehep.net is heavily used within the mathematical physics community).

So, why not simply use Web of Science or Google Scholar as the source of data, given their fuller coverage? Simply because their data is, or is perceived to be, much less robust, and they too do not cover all relevant media. Even within areas of pure mathematics that one expects to be well covered by all these databases, simple comparative checks show wide disparities in the citation counts for a given paper. This absence of a reliable single source of data is one argument (but not the only one) against a mechanistic use of citation data in assessing research outputs in mathematical science.

(b) **Sources of journal rankings**: Within the mathematical sciences, some use is made of the ranking list of journals produced by a committee of the Australian Mathematical Society, sometimes called the ERA ranking, (http://www.austms.org.au/Rankings/AustMS_final_ranked.html), but this is of limited value since it features only 4 categories of journal quality, A*,A,B,C. The existence of a large number of high quality research journals with diverse subject focus, geographic location and editorial control is widely regarded within the mathematical sciences world-wide as a strength, countering any tendency for research to become more narrowly focussed through the pressure of transient fashion, or because of the excessive power which could be wielded by a small number of powerful editors. This diversity might be threatened by an attempt to produce a small list of “top journals”, so we expect that any such attempt would be vigourously resisted by the international community.

The Journal Impact Factor, as calculated by Thomson Scientific from citations in their indexed journals, is often used as a proxy measure of the

quality of a given journal. A comprehensive and damning critique of the value of this tool for assessing journal quality in the mathematical sciences was given in pages 4-7 of [1]. Moreover, in [2], pages 435-6, there is an analysis showing the poor correlation between Journal Impact Factor and the ERA journal ranking, even when restricting to the single subfield of applied mathematics.

[1] *Citation Statistics, A Report of the International Mathematical Union in cooperation with the International Council of Industrial and Applied Mathematics and the Institute of Mathematical statistics*, R. Adler, J. Ewing and P. Taylor, *Statistical Science (2009), 1-14.* Available at http://arxiv.org/pdf/0910.3529.pdf?origin=publication_detail

[2] "Nefarious Numbers", by Douglas N. Arnold and Kristine K. Fowler, *Notices of Amer. Math. Soc.* 58 (2011), 434-437; available at http://arxiv.org/pdf/1010.0278.pdf.

How should metrics be used in research assessment?

- What examples are there of the use of metrics in research assessment?
- To what extent is it possible to use metrics to capture the quality and significance of research?
- Are there disciplines in which metrics could usefully play a greater or lesser role? What evidence is there to support or refute this?
    - How does the level at which metrics are calculated (nation, institution, research unit, journal, individual) impact on their usefulness and robustness?
    - **To what extent is it possible to use metrics to capture the quality and significance of research?**

The use of metrics as a determining measure of the quality of research outputs in the mathematical sciences will inevitably produce many false results, and can never be a substitute for expert human judgement. That is not to say that metrics have no role to play – on the contrary, they can provide useful information, and – properly interpreted – may be able to speed up the reviewing process without severely compromising the quality of the judgements made. We discuss in more detail the possible use of the two sorts of data discussed above, in a large-scale research evaluation exercise such as the REF.

Journal rankings: Within the mathematical sciences generally, the peer-review process is regarded as a reasonably reliable (though not infallible) guardian of the correctness and novelty of published research. Publication in an internationally-recognised journal with a well-established system of peer review can therefore be taken as a reasonably safe assurance of a basic level of quality. But further than that, it is not possible to go with any confidence – that is, there is no reliable map from some ordered list of journals, to the ordered quality of the research articles they contain. As was already explained in our answer in the first box, authors in the mathematical sciences are habitually motivated

by a number of reasons when choosing where to publish research outputs, the perceived quality of the journal only being one reason among many.

We repeat for emphasis the point already made above: any attempt to produce a fine-grained set of rankings for journals in the mathematical sciences would be likely to have unintended damaging consequences for the research culture, and would therefore be strongly resisted by the international community.

Citation indices: The use of citation counts for individual papers as tool to rank the quality or the impact of these papers in an assessment exercise such as the REF is fraught with difficulties. The most important of these are as follows:

1. relatively small numbers of researchers in any given field meaning that random effects have a big influence;
2. long delay times in publishing (often 2 years between submission and a paper's appearance in a journal) mean that citations build up slowly, penalising papers submitted later in the assessment period;
3. the cited half-life of journals in the mathematical sciences is very long in comparison to many other disciplines, so the true impact of a paper cannot be judged within the time frame of a REF assessment interval;
4. cultural factors such as the typical level of scrutiny applied in the refereeing process (and hence the time delay between submission and publication), and the typical number of papers cited in a given article, vary greatly across subfields of the mathematical sciences.

The effects of (1), (2) and (4) being relatively clear, we don't discuss them further, beyond noting that their importance is exacerbated by (3). Factor (3) is intrinsic to the slow-burning nature of research in the mathematical sciences, a consequence of the facts that ground-breaking research in these fields can often take many years to produce, and that the subject is cumulative, the results of today often making use of work done 50 or more years ago. Thus, it is noted in [1, page 5] that an recent analysis of MathSciNet data showed that 90% of the citations to the journals in its reviewing database are to papers published more than 2 years previous to the citing paper, with roughly 50% of citations being to papers which are more than 10 years old. The cited half-life of the Proceedings of the London Mathematical Society, our flagship journal, was 33 years in 2013.[1]

In ways similar to those discussed above for journal rankings, the use of citation data as a proxy for quality in the assessment of research outputs in the mathematical sciences in (for example) a country would be likely to have serious damaging effects for that country's research culture – encouraging herd behaviour in the choice of research fields, discouraging moves into new and as yet small subfields.

Nevertheless, citation data can be a useful adjunct of an expert-based assessment of research outputs in the mathematical sciences. As an illustration, consider the case of

[1] The "cited half-life" of journal P in year X is the median age of the articles in P that were cited by other journals published during the year X.

paper A submitted to the REF, confirming a 20-year old conjecture stated first in paper X. A reviewer of A might well look at the citation data for paper X, to assess the level of interest in the conjecture across the preceding 20 years, and hence the likely impact of the new result. This shows how human judgement can be aided by appropriate citation statistics, but not in an obvious formulaic way.

A reviewer could also use citation metrics as a tool to help focus her work: thus, when reviewing a batch of papers from a coherent subfield which had all been published at least 3 years before, she might bias her detailed reading somewhat towards those which had few citations. That is, she is saving time by taking a bet that high citations imply high impact and hence high quality; but the converse of this bet – that few citations implies low quality – is most definitely false, and must be strongly resisted.

We emphasise finally here our key point, making specific reference to REF2014. Citation data in the mathematical sciences is a useful tool when employed by expert human readers as part of their evaluation process, but can never replace expert reviewers. It is vital for the success of future assessment exercises that the process retains a reasonable level of confidence in the assessed community. For this to happen, a substantial number of expert-person-hours must be devoted to this job. In particular, the situation prevailing in REF2014, where the number of panellists was substantially reduced compared with RAE2008, and where the assessment of outputs was compressed into the first 5 months of 2014, with many panellists having to assess 500 outputs or more, cannot be repeated if confidence is to be maintained.

'Gaming' and strategic use of metrics:

- What evidence exists around the strategic behaviour of researchers, research managers and publishers responding to specific metrics?
- Has strategic behaviour invalidated the use of metrics and/or led to unacceptable effects?
- What are the risks that some groups within the academic community might be disproportionately disadvantaged by the use of metrics for research assessment and management?
- What can be done to minimise 'gaming' and ensure the use of metrics is as objective and fit-for-purpose as possible?

- **What evidence exists around the strategic behaviour of researchers, research managers and publishers responding to specific metrics?**
- **Has strategic behaviour invalidated the use of metrics and/or led to unacceptable effects?**

A comprehensive critique of the opportunities for gaming afforded by the use of bibliometric data in research assessment in the mathematical sciences is given in [2]. This article, as well as making a number of general points about the features of research publication in the mathematical sciences which make it particularly susceptible to gaming, also includes a number of illuminating examples of gaming in action. This analysis makes it completely evident that any useful and reliable employment of metrics in evaluating individual research outputs in the mathematical sciences is only feasible as

an adjunct to expert peer-review of the outputs.

International perspective:

In addressing the issues and questions above, please include relevant evidence and examples from outside of the UK, where appropriate.

Research in the mathematical sciences as practiced in the UK is totally international in its practices, outlook and culture. The evidence which we have cited above, and which we list again here for convenience, is completely international in its perspective and authorship. We take the opportunity also to list [3], the brief commentary by the authors of [1] on the discussion of their paper in Statistical Science. It's worth noting that, among the authors listed below, John Ewing was Executive Director of the American Mathematical Society for 14 years, and is President of Math for America, and Douglas Arnold is past president of the Society for Industrial and Applied Mathematics.

[1] "Citation Statistics, A Report of the International Mathematical Union in cooperation with the International Council of Industrial and Applied Mathematics and the Institute of Mathematical Statistics", R. Adler, J. Ewing and P. Taylor, *Statistical Science 24 (2009), 1-14.* Available at http://arxiv.org/pdf/0910.3529.pdf?origin=publication_detail

[2] "Nefarious Numbers", Douglas N. Arnold and Kristine K. Fowler, *Notices of Amer. Math. Soc.* 58 (2011), 434-437; available at http://arxiv.org/pdf/1010.0278.pdf.

[3] "Rejoinder: Citation Statistics", R. Adler, J. Ewing and P. Taylor, *Statistical Science 24 (2009), 27-28.* Available at http://projecteuclid.org/download/pdfview_1/euclid.ss/1255009007

Journal rankings of the Australian Math. Soc., http://www.austms.org.au/Rankings/AustMS_final_ranked.html

MathSciNet, http://www.ams.org/publications/math-reviews/math-reviews

The Inspire High Energy Physics database, http://inspirehep.net

Would you be interested in participating in a workshop/event to discuss the use of metrics in research assessment and management? Yes.