

Annex G: Discussion of bibliometrics by the Expert Advisory Groups

The use of bibliometrics in the REF

Introduction

1. The second round of meetings of the Expert Advisory Groups (EAG) for the Research Excellence Framework (REF) took place in April and May 2009. We held two two-day events, which covered the use of bibliometrics in the REF, expert review of outputs, how to assess research environment and how to take account of impact.
2. A summary of the discussion of all the EAG meetings can be viewed at www.hefce.ac.uk under www.hefce.ac.uk Research/Research Excellence Framework/Resources.
3. Seventy-eight members attended the first day on the use of bibliometrics in the REF, splitting into six breakout groups. In advance of the meeting, members were given an overview of the results of the bibliometrics pilot, and access to data relating to their specific disciplines. In the morning session, members were given information about the pilot exercise and had the opportunity to ask questions. In the afternoon session, members broke out into small groups to discuss the interpretation of the pilot outcomes and the potential for using bibliometrics in the REF.
4. This annex provides a summary of the discussions on bibliometrics from the first day of the meetings.

Robustness of bibliometrics pilot results

5. In many disciplines (particularly in medicine, biological and physical sciences, and psychology), members reported that the 'top six' model (which looked at the most highly cited papers only) generally produced reasonable results, but with a number of significant discrepancies. In other disciplines (particularly in the social sciences and mathematics) the results were less credible, and in some disciplines (such as health sciences, engineering and computer science) there was a more mixed picture.
6. Members reported that the other two models (which looked at 'all papers') did not generally produce credible results.
7. A number of reasons for the variations in robustness between disciplines, and for the discrepancies in the results within a discipline, were identified:
 - a. Different sets of papers were looked at in the RAE and in the pilot. Some members suggested repeating the citation analysis using only those outputs assessed in the RAE.
 - b. The volume of citations and the time taken to accumulate citations varies between disciplines; citation indicators are more robust in disciplines that publish and cite more frequently.
 - c. The coverage of citation databases is limited in a number disciplines, particularly where non-journal outputs are common.

- d. Citations measure impact on the academic community; this is only one aspect of quality, whereas the RAE results represent a rounded view of quality.
- e. Citations do not provide a good measure of applied research and cannot take into account non-academic impact.
- f. Other limitations of bibliometrics that could distort some of the results, such as negative citations.
- g. Some of the sample sizes were small and these tend to be less stable.
- h. More recent papers have had less time to accumulate citations. Even though publication year is taken into account in the analysis, the results were less robust for papers published in the more recent years.
- i. Limitations with the normalisation process. In particular:
 - i. The categorisation of journals into fields was felt to be problematic in a number of fields (for example where diverse journals are used, such as in Statistics), and for a number of journals (particularly broad journals that cover several sub-fields such as the Lancet, British Medical Journal or Physical Review).
 - ii. Citation rates were normalised against a worldwide 'mean' for the field; yet the distribution of citations is highly skewed.
- j. Differences in the two commercial citation databases (Web of Science and Scopus) led to some marked differences in the results. A few members noted that other databases were more widely used by their disciplines (such as arXiv and Google Scholar).
- k. The way items are categorised within the databases as 'articles', 'review papers' and so on can differ from the way institutions or researchers would classify them. Some material on the databases (for example in 'trade' journals) would not be considered research.
- l. The mix of sub-fields within a submission can affect citation indicators; for example a submission can be dominated by a highly cited sub-field within Physics.
- m. In a few cases members reported discrepancies between RAE outcomes and citation indicators, where the RAE scores appeared to reflect the prestige of the journals papers were published in, whereas the citation rates for the papers provided a different picture.

Use of bibliometrics in the REF

- 8. There was a strong consensus that bibliometrics would be a useful aid to expert review, but could not be used formulaically, due to the range of limitations and discrepancies in the data. Expert review would still be required to take these into account and to ensure the credibility of the process.
- 9. There was a strong consensus that bibliometrics should be applied to selected papers only. Members agreed that the 'address-based' model was undesirable for a number of reasons, not least the substantial problems in associating papers with the

relevant UOA. Of the two 'author-based' models, members felt that selected papers would be more useful and informative, providing a better discrimination of quality. There was no consensus on what value information on all papers would add, and members raised concerns that assessment of all papers would disincentivise speculative research and lead to other adverse behaviours.

10. Members discussed a number of ways in which panels could make use of citation data to enhance the reliability and consistency of expert review and/or to reduce panels' workloads. There was no clear consensus on a single approach and members felt that the particular ways in which panels could make use of the data should vary as appropriate to the discipline. The range of possible uses included:

- a. To inform the reading of individual outputs (most groups supported this approach although some were concerned about using citation data in this way).
- b. As indicators for each submission as a whole, to sense check or provide a 'challenge' to the panel's scores based on reviewing the outputs.
- c. To inform 'borderline' decisions.
- d. To provide benchmarks against international standards and aid calibration against the quality descriptors.
- e. To inform discussions about consistency between panels, or enable comparisons across disciplines.
- f. To enable panels to sample and reduce the number of outputs to be reviewed in detail. Some suggested that bibliometrics could form part of a stratified random sampling procedure; however some members were sceptical about this and many generally doubted that bibliometrics would enable panels to read fewer outputs.

11. Members discussed the type of citation information that would be useful to panels:

- a. Many agreed that all panels that make use of citation data should be provided with the same types of data, but that they could use or interpret the data differently as appropriate.
- b. Limitations with the normalisation method were raised. Members generally agreed that panels would want the 'raw' citation count, in addition to data that enable them to interpret this within an international context. This could be a benchmark for the field, or an indication of where the citation count falls within the worldwide distribution for the field (a centile). Panels would also be interested in a benchmark or centile for all papers submitted to the UOA.
- c. There was also interest in the kinds of contextual data provided from the pilot, relating to the sources of citation (local, national and international) and international co-authorship.
- d. There was some discussion about which citation database(s) should be used and many felt that the REF should not be limited to using a single database across all panels.

12. Some issues about the potential behavioural consequences of using bibliometrics were raised:

- a. If bibliometrics were to be used in different ways across sub-panels, this could influence institutional decisions about where to submit members of staff, or their decisions about which types of outputs to submit to different panels (for example, selecting on the basis of citations for some panels, and the implication that this could favour older papers or disadvantage early career researchers).
- b. Publication behaviours might be affected, for example movement towards more frequently cited journals.

13. Members discussed whether the benefits of using bibliometrics would outweigh the costs. Some found this difficult to answer given limited knowledge about the costs. Nevertheless there was broad agreement that overall the benefits would outweigh the costs – assuming a selective approach. For institutions this would involve a similar level of burden to the RAE and any additional cost of using bibliometrics would be largely absorbed by internal management within institutions. For panels, some members felt that bibliometrics might involve additional work (for example in resolving differences between panel judgements and citation scores); others felt that they could be used to increase sampling and reduce panels' workloads.

Further development

14. A number of areas for further work were suggested:

- most importantly, to develop normalisation techniques including ways of categorising journals
- assessing the accuracy of the databases
- investigating how career stage affects citations (particularly for early career researchers)
- understanding the behavioural implications of the preferred model
- investigating other citation sources (such as Google Scholar)
- understanding coverage of the databases (including, for example foreign language journals)
- extending the analysis to include conference proceedings.