TECHNICAL REPORT

Bibliometrics as a tool for supporting prospective R&D decision-making in the health sciences

Strengths, weaknesses and options for future development

Sharif Ismail, Edward Nason,
Sonja Marjanovic, Jonathan Grant

Prepared with funding support from the Health R&D Policy Research Unit of the Department of Health (England)

RAND EUROPE

The research described in this report was prepared for the Department of Health England.

Published 2009 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
Westbrook Centre, Milton Road, Cambridge CB4 1YG, United Kingdom
RAND URL: http://www.rand.org
RAND Europe URL: http://www.rand.org/randeurope
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

Bibliometric analysis is an increasingly important part of a broader 'toolbox' of evaluation methods available to R&D policymakers to support decision-making. In the US, UK and Australia, for example, there is evidence of gradual convergence over the past ten years towards a model of university research assessment and ranking incorporating the use of bibliometric measures. In Britain, the Department of Health (England) has shown growing interest in using bibliometric analysis to support prospective R&D decision-making, and has engaged RAND Europe's expertise in this area through a number of exercises since 2005. These range from the macro-level selection of potentially high impact institutions, to micro-level selection of high impact individuals for the National Institute for Health Research's faculty of researchers.

The aim of this document is to be an accessible, 'beginner's guide' to bibliometric theory and application in the area of health research and development (R&D) decision-making. The report also aims to identify future directions and possible next steps in this area, based on RAND Europe's work with the Department of Health to date. It is targeted at a range of audiences, and will be of interest to health and biomedical researchers, as well as R&D decision-makers in the UK and elsewhere. The report was produced with funding support from RAND Europe's Health R&D Policy Research Unit with the Department of Health.

RAND Europe is an independent, not-for-profit policy research institute based in Cambridge. For the past three years, RAND has been a designated research unit of the Department of Health, which supported the research for and publication of this document.

For more information about RAND Europe or this document, please contact:

Dr Jonathan Grant
President
RAND Europe
Westbrook Centre
Milton Road
Cambridge CB4 1YG
Tel: +44 1223 353329
Email: jgrant@rand.org

Contents

List of figures

List of tables

List of abbreviations

AMS	Academy of Medical Sciences
BRC	Biomedical Research Centre
BRU	Biomedical Research Unit
CPP	Citations Per Paper
CWTS	Centre for Science and Technology Studies, University of Leiden
DH	Department of Health (England)
HCP	Highly Cited Paper
JIF	Journal Impact Factor
NHS	National Health Service
NIHR	National Institute for Health Research
RAE	Research Assessment Exercise
RQF	Research Quality Framework

Summary

Over the past 10 years, there have been substantial increases in funding for health and biomedical research by public bodies in a number of countries, including Canada, the UK, and the United States, among others. This has coincided with moves towards improved accountability in public service provision in general, and growing demand for both *ex ante* and *ex post* evaluations[1] of research in the health and biomedical research fields in particular. RAND Europe has expertise in both approaches, but has over the past few years undertaken a substantial body of *ex ante* work on behalf of the English Department of Health using a quantitative methodology known as bibliometrics, which has come increasingly to be used in combination with other evaluation approaches to help inform the Department's decision-making.

What is bibliometrics?

Bibliometrics employs quantitative analysis to measure patterns of scientific publication and citation, typically focusing on journal papers. It is one of a set of evaluation methodologies – including case study analysis, peer review, economic rate-of-return analyses and surveys and consultations (among others) – that may be used to help assess the impact of research in the health sciences.

Why use bibliometrics?

Bibliometric approaches offer important advantages over other research evaluation methods. They can be used to generate useful quantitative indicators of collaboration and measures of interdisciplinary research. As the sophistication of analytical tools improves, they are being used to develop more general indicators of 'quality' and even 'excellence'. These analyses are supported by a range of indicators of varying complexity which have been developed over recent years. Robust bibliometric analysis requires a clear understanding of the strengths and limitations of each of these measures, and sensitivity to the contexts in which they are used.

---

[1] There are, broadly speaking, two types of evaluation approach. On the one hand, an *ex ante* approach seeks to evaluate research proposals before they are conducted: i.e. with a view to selecting the most appropriate proposals for funding support. This is the form of evaluation involved in grant peer-review processes. On the other hand, *ex post* evaluation involves assessing the impact of research once it has been completed. *Ex ante* evaluation typically involves some element of *ex post* analysis, since the prior record of researchers is often reviewed as part of the assessment of whether or not to provide them with further funding support.

Ultimately, the analytical power of bibliometrics derives mainly from use in *combination* with other methods, rather than independently. For example, there is growing consensus that bibliometric analysis can be used successfully to complement peer-review decisions. This is particularly the case for large-scale peer-review exercises, where the volume of material to be analysed, and indeed its complexity, may be such that some form of quantitative validation may be useful.

What does *ex ante* bibliometric analysis look like in practice?

RAND Europe has undertaken a substantial body of bibliometric work for the English Department of Health over the past few years, in partnership with the Centre for Science and Technology Studies (CWTS) at the University of Leiden, the Netherlands.[2] Specifically, bibliometric analysis has been used to support the selection of appropriate academic institutions in the UK as biomedical research centres, academic research departments as biomedical research units, and individuals as faculty members at the National Institute for Health Research. In all three cases, bibliometric analysis was used to evaluate the prior academic performance of applicants, with a view to identifying high achievers and assisting in the selection process.

How robust is bibliometric analysis?

There are considerable advantages to a bibliometric approach, especially as the power and range of indicators available improves – however, a clear understanding of limitations and caveats is required. From a theoretical perspective, some doubts remain as to the ability of bibliometric methods to capture abstract concepts such as research 'quality'. Methodological challenges include issues of journal coverage in major bibliometric databases, adequately identifying author affiliations and choosing the right timeframe for analysis. Caveats to bibliometric analyses include variations in citation behaviour between fields and individuals, and a perennial difficulty in evaluation: attribution. While it is usually possible to determine whether research work *contributed* to the content of particular publications, often *attributing* publications solely to particular bodies of research is very difficult, if not impossible.

How could the use of *ex ante* bibliometric analysis be improved?

Further development in some discrete areas could strengthen significantly the analytical power of *ex ante* bibliometric assessments. First, by investigating the linkages between publication and citation patterns and broader impacts of individual researchers, groups or institutions (such as the health gains resulting from research work, or economic benefits accruing to the wider economy), it may be possible to inform strategic funding decisions to maximise economic returns. Second, new and robust indicators may be developed to identify up-and-coming researchers. Third, cross-checking systems may help to identify the small number of researchers, groups and institutions who effectively 'play the system' by manipulating their publication profile – and thus their bibliometric indicators – without any commensurate change in the quality of their research. Other work may include developing indicators of close-to-patient work for funders looking to maximise the impact of their funding on health outcomes.

---

[2] See: http://www.cwts.nl/cwtsbv/index.html (accessed 7 August 2009).

Acknowledgements

The research for this report was undertaken with funding support from RAND Europe's Health R&D Policy Research Unit with the Department of Health. The authors would like to thank Steven Wooding for his contributions to formulating and determining the focus of this report. We would also like to thank Linda Butler, Stijn Hoorens and David Kryl for their very valuable comments on earlier drafts of this document.

Any errors or omissions that remain are our own.

CHAPTER 1 Introduction

Chapter summary

- Bibliometrics is one of a set of evaluation methodologies, including case study analysis, peer review, economic rate of return analyses and surveys and consultations (among others), which may be used to help determine the impact of research in the health sciences.
- The growing use of bibliometric approaches to measure patterns of scientific publication comes against the background of stronger demands for evaluation in public policy in general, and – for the purposes of this report – research policy in particular.

Bibliometrics uses quantitative analysis to measure patterns of scientific publication, typically focusing on journal papers. Over the past 40 years, it has emerged as a branch of the wider field of infometrics, and has become particularly prominent as an evaluation tool over the past 20 years. What explains the rising prominence of this methodological approach, and what particular advantages does it offer? This chapter reviews some important changes in the public administration landscape in general, and the administration of scientific research in particular. These changes have spurred an increasing focus on evaluation of scientific research, and with it, the search for new, supporting methodologies. Bibliometrics has emerged as an increasingly important tool in this context.

1.1 Setting the scene: why evaluate?

The growing popularity of bibliometric approaches must be set against a background in which evaluation is encouraged increasingly in public policy in general, and research policy in particular. At general public policy level, governments increasingly encourage the development of strategic goals and performance indicators to guide reform and organisational change. In the USA, the Government Performance Results Act of 1993 requires federal agencies – including those that fund research – to set research goals and use performance measures for management and budgeting.[3] In Canada, a report by the

[3] The text of the Act can be found here: http://www.whitehouse.gov/omb/mgmt-gpra_gplaw2m/ (accessed 7 August 2009).

Auditor General concluded that departments and agencies should establish the mechanisms and practices that they need to demonstrate the results of their science and technology activities and to ensure that their resources are allocated effectively (Auditor General of Canada, 2004). In the UK, by contrast, the National Audit Office noted in 2003 that government departments "have no systematic mechanisms for measuring the overall impact of their research effort" (National Audit Office, 2003: 7).

The development of an increasing emphasis on efficiency and effectiveness applies to government-funded research in most Organisation for Economic Co-operation and Development (OECD) countries, where there has been growing pressure for information-gathering to support research allocations, reorientations of research support, rationalisation or even wholesale restructuring (Moed, 2007). In the UK health sector, for example, there have been important moves to address a perceived impact measurement deficit over the past five to 10 years. In 2004, the Government published a 10-year Science and Innovation Investment Framework (HM Treasury, 2004). A key requirement under the Framework was that research councils should feed into a performance management system run by the Office of Science and Innovation, which was intended to demonstrate the contribution that each council makes to achieving government targets. This system included delivery plans, an outputs framework of performance metrics, and a scorecard of targets and milestones.

Increasing emphasis on health research evaluation in the UK echoed similar moves elsewhere (not least in the USA), and – besides the impetus provided by legislation – responded to two broad changes in the scientific research arena. The first was a perception that public expenditure on research had been in decline since the early 1980s, and that competition for available funding was now stronger (Verbeek et al., 2002); this was despite a notable increase in public expenditure on supporting research in the UK since the general election of 1997 (MacLeod, 2005). The second was a growing sense that the peer-review system – which had long been regarded as the best way of selecting research proposals for funding – was now under considerable administrative strain as a result of the burden of applications submitted. In short, its limitations were exposed increasingly (Verbeek et al., 2002).[4]

In addition, there has been a strategic dimension to these changes. Specifically, there is a growing demand for funding agencies to evaluate the impact of the research that they fund, in order to provide enhanced impetus for research translation exercises – in other words, to promote the translation of basic research into practice (Council on Health Research for Development, 2000; Grant et al., 2004; National Audit Office, 2003; Smith, 2001). An important reflection of this is the growth of interest in improving systems of *ex ante* evaluation as a means of improving funding decision-making. Overall, however, research designed to evaluate research contributions in a range of areas (e.g. knowledge production, capacity building, health impacts and wider economic impacts) remains in short supply.

[4] More prominent critiques are briefly discussed in Chapter 2, and include problems of cost and administrative burden, and accusations that the peer-review system is biased (with respect to gender and age, among other issues) and even anti-innovation.

Partly in recognition of this deficit, the Academy of Medical Sciences, the Wellcome Trust and the Medical Research Council established the UK Evaluation Forum in 2005. The next year, the Forum delivered a report entitled *Medical Research: Assessing the Benefits to Society* (UK Evaluation Forum, 2006). A key contention of the report was that measuring the performance and results of research in practice is a challenging and complex exercise. First, many research impacts are not easily quantifiable. Second, it is difficult to attribute a policy or clinical impact to a particular research project (the problem of attribution is one which we will return to in greater depth in Chapter 2). Despite these difficulties, the report outlined a number of strategies – ranging from assessment of outputs and outcomes to full-scale economic evaluation – to help funding bodies understand the impact of the research that they support.

1.2 Doing evaluation: the research evaluator's toolkit

Table 1.1: Key methodologies in the research evaluator's 'toolbox'

Evaluation method	Characteristics	Advantages	Disadvantages
Bibliometric analysis	Can be narrow and deep or broad and shallow	• Quantitative, measuring volume of output • Can be used to indicate quality of output • Enables analysis of global trends	• Estimates of quality based on citations can be misleading • Data must be normalised to compare across research fields
Case study analysis	Narrow and deep	• Provides in-depth analysis of the process of discovery • Can demonstrate pathways from basic science to application	• Selection bias: how do we know that the chosen cases are representative? • Highly resource-intensive to do well
Systematic peer review	Narrow and deep	• Well-understood component of research management • Widely accepted by both the 'establishment' and researchers themselves	• Time-consuming for experts involved • Concerns over the objectivity and reliability of findings
Surveys and consultations	Can be narrow and deep or broad and shallow	• Can identify outputs and outcomes associated with particular pieces of funding or research • Provides qualitative analysis of outcomes	• Dependent on contact details being available for researchers in question • Poor response rate can limit findings
Economic rate of return 1: micro-economic analysis	Broad and shallow	• Can be applied to different sectors • Comparative potential, e.g. cost–benefit analyses	• Difficult to put a financial value on many of the influences involved
Economic rate of return 2: macro-economic analysis	Broad and shallow	• Quantitative • Provides 'big picture' and context of research	• Difficult to identify the contribution of an individual sector or funder

Source: Adapted from UK Evaluation Forum (2006)

Demands for improved evaluation can be translated into practice in a number of ways. The UK Evaluation Forum's report included an assessment of a cluster of key research evaluation methodologies. Its findings are summarised in Table 1.1 above.

An important factor when selecting the most appropriate form of research evaluation methodology is the *level* and *scope* with which it is to be applied. The present report distinguishes between macro, meso and micro-level analyses:

- macro-level analyses refer principally to entire research fields;
- meso-level analyses refer to institutional or organisational assessments;
- micro-level analyses refer to assessments of individual researchers.

With respect to scope, the report considers two approaches: 'broad and shallow' and 'narrow and deep'. Broad and shallow approaches aim to quantify the large-scale effects or quality of research. Narrow and deep evaluations focus on understanding how research funding could be improved, and how the process of translation could be accelerated.

1.3 Organisation and structure of the report

This report is divided into four further chapters. Chapter 2 focuses principally on theoretical aspects, examining the evolution of bibliometrics as a branch of the wider field of infometrics within the sociology of science, before discussing a variety of common measures and indicators of research output. Chapter 3 outlines some examples of recent RAND Europe work in this area, focusing specifically on *ex ante* evaluation in support of decision-making at the Department of Health (England). The chapter looks at the use of bibliometrics to select biomedical research centres, biomedical research units, and individual members of the National Institute for Health Research's (NIHR) Faculty of Senior Investigators. Chapter 4 reviews some of the advantages, disadvantages and most potent challenges to bibliometrics posed in the academic literature, and highlights a number of overarching caveats to any analysis carried out in this way. This discussion provides a preface to Chapter 5, where a number of suggestions are made as to the ways in which current *ex ante* bibliometric assessments could be developed and improved to support more sophisticated analysis. This discussion builds on RAND Europe's experiences to date.

CHAPTER 2 Structuring bibliometric analyses

Chapter summary

- Bibliometrics uses quantitative analysis to measure patterns of scientific publication, typically focusing on journal papers.
- Bibliometric approaches offer important advantages over other research evaluation methods, but their analytical power derives mainly from use in *combination* with these methods, rather than independently.
- Specifically, bibliometric methods can be used to generate useful quantitative indicators of interdisciplinarity and collaboration, as well as more general indications of 'quality' and even 'excellence'.
- These analyses are supported by indicators of varying complexity which have been developed over recent years. Robust bibliometric analysis requires a clear understanding of the strengths and limitations of each of these measures, and sensitivity to the contexts in which they are used.

This chapter provides an overview of bibliometric methods, exploring their theoretical bases, characteristics of the data typically used in analysis, and the range of bibliometric indicators that are used most commonly. Perhaps most importantly, it attempts to situate bibliometrics within the range of other evaluation methodologies used most commonly, while highlighting the areas in which it can offer greatest analytical power.

2.1 What is bibliometrics?

2.1.1 Bibliometrics is part of a family of methods for assessing scientific communication

Bibliometrics has emerged as a branch of the wider field of infometrics, a field devoted to quantitative studies of science and technology. It uses quantitative analysis to measure outputs and the scientific impact of publications. The earliest definition was offered in 1969 by Alan Pritchard, one of the pioneering researchers in this field, who described bibliometrics as "the application of mathematics and statistical methods to books and other media of communication" (1969: 349). However, bibliometric analysis is applied more commonly now to journal papers than books.

Bibliometrics offer insights principally along four dimensions (adapted from Narin et al., 1994):

1. activity measurement – counts of articles as a measure of the *volume* of outputs in a given research field. This is, in effect, a measure of the size of scientific activity;
2. knowledge transfer measurement – on the basis that the citation process reflects the communication of knowledge within the scientific community and provides an indirect measure of research quality. Analytical approaches here can include attention to the number of times that articles are cited subsequently in the research literature. However, there are a considerable number of possible approaches to this (many of which will be reviewed later in this report);
3. linkage measurement – involving the assessment of links between individuals and research fields as an indication of the way in which the social and cognitive networks of scientific research are developed and sustained. Lately, this has been developed as part of field mapping exercises, demonstrating cross-disciplinary links and collaboration between researchers or research groups;
4. citation analysis – as a proxy for one or more dimensions of the *quality* of scientific output.

2.1.2 Publication and citation data have important characteristics

Bibliometric analysis involves some important assumptions about the pattern of journal publication, and how researchers cite others' work once it has been published. First, it is assumed that citations begin accumulating once a paper has been published; and second, that it is very rare for a paper to stop accumulating citations altogether – even if citation rates tail off at a given time after publication. Given these underlying assumptions, there is growing interest in examples of papers classified as 'mayflies', which show exceptional early impact, and then tail off, or 'sleeping beauties', which may be ignored for years before suddenly being found to be of major significance (Evidence Ltd, 2007; Moed, 2005).

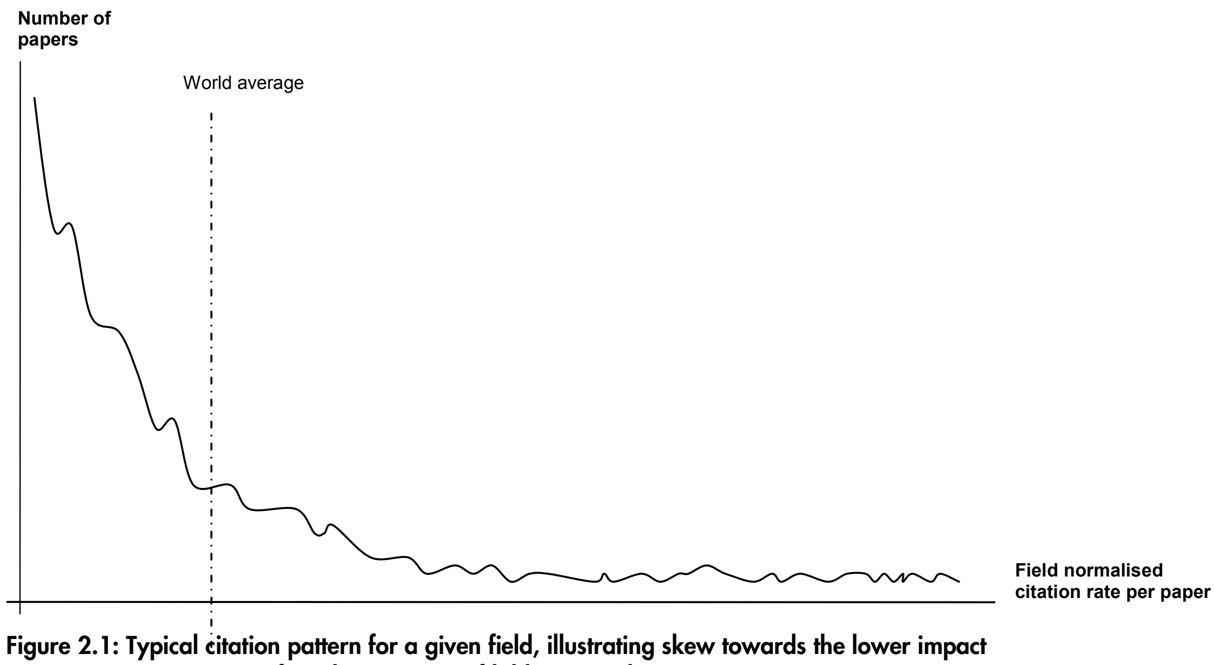

Figure 2.1: Typical citation pattern for a given field, illustrating skew towards the lower impact categories often characteristic of bibliometric data

Source: Adapted from Evidence Ltd (2007)

Third, there is an understanding among many working in this field that data distributions are always skewed, with vastly more papers in any given research field at the lower end of the impact spectrum (Evidence Ltd, 2007). This phenomenon is illustrated in Figure 2.1.

2.2 Why do we need bibliometrics?

2.2.1 Well-recognised problems with other evaluation methodologies

The increasing popularity of bibliometric analysis derives at least in part from unease with respect to the robustness of more widely-used research evaluation methods. Foremost among these is systematic peer review, by far the most widely-practised form of research evaluation. Well-understood and broadly accepted by large swathes of the research establishment, it is also the process by which around 95% of the £2 billion spent annually on medical research in UK is distributed.[5] However, the general term 'peer review' covers a multitude of sometimes quite different processes – some more effective than others. There are some notable criticisms of peer review-based approaches to research evaluation and funding distribution, although it is important to note that the evidence supporting some of these contentions is not particularly strong.[6]

In view of these challenges, there is growing interest in examining the potential of alternatives to conventional peer-review processes,[7] or at a minimal level, improvements to existing systems to improve their efficiency and effectiveness. A major difficulty when looking for alternatives is that the methodologies described in Table 1.1 above also have clear limitations. Case study research, for example, may provide extremely useful insights into the translation of research from bench to bedside (in the case of healthcare), but it is unlikely to involve sample sizes large enough for statistically significant findings to be made. In any case, using this approach for very large-scale research evaluation exercises would impose an impossibly heavy administrative burden on those participating. Similarly, economic rate-of-return analyses are constrained by the fact that they are specifically formulated at either macro or micro levels; macro-level analyses provide a good overview but cannot tell us much about institutional or individual performance, and deal mainly with a particular kind of impact.

2.2.2 Bibliometric analysis offers some clear advantages over alternative approaches

There are two schools of thought on the potential role of quantitative indicators in research evaluation. On the one hand, they may be used to enhance conventional peer review – and on the other, more radically, that they might come to replace peer review altogether (Research Evaluation and Policy Project, 2005).

---

[5] Calculated using figures published on the website of the Association of Medical Research Charities (AMRC) and combining these with known expenditures by other third sector and public sector organisations in the UK.

[6] See Ismail et al. (2009) for a fuller review of the evidence on the strengths and weaknesses of grant peer review.

[7] In the case of grant peer review, this process typically involves external (and sometimes internal) academic reviewers in the process of deciding which applications to a funding body are rewarded with financial support, based on assessments of funding applications.

The latter view was advanced most publicly in the late 1990s, mainly in response to data on the spiralling cost of the Research Assessment Exercise (RAE)[8] in the UK, and a feeling that a quantitative approach would help to generate significant cost savings (Oppenheim, 1997). Although several recent studies show a reasonably good correlation between RAE scores in the UK and bibliometric scores for research in a range of disciplines, there are deviant cases, and in general support for this more radical approach to reforming the research evaluation system has waned (see Warner, 2000, among others). Partly, this is because institutional support for peer review remains strong. A recent report on the use of bibliometrics as part of research evaluation, commissioned by the Higher Education Funding Council for England, found that "peer review is and has to remain the principal procedure for judgement of quality" (Centre for Science and Technology Studies, University of Leiden (CWTS), 2007: 36). However, there are also conceptual difficulties with an approach based purely on metrics. The strengths and weaknesses of bibliometric analysis will be discussed in depth in Chapter 4, but it is worth highlighting some of the more prominent issues here.

There is now widespread agreement that bibliometric methods have much to offer as a *complement* to conventional peer-review processes for research evaluation with a view to strengthening them (Moed, 2005, among others). When compared to alternative methodologies, bibliometrics offers the particular advantage of using apparently objective data which can be collected with minimal involvement from the researchers themselves. This helps to reduce the administrative burden posed by many of the alternatives – particularly peer review – for researchers, while ensuring that the data gathered is more likely to be representative because it is gathered for entire research communities, not simply those researchers who give their consent (Smith, 1981).[9]

Following Narin's definition in section 2.1.1 above, we have seen that a key advantage of bibliometric analysis is that it can be used at a number of levels to support a variety of analyses, depending on the size and characteristics of the dataset in question. Although viewed initially as a means for building system-level understanding of outputs and impacts in research fields, recent refinements have meant that it is now possible to apply some of the tools of bibliometric analysis to institutions, research groups and even individuals (which will be a key focus of the discussion in Chapter 3).

Combinations of bibliometric analysis and peer review appear to offer greatest potential in overcoming areas of weakness in the peer-review system (Moed, 2005). First, the results of bibliometric analysis may be used to challenge peer reviewers, seeking explanations for unusual or unexpected patterns. Second, they may be particularly useful for high-level or multi-factorial analysis during research evaluation exercises – when the number of factors

---

[8] The RAE is a large-scale exercise undertaken approximately every five years to evaluate the quality of research conducted by higher education institutions in Britain. To date, the RAE has relied largely on peer review of institutions by subject specialist panels, on the basis of which a scaled quality rating is awarded. Previous exercises were carried out in 1986, 1989, 1992, 1996, 2001 and 2008.

[9] There is, of course, usually some involvement from researchers, if only to validate the lists of publications attributed to them that are analysed. However, on balance, the reduction in burden is significant compared to conventional peer review-based approaches, which often require a host of qualitative information collection in addition to details of publications arising from research projects.

under consideration becomes so high that peer reviewers may have difficulty analysing them all in detail, or where some alternative form of validation of peer review decisions may be required.[10]

Although still an area for development, the potential of bibliometrics as a mapping tool is winning greater interest, with a view to supporting longer-term, strategic decision-making. (Cartographic approaches to examining research output are discussed in greater depth in Chapter 3.)

2.3 Some key issues in evaluation: what can bibliometrics measure?

2.3.1 Measuring scientific collaboration

One of the areas in which bibliometrics may offer particularly powerful insights is analysis of collaborative scientific activity: measured typically through the institutional affiliations of authors on publications. This also allows for some comparison of national and international level cooperation, although it should be noted that analysis of collaboration is subject to methodological issues around attribution of research work (a point to which we will return in some depth in Chapter 4).[11]

2.3.2 Assessing interdisciplinary research

The research community has long regarded the assessment of interdisciplinary research as a particular weakness of the peer-review system (Porter and Rossini, 1985, among others). Over the past few years, various bibliometric research groups have experimented with indicator-based approaches which might help to overcome this acknowledged deficit (Cooksey, 2006; Feller, 2006).

Moed (2007) describes two distinct approaches to assessment of interdisciplinary research. First, he describes a forward-looking, raw publication analysis, building a profile of the contribution of a particular research institute or individual across a number of fields in order to assess research impact. Then a measure of multidisciplinarity can be derived from the distribution of the group or individual's papers among fields. Second, an alternative approach involves looking *back* at the fields and disciplines of papers that authors draw information from, as inputs to interdisciplinary research (Moed, 2007). Third, a rougher approach involves looking simply at the fields into which journal publications produced by an author or research group fall.

A further study (Adams et al., 2007b) suggests that there is no need for separate bibliometric measures for interdisciplinarity, since observed patterns are broadly similar between and across research fields. The report authors stress, however, that "it is important

---

[10] This may apply, for example, where peer-review panels are too small to cover the full breadth of research disciplines from which they are asked to assess proposals. In Australia, the Research Quality Framework (RQF), which was until recently the primary means of research quality assessment, employed 13 subject area panels, each of 12 members. For some of these panels, the range of field expertise required was considerable.

[11] In bibliometric analysis, of course, analysis of collaboration is complicated further by the difficulty of linking institutional affiliations to paper authors. Journals do not necessarily provide addresses for all authors; where addresses are provided, it is not always clear to which authors they relate.

… to exercise care in choosing the appropriate field against which to normalise the citations of more interdisciplinary outputs" (2007: 2).

2.3.3 The quest for 'quality' and 'research excellence'

The concepts of 'quality' and 'excellence' in research evaluation are fraught with difficulty. A fuller discussion of some of the conceptual debates in this area is provided in Appendix A to this report, but it is worth highlighting here that there is emerging interest among bibliometricians in developing objective measures for quality and excellence (Noyons et al., 2003). One approach takes into account the skewed nature of citation distributions in order to provide a measure of the number of publications produced by an institution that are in the top 10% of the worldwide citation distribution of the field concerned (CWTS, 2007). This measure has been used to monitor trends in the position of research institutions ands groups at the top levels of their respective fields

Finally, various groups have experimented with measures that allow for top-quality research to be identified.[12] This is particularly important, given recent evidence of an emerging interest in the funding world in increasing focus on 'research excellence' (Tijssen, 2003). As we shall see in due course, this shift has led to the development of a range of bibliometric indicators tailored specifically to identify top-quality research (Van Leeuwen et al., 2003).

2.4 Integrating bibliometric analysis with other performance analysis tools

In Section 2.2.2 it was suggested that the most beneficial use of bibliometrics lies in combining it with other performance analysis tools, including peer review (some prominent examples of integrated approaches are outlined in Appendix B to this report, including the university ranking system adopted by *Times Higher Education*, 2008). More significantly, however, there have been important moves both in Australia and the UK towards integrating bibliometrics into wider research assessment exercises, although this has proved deeply controversial in both cases. Perhaps the Australian example is particularly instructive, and the evolution of research assessment there in recent years is considered in depth below.

2.4.1 Australian Research Quality Framework

The Australian research assessment system underwent major restructuring in 2006 towards a framework based explicitly on assessments of quality and relevance: the Research Quality Framework (RQF). Previously, there had been an overt focus on the *quantity* of academic output, including measures of publication output, the amount of external research income gathered, higher degrees earned relative to the number of research students in any given department, and the rate at which higher research degrees were being completed. The aim

[12] Here, we understand 'top quality' to mean research that is not only methodologically robust and soundly formulated, but that *also* achieves high impact – as measured through journal publications in this case. High quality research may be methodologically very robust and soundly formulated, but may not necessarily achieve high impact. In any event, we acknowledge that the concept of research 'quality' is complex, and may be understood in many different ways. (For a fuller discussion of the concept of 'quality', we refer the reader to Appendix A to this report.)

of the new system was to integrate metrics-based approaches much more fully into the analysis, to which there were three elements:

1. a context statement from each research group, looking at the culture of research in their area and some of the most significant advances;
2. an outline of the full body of work produced by the research group, listing all publications and other forms of analysis;
3. an indication of the best outputs produced by the group, as nominated by the researchers themselves.

The RQF used two bibliometric approaches to support the analysis of data collected: standard analysis, based on recognised journal publications by the group; and non-standard analysis, assessing the impact of book chapters and other publications of this kind. The administrators of the new system recognised the potential for quantitative information to skew judgements on research quality, since it is often more readily assimilated than complex qualitative information, and took measures to try to reduce this effect by, for example, explicitly avoiding the use of composite bibliometric indicators in their analysis of publications (Butler, 2008).

2.4.2 Moving towards Excellence in Research for Australia

From an early stage, it was apparent that there were problems with the RQF approach. First, it was regarded as complex – more so than rival systems used elsewhere such as the RAE in the UK. Second, the RQF focused assessments of research at research group level (as opposed to departmental or university levels): because these units are quite fluid and are not recognised legal or administrative entities, there were concerns that assessment over a six-year cycle (the proposed cycle length for the RQF) would be impossible as group composition and focus changed (Hicks, 2009).

In response to these criticisms, the Australian Government announced in early 2008 that a new research quality and evaluation system would be replacing the RQF. The new system, Excellence in Research for Australia, incorporates a combination of metrics-based research assessment and expert peer review (Hicks, 2009). However, the emphasis on metrics-based approaches in Excellence in Research for Australia is significantly greater than under the RQF, and incorporates a tentative ranking system for journals which has proven to be controversial (Corbyn, 2008). The new system also abandons the previous requirement for groups to identify their best research outputs.

The integration of bibliometric tools as part of wider research evaluation exercises is still in its infancy but seems to offer major promise for the future, especially as the peer-review process is streamlined to improve both efficiency and effectiveness.

2.5 Using bibliometrics to map scientific research

Thus far, performance analysis-based approaches have been by far the most common application of bibliometric methods in research evaluation. This stems partly from the increasingly comprehensive range of indicators available to support analysis, and the ease with which most of these 'objective' measures can be understood by the research evaluation community. However, mapping exercises seek to capture complex inter-relationships and open up interesting opportunities for a more strategic approach to decision-making in research policy over the long term.

Mapping exercises support three broad kinds of high-level analysis. First, they can help to build a sense of activity across the sciences, using journal-to-journal citation maps. This form of analysis builds on links between journals – the key assumption underpinning it is that inter-journal citation frequencies reflect the magnitude of relations between journals (Verbeek et al., 2002). Second, keyword analysis can be used in a variety of ways to build understanding of core areas of scientific output, and to explore the structure of scientific fields and subfields, their evolution over time and the main actors within them. Domains are constructed using combinations of related keywords, and trends are monitored in terms of interactions between sub-domains. Importantly, the maps produced in this way need to be validated by experts in the field. In particular, there are concerns that domain clusters identified on the basis of keywords may not always correspond to the major themes that researchers would identify themselves, and this is where expert validation is particularly useful (Noyons et al., 1999). Finally, citation analysis can be used to support analysis of linkages between fields, domains and sub-domains.

However, mapping exercises can support strategic decision-making in important ways. First, they can be used to assess changes in the activities of research groups or individual researchers over time. Correlated with various input and process variables, this kind of analysis can offer valuable guidance to funders looking to improve funding strategy. Second, they can be used to identify new or emerging fields by correlating changes in volumes of publications and citations over time with changes in levels of financial support for investing in R&D in those fields (Moed, 2005).

2.6 Bibliometrics for research evaluation: measures and indicators

Bibliometrics has come to encompass a vast range of approaches and indicators, and researchers seeking to use it will need to engage carefully with the evidence base around each to understand precisely the kind of information each can provide, and what limitations they may be subject to. In this section, we provide a review of some of the most common approaches and measures, and ways in which they have been used to support a range of research projects. The range of indicators currently in use is outlined in Table 2.2 below.

Table 2.2: Some of the key bibliometric indicators used to support performance analysis

Category of measure/ indicator	Particular measure/ indicator	What is it?	Key advantages	Key disadvantages
Volume measures	Number of publications (P)	The number of publications produced within a specified time period by an individual, research group or institution	• Basic measure of research output • Data is easy to collect	• Crude – gives no indication of impact, only of the level of activity
	Number of citations (C and C_s)	The total number of citations to all papers published by an individual, research group or institution (C); or the total number of citations with *self*-citations removed (C_s)	• Provides basic information on impact • Usually works well on a five-year timescale for the life sciences	• Not adjusted by field or journal set: cross-comparison is difficult, since citation rates in some fields and journal sets are greater than others
Journal impact measures	Journal Impact Factor (JIF)	A measure of research journal quality, based on the number of times that articles within a particular journal are cited by others on average	• Provides a journal-based indicator of impact • Easy to understand and communicate • Data is readily available, whereas data for actual citation counts of articles often has to be purchased from commercial suppliers (particularly large-scale analyses) • Regarded as a timely measure by bibliometricians, since it is based on publications from the most recent two years, and is recalculated annually	• Citation rates between journals and fields vary: comparison is hard • Sometimes criticised for relying on too short a time window for citations to accrue (typically, five years) • Subject to manipulation by editors: e.g. it often includes citations in editorials. It also often includes citations by authors to their own work (self-citations)
Citation-based indicators	Citations Per Paper (CPP)	Measures the average number of citations per paper within a given time period	• Gives indication of average citation performance across a field, organisation or even for an individual	• Does not normalise by field or journal, so it cannot be used to compare within a field or between fields
	Citation rate per paper normalised by journal set (CPP/JCSm)	Adjusts the average number of citations per paper by normalising against the citation rate for other journals in the same journal set.	• Describes whether a researcher, group or institution is above or below the citation averages for the specific journals they publish in • Normalises by adjusting against the average citation rate for the journal set	• Does not provide an indication of the researcher, group or institution's performance *in a field*. This is important because researchers who publish primarily in high-impact journals will be disadvantaged because it will be considerably harder for them to have a CPP/JCSm value much above 1, whereas it is considerably easier for researchers who publish most of their output in low-impact journals. So this measure is not usually used in isolation
	JCSm/FCSm	Provides a comparative measure of journal impact against the field impact	• Provides an indication of the ambition of a researcher, group or institution when it comes to submitting their publications	• Measures potential impact (based on the relative impact of the journals in which the researcher is publishing), rather than the actual impact achieved by the articles

Category of measure/ indicator	Particular measure/ indicator	What is it?	Key advantages	Key disadvantages
	Citation rate per paper normalised by research field. (CPP/FCSm)	Adjusts the citation rate by the citation rates for all papers published in journals in the same field	• Normalises by adjusting against the average citation rate for the field, thus providing a robust comparative measure against other researchers in the same field	
	Highly-cited Papers (HCP)	A measure of excellence based on identification of the top performing papers in a field (can focus on the top 50%, 20%, 10% or 1% of papers)	• Identifies strongly performing individuals or institutions • Timely, since it can be calculated on a year-by-year basis, rather than across a four to five-year window	• Does not reflect performance across the full range of published outputs (many of which will be low impact)
Composite indicators	H-index	A composite measure assessing both the productivity and apparent impact of research papers	• Convenient composite measure for individual researchers	• Works best for high-impact or senior researchers; is a poor indicator for early-career researchers, largely because it is dependent on publication volumes • Cannot deal with issues of attribution to specific grants, programmes, or funders • Discipline norms have not yet been established, so it is not possible to normalise for different citation practices across fields
Distribution-based approaches	Centile distribution	A measure of the distribution of publications across impact bands (i.e. top 50%, 20%, 10%, 1%)	• Can combine HCPs while simultaneously mapping the distribution of all publications from a particular unit under study across impact bands.	

2.6.1 Publication volume measures

The simplest bibliometric indicators are those that measure raw output – often in terms of the number of publications produced by a particular individual, research group or university. In practice, this kind of measure suffers from profound limitations. First, the quantity of outputs produced by an individual or group does not necessarily have a bearing on the impact of their research (Martin, 1996). Consider, for example, the fact that over the past few years, the UK's contribution to the total number of publications produced worldwide has declined in absolute terms; there is no sound evidence to suggest that this has had any detrimental effect on the quality of output; on the contrary, it would appear that the UK is producing fewer uncited papers than in the past (Evidence Ltd, 2007).[13]

Second, while the prospect of measuring the diversity of research output has been suggested, there are important difficulties when trying to make analytical judgements on this basis. Principally, capacity is a key underlying factor which it is difficult to control for, because it multidimensional. It is likely that larger institutions will have more capacity to produce a wider and more diverse range of outputs, so diversity in itself cannot be used as a measure of quality, as it measures only one form of impact.

2.6.2 Journal-linked performance measures

Journal performance measures have been in use since the 1960s, when the most widely-known measure – the Journal Impact Factor (JIF) – was developed. Originally, JIF was developed to provide a sound indicator of the quality of a research journal and, as a consequence, a measure of the likely quality of the papers published within it. It is based on a crude measure of the number of times that articles within a journal are cited by others.

Unfortunately, JIF carries with it important methodological problems. First, citation rates vary considerably between disciplines, making cross-disciplinary comparison difficult. For example, a recent report commissioned by Universities UK found that papers in molecular biology and genetics received an average of 16.15 citations, whereas in engineering – a discipline with very different conventions in terms of dissemination of research findings – typically, papers would be cited only 1.96 times (Evidence Ltd, 2007). Second, citation rates have been found to vary in important ways within broad subject areas – within the broad field of molecular biology and genetics, for example – affecting journal citations in the process, and making clear inferences difficult from this data. Third, there is evidence to suggest that JIF measures may be subject to manipulation by editors through, for example, inclusion of self-citations, or inclusion of articles cited in editorials. This helps to explain ongoing concerns about the way in which JIF measures are calculated and the potential for misrepresentation of analysis results (Moed, 2002).[14]

---

[13] This issue has been exposed to some extent by problems during recent rounds of the RAE in the UK with respect to publication volume per research full-time equivalent. University departments occasionally reduced the number of research full-time equivalents they submitted to the review process, in order to ensure that they achieved a higher 'per full-time equivalent volume'.

[14] However, it is difficult to say with certainty what this all means for the strength of JIFs as indicators of research quality. There is, for example, a strong correlation between JIFs and journal quality as assessed by discipline experts – but of course this comparison is restricted to within-discipline assessments.

In response to some of these concerns, recently the developers of JIF have put forward a new measure, the Journal Performance Indicator, in response to some of these criticisms (Garfield, 2006). Although untested, this measure seems to offer greater promise as a macro-level indicator of impact, although there has been something of a move away from journal-based indicators in bibliometrics over the past decade or so as access to citation data at the article level becomes more widespread (Moed, 2005).

2.6.3 Citation analyses as measures of research impact

Citation analysis is without question the area in which the most extensive range of measures of research performance has been developed. We have reviewed some of the key assumptions underpinning bibliometric analysis previously, but it is worth reiterating these specifically as they apply to citation analysis. The construction of citation measures is underpinned by some important assumptions (Verbeek et al., 2002):

1. citation of a document implies the use of that document by the author in question;
2. citation of a document reflects the merit of that document in terms of the quality, significance or impact of that document;
3. citations are made to the best possible work;
4. a cited document is related in content to the citing document;
5. all citations are equal, although there are field adjustments which can be made to take account of the number of citations that a paper receives relative to others in the same field.[15]

Although robust and comprehensive theory to explain patterns of citation remains elusive despite a number of recent studies investigating citation behaviour (e.g. Hanney et al., 2005; see section 4.3.1 for a fuller discussion), the range of indicative measures of research performance that have been developed is impressive. They include everything from very simple volume measures to complex, field-adjusted indicators which can provide powerful information on research performance relative to others in the field.

Citation volume

The most basic indicator is citation volume (*C*). In some cases self-citations are excluded to give an adjusted measure (C_s), although this practice is not universal. An important question for those using citation volume indicators is what the most appropriate window for data collection is, since this may have a significant skewing effect on results. In the natural and life sciences, it has been found that journal paper citation volume typically peaks in the third or fourth year post-publication, and on this basis a window of five years has been suggested as the most appropriate one for research assessment (Van Leeuwen et al., 1999).

[15] It should be noted that a number of these assumptions are contested. For example, with respect to assumption 3, authors may well cite publications that they in fact criticise (e.g. for methodological reasons) rather than referring to the 'best possible work' (negative citations). Equally, it may be that authors cite a particular paper because it was the most robust document that arose using a particular set of keyword searches, rather than because it is necessarily the most robust piece of work available on the subject.

Citation volume indicators are hampered by very similar limitations to those identified for publication volume earlier in this section. First – as for publication volume – citation volume gives an indication of 'knowledge market share' but not of performance, since it cannot provide us with information about impact other than in the narrowest sense (as it is not normalised in any way). Second, publishing a greater number of papers – irrespective of the quality of the research that they contain – is likely to result in more citations simply because there will be cross-referencing and more papers available to be cited, irrespective of quality. Citation volume indicators cannot provide any meaningful information about top-quality research.

Citations per paper and normalised measures

The average number of citations per paper (CPP, or *c*) is a more advanced measure with the potential for use as an indicator of performance (Van Leeuwen et al., 2003). Typically, this measure includes a normalisation element by adjusting to the field and year in question. This is done partly due to the observation that worldwide citation rates are increasing; it also helps to enable fair comparisons between units of different sizes. One normalised measure involves dividing the average number of citations per paper for, say, an institution, by the average citation rate of all papers in the journal set in which the institution's investigators publish (JCSm). This measure – the citation rate per paper normalised by the journal set (CPP/JCSm) – indicates whether the impact of the institution or individual is above (>1) or below (<1) the international average for those journals in which they publish (CWTS, 2007; Van Leeuwen and Moed, 2002). However, an important problem with this approach is that it disadvantages individuals and units that typically publish in high-impact journals.[16]

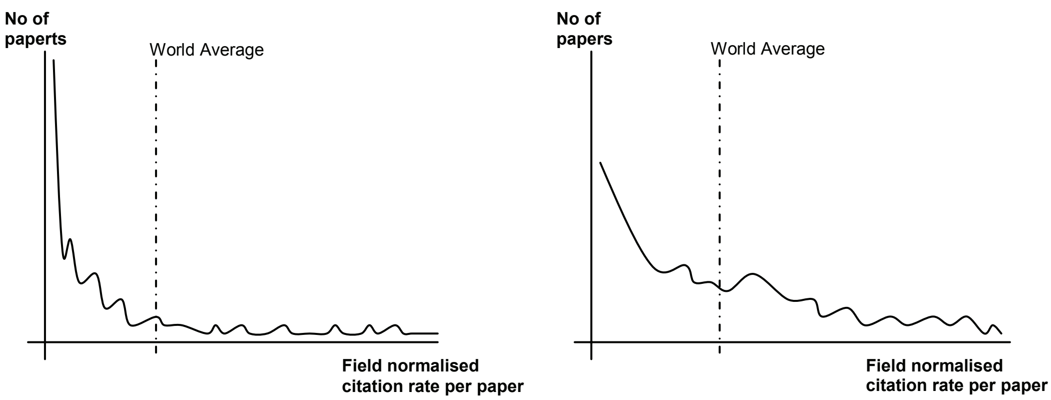

Figure 2.2: Similar world average impact figures disguise different distribution of publication impact

[16] For example, authors and units that regularly publish in journals such as *Nature* will find it difficult to achieve a CPP/JCSm value much above 1, because the average impact of the journal is so high – even though regular publication in a journal of this kind indicates that their research is of very high quality.

An innovation designed to improve the effectiveness of this measure has built in field-specific international references (FCSm), based on the citation rates of all papers published in journals in the fields in which the institution is active.[17] This adjusted measure, the citation rate per paper normalised by research field (CPP/FCSm), allows more discriminating judgements to be made about the standing of the journals in which an institution's researchers typically publish (Van Leeuwen and Moed, 2002; Van Leeuwen et al., 2003). As used by CWTS, this measure allows for a four-level analysis of research performance, whether a research institution performs:

- well below the international average (<0.5);
- below average (0.5–0.8);
- at the mid-range (0.8–1.2);
- above average (1.2–1.5); or
- well above average (>1.5).

Empirical evidence suggests that a CPP/FCSm value above 2 denotes a very strong research group, while a value of 3 or above suggests that a group "can generally be considered to be excellent and comparable with the top groups at the best US universities" (CWTS, 2007: 24). Because of the perceived robustness of CPP/FCSm as a measure of research quality, it is sometimes referred to as a 'crown indicator' in the literature on bibliometrics (Van Leeuwen et al., 2003).

However, because of the extent to which the distribution of publications is skewed towards the low end of the spectrum, and the fact that single, highly-cited publications can have a drastic influence on the citation average,[18] it can be difficult to draw clear conclusions about the quality of publications at either end of the spectrum in a given field (as Figure 2.2 above illustrates) – the average simply provides an indication of where the mid-range may lie.

Highly-cited papers

The clear advantage of the HCP measure is that it acknowledges that the distribution of citations is highly skewed, and that averages can be affected hugely by a single, very highly cited paper. Instead, the focus is overtly on the top performing papers, on the assumption that these are the ones that are likely to have made the greatest contributions to their field (generally the papers in the top 10% in terms of citations, in a field). (In Section 2.3.3 the use of adjusted HCP measures to identify top-level researchers and institutions was discussed.)

While it is a useful indicator of research excellence, HCP cannot give a clear sense of research performance across the board (including the quantity of low-quality research). This means that it is not a particularly useful measure for *post hoc* research evaluation,

[17] Research 'fields' are themselves determined and specified by publication databases such as Thompson Reuters Scientific.

[18] Of course, this will be a particular problem when the publication sample size is small. In a small sample, any HCP is likely to pull the average up much more markedly than for a larger sample.

except where: (1) the aim is to identify outstanding performers; or (2) they are used in tandem with average measures, such as CPP/FCSm (described above).

2.6.4 Composite measures

There is growing research interest in the potential of composite measures as a way of overcoming some of the difficulties inherent in the discrete indicators identified in the previous section. One of the most high profile of these measures is the Hirsch-index (H-index) developed by an academic physicist in 2005 (Hirsch, 2005). Designed for micro-level studies of individual publication activity and citation impact, it is intended to take into account some of the skewing factors that indicators may be subject to at the individual level. First, the key advantages of the H-index include the fact that it is a very simple, composite of a number of forms of analysis which can be applied at a number of levels, but is particularly useful at micro-level: a level at which the reliability of many of the other indicators is most questionable. Second, it normalises for productivity (i.e. raw output) in a way that ensures that researchers are not rewarded simply for the volume of papers that they produce (and by implication, the volume of citations). Finally, it focuses on measuring durable citation performance over a sustained period, rather than single peaks.

On the one hand, several authors have pointed out important disadvantages of this measure. First, the H-index tends to be most effective for high-output and high-impact researchers, and because it is a composite measure, it may not be particularly effective for comparing different dimensions of research performance between individuals (Evidence Ltd, 2007). Second, it does not work particularly well for early-career researchers, who are unlikely to have either a high publication output or particularly high impact. On the other hand, the charge of discrimination against early-career researchers is one that can be levelled against almost all of the currently accepted research evaluation methodologies.[19] Third, because of the H-index's explicit focus on long-term performance, generally it will not highlight declining output over time, and therefore is unlikely to provide a useful indication of the career stage of an individual researcher. Finally – and perhaps most strikingly – the H-index may underprivilege individuals who produce a single, extremely high-impact paper.[20]

2.6.5 Distribution-based approaches

A key theme in this chapter has been that information from bibliometric analysis needs to be treated with caution if it is being used in isolation. Specifically, many of the measures examined thus far represent averages that may not represent accurately the range of values within a given field – we have noted already that citation distributions tend to be skewed towards the low impact end of the spectrum. Therefore several studies assert the importance of considering distributions of bibliometric data in order to overcome the extent to which averages disguise skewed data (Evidence Ltd, 2007). While field-adjusted

[19] This is an accusation commonly levelled against peer-review processes, although in fact the evidence for this is inconclusive (see Jayasinghe et al., 2001, 2003).

[20] Conversely, other measures based on averaged assessments of impact may unduly privilege researchers who have produced only a single or small numbers of high-impact publications. This is an important motivation for using a suite of indicators as part of research performance assessments, rather than relying on individual measures.

values help to overcome some of these problems, distributions offer particular analytical advantages, namely:

- the ability to determine what proportion of the total number of papers are uncited;
- the proportion of papers cited that are below a certain benchmark of impact (e.g. the benchmark for average performance across the UK research establishment);
- the proportion of papers above the benchmark;
- the proportion of papers that are cited at extremely high levels;
- some indication of improvement or decline, as measured by the change in distribution over time;
- enabling comparisons to field medians, rather than means (i.e. above or below the 50th centile benchmark). Many statisticians favour this approach because of the highly skewed nature of the citation distribution.

An indication of the increasing popularity of distribution-based approaches is provided by the fact that they have been included in the Excellence in Research for Australia initiative. Centile distribution measures (see Table 2.2) form an important part of the Excellence in Research for Australia assessment.

CHAPTER 3 Using bibliometrics for selection procedures

Chapter summary

- RAND Europe has undertaken a substantial body of bibliometric work for the English Department of Health over the past few years, in partnership with CWTS.
- The focus of this work has been on *ex ante* evaluation – to help inform selection procedures at the NIHR.
- Specifically, bibliometric analysis has been used to support the selection of appropriate academic institutions as biomedical research centres, academic research departments as biomedical research units, and individuals as faculty members at the NIHR.

This chapter moves away from theory to examine how bibliometrics can and has been used in practice. It does so by reviewing a series of examples of recent RAND Europe project work conducted in partnership with CWTS in the Netherlands, and the Research Evaluation and Policy Project in Australia. Throughout this chapter, the emphasis is on how bibliometric methods can best support *prospective* (i.e. *ex ante*) decision-making in the health sciences.

3.1 Background

As discussed, bibliometric methods have been used primarily to assess the findings of research as part of an *ex post* evaluation process. However, in the last three years, RAND Europe has been working closely with the English Department of Health to bring bibliometric methods to the selection process for research funding distribution, *ex ante*. This has been a complementary process to peer review, with the bibliometric analyses undertaken providing a shortlist of 'high-quality' funding applicants that can then be assessed with complementary peer-review to determine ultimate funding decisions.

Funding for health research can be distributed at different aggregations: from funding large generic centres of excellence and centres for specialist research right through to funding individual researchers. RAND Europe's work with the NIHR – a body created in 2006 as part of the Government's new R&D strategy, *Best Research for Best Health* (Department of

Health, 2006) – has transferred this bibliometric shortlisting method to all of these aggregation levels by modifying the method and tools appropriately to reflect the outputs of each level accurately.

Three brief case study vignettes are provided below. The first and second describe the application of bibliometrics to selecting centres of research excellence. The third describes the application of bibliometrics to rewarding top researchers.

3.2 Selecting centres of excellence (biomedical research centres)

As part of the NIHR suite of funding streams, the Department of Health pledged circa £500 million to support research in biomedical research centres (BRCs) – centres of excellence for medical and health research across a variety of subject areas. NIHR describes the centres, based within the most outstanding NHS and university partnerships in this country, as "leaders in scientific translation … Translating fundamental biomedical research into clinical research that benefits patients" (NIHR, 2006).

BRCs would be comprised of at least one NHS trust and an associated academic partner, where clear collaborative links already existed. Prospective BRCs had to show a strong track record in research in a variety of subject areas (research fields). RAND Europe's role in aiding the selection of BRCs was a two-step process at the pre-application and post-application stages (for a fuller description of the process involved, see Van Leeuwen et al., 2009). The key innovation in RAND Europe's approach (subsequently built on in the further studies discussed below) was the use of bibliometric methods to help reduce the transaction costs involved in a large-scale assessment exercise, which otherwise might have relied more heavily on cost and time-intensive peer review processes – although these still formed an important part of the overall process conducted by the NIHR.

3.2.1 Pre-application

At the pre-application stage, NIHR wished to try to encourage the strongest applications by identifying which institutions in England and Wales were centres of excellence for particular research. To identify these institutions, RAND Europe and CWTS produced a matrix of all university and NHS trusts that perform research in England and Wales, and the 72 health research fields contained in the Institute for Scientific Information (ISI) database. Then, this matrix was populated with the organisation's share of the top 20% most HCPs in health research fields (as identified through addresses on publications). For reasons of visual clarity, the matrix has not been included here; instead, a table identifying those organisations producing the most HCPs in a year, aggregated across all health research fields, are identified in Table 3.1.

Using the matrix, we were able to readily identify those organisations (both academic and NHS) producing a large proportion of HCPs, and the breadth of fields that their research covers. The findings from this stage of the analysis demonstrated that certain organisations produce sizeable proportions of HCPs in a number of research areas, such as University College London – the institutions that one would expect to apply for BRC status as high-quality generalists. Organisations with small proportions of HCPs or with high proportions in only one or two research fields are unlikely to be successful in achieving

BRC status if competing against high-quality generalists, so they would be discouraged from using resources to apply for BRC status.

In addition to identifying organisations with high proportions of HCPs, bibliometric techniques were used to identify the number of HCPs produced by organisations over the period 1995–2001.[21] This allows a simple visualisation of the continuing success of organisations in health research, and to identify which organisations are improving the quality of their research.

To identify which organisations would be likely to provide a strong consortium of academic and NHS partners, RAND Europe and CWTS produced a network analysis of collaborations between different organisations. Organisations were considered to have collaborated where multiple addresses were identified on publications. Figure 3.1shows the collaborations between organisations with 20 or more co-publications. To account for the quality of the co-published research (a key to identifying collaborations that further research), the size of nodes in the figure relates to the number of HCPs produced through co-publication by the organisation. The thickness of connecting lines indicates the strength of the collaborative relationship between the organisation; and the colour of the nodes indicates whether the organisation is academic (blue), NHS (red), or 'other' (green).

Chapter 4 of this report covers the generic strengths and weaknesses of bibliometric methods, and there are specific challenges and caveats associated with using HCPs and network analysis in this work. HCPs provide information on the very best work produced by an organisation, but may not reflect accurately the range of output quality of the organisation. This means that it can mask the quality of the remaining output of the organisation. Network analysis can be difficult to perform where address details are not standardised, but in this case we used data that had been cleaned to identify the location of authors of the publications. It is true that network analysis does not indicate the relative effort made by each author on the publication, but for the purpose of indentifying which organisations work together, this analysis is entirely appropriate.

[21] 2001 was the most recent period for which validated citation information could be obtained from the ISI database, since citations typically take three to four years to accumulate.

Table 3.1: Total number of highly-cited papers in health research fields for candidate biomedical research centres, 1995–2001

Institution	Type of institution	HCP (top 20%)
University College London	University	4650
University of Oxford	University	3398
University of Cambridge	University	3084
King's College London	University	2407
Imperial College of Science, Technology and Medicine	University	2033
University of Manchester	University	1612
Oxford Radcliffe Hospitals NHS Trust	NHS	1367
University of Bristol	University	1314
Guy's and St Thomas's Hospital NHS Trust	NHS	1297
Hammersmith Hospitals NHS Trust	NHS	1281
Cancer Research UK	Research centre	1237
University of Birmingham	University	1222
University of Liverpool	University	1032
Royal Free Hampstead NHS Trust	NHS	1022
University of Newcastle-upon-Tyne	University	1012
Barts and the London NHS Trust	NHS	1005
University of Nottingham	University	924
St George's Healthcare NHS Trust	NHS	895
University College London Hospitals NHS Trust	University	884
University of Sheffield	University	882
University of Leeds	University	831
London School of Hygiene and Tropical Medicine	University	829
University of Leicester	University	790
University of Southampton	University	688
Institute for Cancer Research	Research centre	622
Leeds Teaching Hospitals NHS Trust	NHS	596
National Institute for Medical Research	Research centre	588
Addenbrooke's Hospital NHS Trust	NHS	470
University of York	University	469
St Mary's NHS Trust	NHS	467
Royal Brompton and Harefield Hospitals NHS Trust	NHS	458
King's Consortium	NHS	429
Newcastle-upon-Tyne Hospitals NHS Trust	NHS	428
Central Manchester and Manchester Children's University Hospitals NHS Trust	NHS	425
Royal Marsden NHS Trust	NHS	388
Central Public Health Laboratory	Research centre	372
Southampton University Hospitals NHS Trust	NHS	364
Christie Hospital NHS Trust	NHS	348
Queens Medical Centre, University Hospital Nottingham NHS Trust	NHS	333
Sheffield Teaching Hospitals NHS Trust	NHS	325
University of Reading	University	318
University of Sussex	University	299
Great Ormond Street Hospital NHS Trust	NHS	295
North West London Hospitals NHS Trust	NHS	282
Queen Mary, University of London	University	274
University Hospital Birmingham NHS Trust	NHS	253
South Manchester University Hospitals Trust	NHS	247
University of Bath	University	243
United Bristol Healthcare NHS Trust	NHS	222
Nottingham City Hospitals NHS Trust	NHS	210

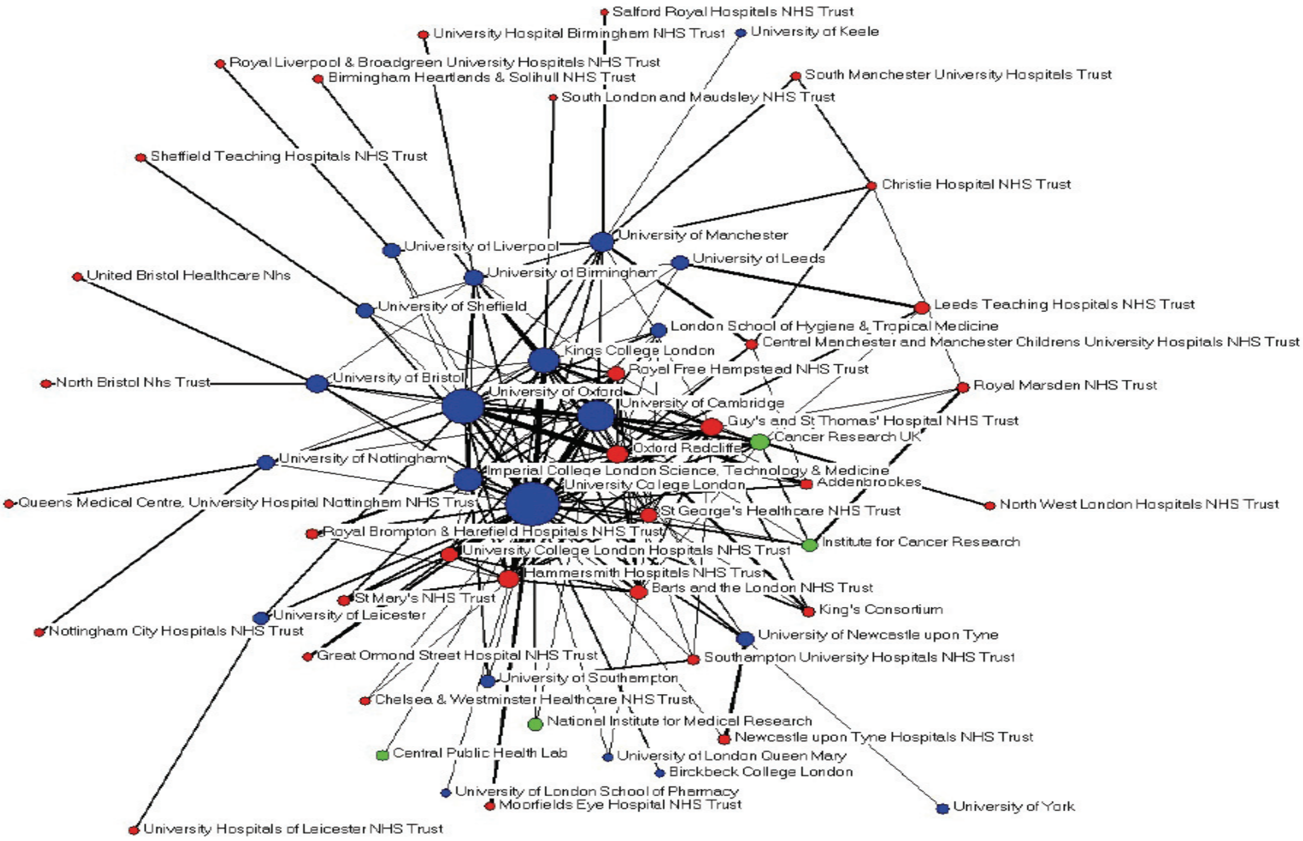

Figure 3.1: Network analysis conducted for candidate biomedical research centres as part of RAND Europe's work for the Department of Health

3.2.2 Post-application

Once prospective centres had applied for BRC status, the NIHR conducted a series of site visits. To support these site visits, RAND Europe and CWTS provided a citation analysis of publications submitted by the applicants. These analyses were conducted upon all submitted publications by the applicant (sorted by research theme) and upon a subset of publications authored by designated 'theme leaders' and the 'centre director' (designated CVs). For each theme, applicants supplied:

- the 10 most important publications specifically from this research area in the past five years' ('themes'); and
- a maximum of five relevant publications for each director and research theme leader ('CVs').

As with the pre-application process, citation analysis was performed using the proportion of submitted publications in the top 20% HCPs in the world. The analysis also produced a value for CPP/FCSm for each theme submitted by the applicant. Values of CPP/FCSm were divided into impact levels:

- CPP/FCSm<0.8 was defined as low impact (less than the world average citations in the field);
- CPP/FCSm: 0.8 to 1.2 was defined as average impact (around the world average citations in the field);
- CPP/FCSm>1.2 was defined as high impact (greater than the world average citations in the field).

BRC selection decisions were performed by an international panel of experts based on a number of criteria, of which citation analysis was one.

3.3 Selecting biomedical research units

Having selected five generalist BRCs and seven specialist BRCs in 2006, NIHR identified eight health research fields that were underrepresented in BRCs:

1. cardiovascular disease;
2. deafness and hearing problems;
3. gastrointestinal disease (including liver, peptic ulcers and dyspepsia);
4. musculoskeletal disease;
5. nutrition, diet and lifestyle (including obesity and blood pressure);
6. respiratory disease;
7. infection (including *Clostridium difficile* and hepatitis C); and
8. pancreatic disease.[22]

To ensure that translational clinical research would take place in these fields, NIHR ran a competition along the lines of the specialist BRCs to provide funding for NHS or university partnerships, in order to achieve biomedical research unit (BRU) status. BRU status brings with it £3.75 million over four years per unit, a significantly smaller sum than the specialist BRCs (which share a pot of £450 million over five years).

As with the BRC competition, RAND Europe provided a two-stage analysis of prospective centres: pre-application identification of research strengths in identified target health research themes, and post-application analysis of submitted publications for each theme.

3.3.1 Pre-application

The pre-application phase of BRU selection mirrored that for BRCs, identifying the proportion of the world's top 20% HCPs held by organisations in the selected research

[22] These are the research areas in which BRUs were selected. At the pre-application stage the list also included microbiology, pathology and stroke as potential BRU health research areas.

areas. This required identifying ISI journal subject categories (JSCs) aligned with specific research areas (Table 3.2).

Table 3.2: Linking research themes to ISI journal subject categories (JSCs)

Research theme	JSCs associated
Cardiovascular disease	Cardiac and cardiovascular systems Critical care medicine
Deafness and hearing problems	Otorhinolaryngology
Gastrointestinal (including liver) disease	Gastroenterology and hepatology
Musculoskeletal disease	Orthopaedics Rheumatology
Respiratory disease	Respiratory system Allergy
Nutrition, diet and lifestyle (including obesity)	Nutrition and dietetics Endocrinology and metabolism Food science and technology
Stroke	Peripheral vascular disease Clinical neurology Neuroimaging Critical care medicine Rehabilitation Geriatrics and gerontology
Pathology	Pathology Medical laboratory technology
Microbiology	Microbiology Biochemistry and molecular biology Biotechnology and applied microbiology Infectious diseases

The individual NHS trusts, universities and other research organisations were assessed for their proportion of top 20% HCPs in the research themes only. A matrix of research themes and organisations highlighted HCP proportions. Unlike the BRC pre-application analysis, it was felt unnecessary to conduct a second network analysis to identify collaboration, since these collaborations were unlikely to have changed much since the previous year's analysis.

3.3.2 Post-application

The pre-application stage highlighted that there was insufficient strength (in bibliometric terms) in the applicant organisations to warrant a BRU for three of the research themes (pathology, stroke and microbiology). Therefore these three areas were ignored in the post-application stage. As with the BRC post-application analysis, the bibliometric information was a contributing factor to the selection decisions, but *not* the deciding factor.

In the analysis 52 applicants were considered, although only 10 to 15 BRUs were likely to be awarded.[23] The aim of the bibliometric analyses conducted was to aid the shortlisting process for potential BRUs. For each applicant's submitted publications, the citation rate per paper adjusted by research field (CPP/FCSm) value was identified, in order to show whether the applicant was producing publications that were cited more often than the world average. If an applicant had a CPP/FCSm value of fewer than one, they were excluded immediately from the shortlist. In addition, those applicants whose value was not statistically significantly greater than the world average were not considered for shortlisting.

[23] In the end, BRU status was awarded to 15 consortia, with five other consortia receiving development grants to support their clinical research activities which support translational research.

In addition to the CPP rate adjusted by research field, the numbers of HCPs produced by each applicant were identified. In contrast with the BRC competition, HCPs were not only considered as the top 20% of publications, but also the top 10%, 5%, 2% and 1% of the world's most HCPs, because the applicants were aiming to be centres of excellence in a specific subject area.

Given the number of applicants for the scheme, it was necessary to reduce the potentially successful BRUs for each theme to a number small enough for site visits by the Department of Health to be practical. For each theme, the number of HCPs for each applicant were compared, and those with the lowest number of HCPs were removed. Using the increasing excellence of HCPs (starting at 20% and working up to 1%), the applicant with the lowest number of HCPs for each research theme was excluded. In the event of a tie (i.e. two or more applicants with the same number of HCPs), the ratio of observed HCPs to expected HCPs[24] was considered, with the lower ratio removed from the process. Performing this process at each level of HCP successively removed one applicant for each research theme at each level, until there were *at least* two applicants left for each theme. In two research themes, there were still relatively large numbers of applicants after this process. To reduce the numbers further, the applicants with the smallest number of the top 1% of HCPs were removed from the process for each theme.

This information fed into the selection process by the Department of Health, which also included site visits and peer review by an independent expert panel.

There are a number of challenges and caveats that need to be taken into account when considering this shortlisting process. First, smaller organisations are disadvantaged by using the number of HCPs in shortlisting. This is a concern, but is not unacceptable, since the aim of funding excellent research organisations is to fund those doing the best work in a research area. Second, it is possible that by submitting a smaller number of only high-quality publications, an applicant can make their impact on the field appear larger than it actually is (i.e. their CPP/FCSm value will be higher than if they include all publications). However, it was concluded that the focus on HCP numbers should resolve this issue, since by reducing the numbers in the analysis, an applicant also reduces their chance of having a larger number of HCPs.[25] Third, it is also true that the smaller the percentage used for HCP analysis, the smaller the number of publications there are to compare. However, with an organisation, this is not a big issue, since there were a large number of publications submitted for analysis.

[24] The expected number of HCPs is the number of HCPs one would expect to see if HCPs were allocated at random to research: thus for the top 20% of publications, one would expect 20% of the number of submitted publications to be HCPs.

[25] Of course, it is possible for researchers to boost their perceived impact artificially through careful selection of papers, if they have conducted an analysis of their publications already to identify which are HCPs. However, we considered that this was highly unlikely to be the case in this application process.

3.4 Selecting top researchers conducting close-to-patient research

Aside from funding collaborations between research organisations, the NIHR is interested in supporting the careers of individual researchers who embody excellence in 'close-to-patient' research. There are a number of initiatives through which this is done, and one of them is the NIHR Faculty. The Faculty consists of researchers in patient-based and applied health research fields, funded by the NIHR or the Department of Health Policy Research Programme. There are five categories of NIHR Faculty: members, investigators, senior investigators, honorary investigators, associates and trainees.

The explicit focus of our work for the NIHR was supporting the process of selecting individuals in one of these categories: senior investigators. A tranche of NIHR senior investigators is selected annually. The award of senior investigator status is meant to recognise the UK's most prominent and accomplished researchers who lead the most prestigious health and social care research projects. The aim of the awards is to be a source of prestige, acknowledge senior researchers and their work, enhance esteem for applied clinical research in the UK, and reward internationally prominent, people-focused health researchers. To date, two rounds of awards have been made: Senior Investigators 1 (SI1) and Senior Investigators 2 (SI2).

3.4.1 Introducing the selection process

RAND Europe was asked to conduct a bibliometric analysis of publications from NIHR senior investigator applicants, for both rounds of awards (SI1 and SI2) which have been made to date. This bibliometric analysis was requested to inform and support the decision-making of a selection committee consisting of a panel of experts convened by the Department of Health, for the specific purpose of identifying those applicants that combine research excellence with the ability to translate their research into benefits for patients and the health and well-being of the public.

The bibliometric analysis provides an indication of the research excellence of individuals. However, bibliometric analysis of individuals does not provide in-depth contextual information on other criteria on which senior investigator awards are made. These include:

- the relevance of the research portfolio to the health of patients and the public;
- the impact of the research on improvements in healthcare and public health; and
- the impact of the individual's leadership on patient and people-based research, including capacity building in research and training.

Bibliometric analysis also brings into play a large number of fields of research which have different publication and citation habits. Hence, results across fields needed to be interpreted and compared with sensitivity to the different bibliometric patterns.

A panel of international health research experts reviewed all the applications and made the ultimate senior investigator award decisions. The panel considered the bibliometric performance of applicants as a proxy for research quality and impacts, then assessed applicants on the additional criteria (as discussed above). The panel assessment scores were averaged for each applicant, and then applicants were ranked according to their mean score.

3.4.2 Gathering suitable data for the selection process

Candidates[26] for NIHR senior investigator status were asked by the Department of Health to submit an application form that requested information on the following:

- the type of research that they do (i.e. fields of research);[27]
- a list of publications from the past 10 years that were relevant to their application;
- information on the relevance and benefits of their research for patients and the wider public;
- evidence of public engagement efforts;
- research leadership roles; and
- other information.

3.4.3 Conducting the bibliometric analysis

Two key bibliometric indicators were used to assist in the profiling and shortlisting of applicants on the basis of the quality of their publications:

1. the 'normalised field score citation indicator' (CPP/FCSm); and
2. the number of HCPs – focusing in this case on the top 10% of publications worldwide.

In the first funding round, the NIHR decided to fund approximately 120 senior investigators; 298 individuals applied for senior investigator positions, and ultimately 99 awards were made. In the second award tranche, 271 individuals applied and the NIHR decided to make between 60 and 70 awards (ultimately 64 were made).

In SI1, bibliometric indicators were used to identify the applicants ranking in the top half of the field for both key bibliometric indicators for further consideration by the panel. In SI2, applicants were ranked into percentiles, identifying candidates in the top 10%, top-quarter, top-third, top-half and bottom-half of the applicant pool for both bibliometric indicators. This was meant to assist the panel in the context of the smaller number of awards being made in SI2.

To aid the panel, RAND Europe and CWTS also provided the following information:

1. volume of publications – to help signal whether applicants were only submitting a proportion of their output to manipulate the citation analysis;

---

[26] Candidates for senior investigator status were identified by the Department of Health in two ways: first, through logs of all institutions, programmes and groups currently receiving NIHR financial support (from which the names of lead investigators were gathered); and second, by placing an open advertisement for the senior investigators scheme on the NIHR website. However, it should be noted that applications arising from the website advertisement were open only to those individuals already in receipt of some form of NIHR financial support for their support, as above.

[27] The field option categories that an applicant could select from were specified on the application form, and specified by the NIHR, based on ISI Web of Science categories.

2. JSCs of publications – to identify whether applicants were applying in the research themes (primary and secondary) that relate to their publication output;
3. the normalised journal citation score indicator (CPP/JCSm), which describes whether researchers score above or below the citation averages for the specific journals in which they publish;
4. the journal impact compared to the field impact (JCSm/FCSm), which provides an indication of the ambition of an applicant when it comes to submitting their publications.[28]

It is reassuring that there is a fairly close correlation between the bibliometrically best performing applicants (those ranking highest), and those ultimately selected by the panel, in both SI1 and SI2 selection rounds (Figures 3.2 and 3.3). Exceptions do exist, and these included applicants for which the bibliometrics was not reliable (e.g. due to low journal coverage of their field in the bibliometric database; a small number of submitted publications being analysed due to factors such as errors in submitted publication codes, document types that cannot be analysed, publications outside a suitable time window for analysis), or where the panel of experts felt that the applicant made significant other impacts (e.g. through research leadership, public engagement) in order to justify selection.

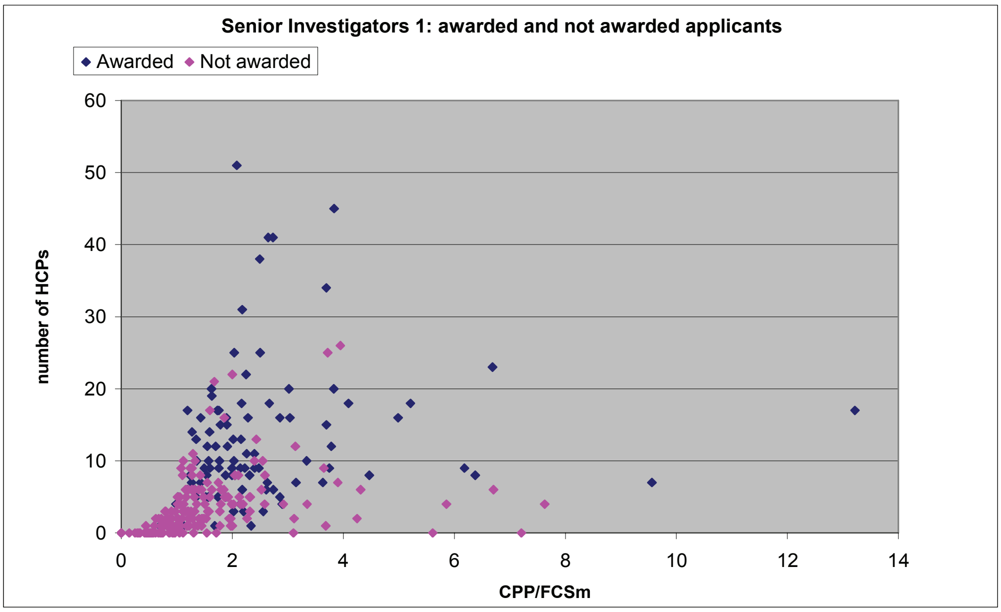

Figure 3.2. SI1: Bibliometric rankings of awarded and not awarded applicants

[28] For example, if a researcher's JCSm/FCSm score is above 1, this indicates that they target journals with above field average impact – i.e. they target top-flight journals and can be considered to have high ambitions.

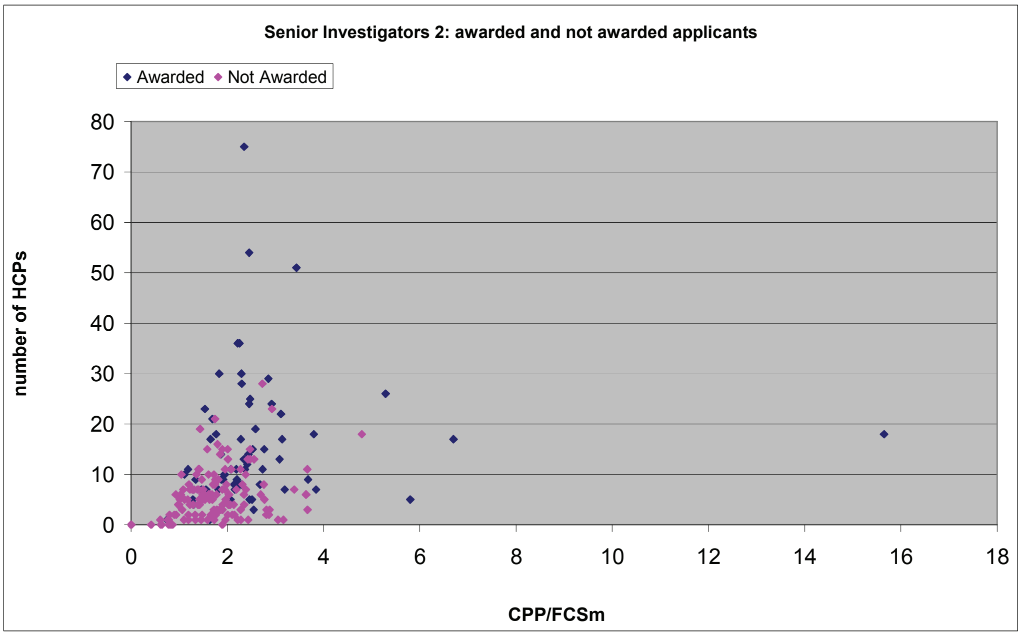

Figure 3.3. SI2: Bibliometric rankings of awarded and not awarded applicants

In addition, the 'research quality/impacts' were compared (as proxied by citations per paper normalised by field, and by the number of HCPs) between applicants in the two rounds (SI1 and SI2), in order to assist the Department of Health in assessing whether standards are being maintained over time. Figure 3.4 shows that the selected senior investigators in both rounds had similar bibliometric performance.

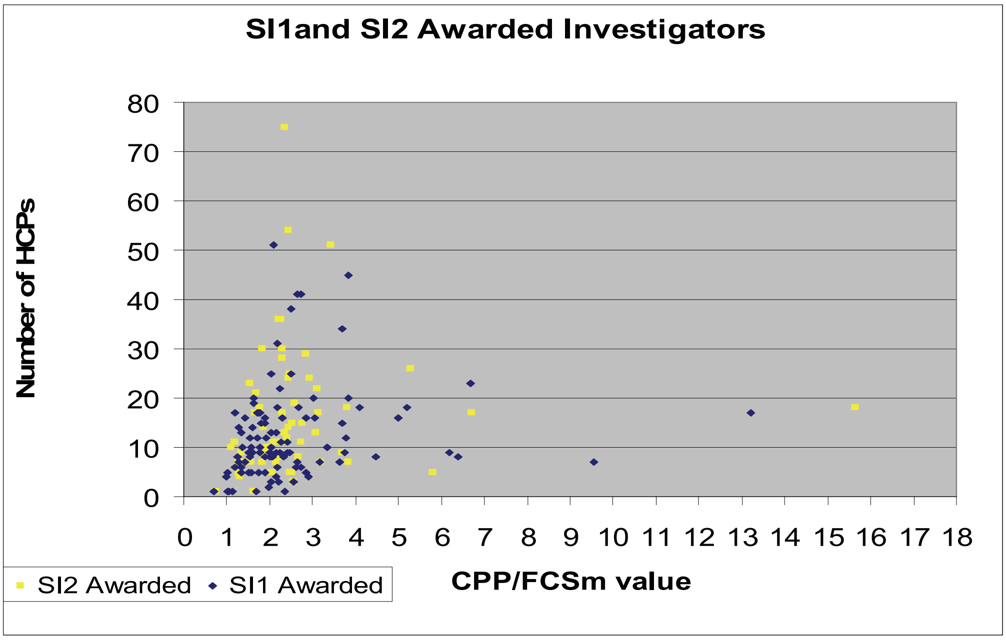

Figure 3.4. Comparison of bibliometric performance of awarded applicants in SI1 and SI2

CHAPTER 4 Assessing bibliometric techniques: strengths, weaknesses and caveats

Chapter summary

- There are considerable advantages to a bibliometric approach, especially as the power and range of indicators available improves. However, a clear understanding of limitations and caveats is required.
- From a theoretical perspective, some doubts remain as to the ability of bibliometric methods to capture abstract concepts such as research 'quality' – but these apply equally to accepted approaches, including peer review.
- Methodological challenges include issues of journal coverage in major bibliometric databases, adequately identifying author affiliations and choosing the right timeframe for analysis.
- Caveats include variations in citation behaviour between fields and individuals – although these can be offset to some extent by ensuring appropriate publication sample sizes – and a perennial difficulty in evaluation: attribution.

Bibliometric approaches have strengths and weaknesses, many of which have been described in detail elsewhere (see for example, Research Evaluation and Policy Project, 2005). Some of these relate to changes in the way in which scientific knowledge is disseminated: over the past 10 to 15 years, the primacy of print journal articles in some fields has been challenged increasingly by new types of publication behaviour, notably electronic publishing (Van Raan, 2001) and open-access journals (Harnad et al., 2003).[29] Others relate to precisely what kind of impact these approaches can measure; clearly, assessments of the socio-economic impact of research would draw more profitably on other approaches, although bibliometrics may still be used as part of the overall analysis. At a more fundamental level, there is the question of precisely what kind of measure of intellectual 'influence' bibliometrics truly provides (CWTS, 2007).

---

[29] It should be noted that electronic-only and open-access journals *can* be used in bibliometric analysis; our contention is that as publishing through non-indexed scholarly dissemination channels (e.g. non-peer-reviewed working papers issued through institutional repositories or by self-publishing) increases, bibliometricians will face growing challenges in analysing the spread of scientific knowledge.

Broadly speaking, assessments of the strengths and weaknesses of bibliometrics fall into three categories (UK Evaluation Forum, 2006): theoretical issues; those that explore general methodological challenges; and caveats to bibliometric analysis. In the discussion that follows, key criticisms of bibliometric methods from the research literature are described and evaluated. However, this chapter begins with a discussion of some of the epistemological difficulties with bibliometric analysis, and the kind of information that it generates.

4.1 Theoretical issues

4.1.1 Measuring research 'quality' probably requires the use of suites of indicators

Donovan (2007b) highlights a number of stock criticisms to which bibliometric analysis often is subjected, the most damning of which is that indicators do not actually measure research quality. While indicators may provide useful indications of impact, there is no direct correspondence between this and research quality (a complex and multifaceted concept that is discussed in greater depth in Appendix A to this report). A key issue revolves around the fact that the concepts of 'citation impact', 'intellectual influence' and broader notions including 'quality' and 'excellence' do not necessarily coincide. 'Intellectual influence' for example, is an essentially theoretical, qualitative construct, which can be assessed only by taking into account the cognitive content of the work under consideration (Moed, 2007).

As we have seen, critics argue that proper evaluation of scientific research requires a broader, qualitative understanding of the content of the research. However, the evidence for this contention is uncertain: some studies suggest that in fact correlations between peer-review judgements and bibliometric measures are quite strong (Rinia et al., 1998).

Moreover, the use of *suites* of bibliometric indicators (rather than individual measures) offers real potential in overcoming some of the challenges associated with quality measurement. By focusing attention on a *range* of bibliometric measures (embodied by the various indicators described in Chapter 2) rather than single measures, it is possible to reflect a number of aspects of research impact, and thereby move closer to robust measures of 'quality' (Moed, 2005).

4.1.2 Bias against early-career researchers is a concern, but the evidence is equivocal

Bibliometrics depends heavily on an established track record of publication in order for analysis to be effective. As mentioned previously, a key problem with this approach is that it *may* discriminate against early-career researchers who do not have a substantial body of publications to their name, unless adjusted measures are used. Similar accusations have been levelled against grant peer-review systems in the past (see for example, Bazeley, 2003), although importantly, no clear evidence for this is forthcoming.

This problem can be reduced significantly by shortening the time windows from analysis (e.g. down to a single year from the usual five). Reducing the time window ensures that early-career researchers without a significant number of older papers are not disadvantaged. This approach has been proposed as part of the methodology for Excellence in Research for Australia. However, it is important to remember that as publication sample sizes decrease, the likelihood of errors of analysis increase, and judgements of impact for early-career researchers are less likely to be reliable than for those with large portfolios. This applies even where field and publication volume-adjusted measures, such as CPP/FCSm, are used.

4.2 Methodological issues

This section reviews some of the issues to consider when constructing methodologically robust bibliometric analyses. Much fuller accounts are provided elsewhere (e.g. Moed, 2005).

4.2.1 Gathering accurate publications information is time-consuming

From a practical perspective, there are challenges with collecting accurate publications information. First, while a number of databases collect basic information on scientific publications (in the medical research field, Medline, PubMed and Web of Science are notable examples), a large number of the articles included are not available in full-text format. This means that gathering key publication data, including details such as author addresses and funding acknowledgements, can be difficult and time-consuming.

Second, there is no standard way for researchers to acknowledge their funders, resulting in inconsistencies between journals and research fields. This is a challenge for attribution in research evaluations. Some research has shown that only around two-thirds of biomedical publications acknowledge a funding source; clinical papers were found to be much less likely to cite a funding source than basic ones (Webster et al., 2004; Wellcome Trust, 1998). However, bibliometric investigations of specific fields have tended to return more encouraging results: a study of acknowledgement patterns by rheumatology researchers, for example, found that at least one funding source was acknowledged per paper on average (Lewison and Devey, 1999).

Finally, for papers with a large number of authors, there is no clear way of distinguishing relative contributions made by each individual, an observation that hints at one of the key caveats to any bibliometric analysis: the issue of attribution. In practical terms, citations on papers are attributed typically to all named authors during evaluations.

4.2.2 Database journal coverage is variable, but improving

Similarly, the databases currently available to support bibliometric analysis, their coverage and broad trends in the data that they hold are open to challenge. The two major indexes of publications currently available internationally are provided by Thompson Reuters Scientific and Scopus. The Thomson Reuters Scientific database covers some 9,000 of the most important journal titles, mostly selected on the basis of their citation impact.[30] Scopus covers 15,400 journal titles across the natural and social sciences, art and literature.[31]

---

[30] For further details of Thomson Reuters Scientific's coverage, see Testa (2009).

[31] For further details of Scopus' coverage, see Scopus (2008).

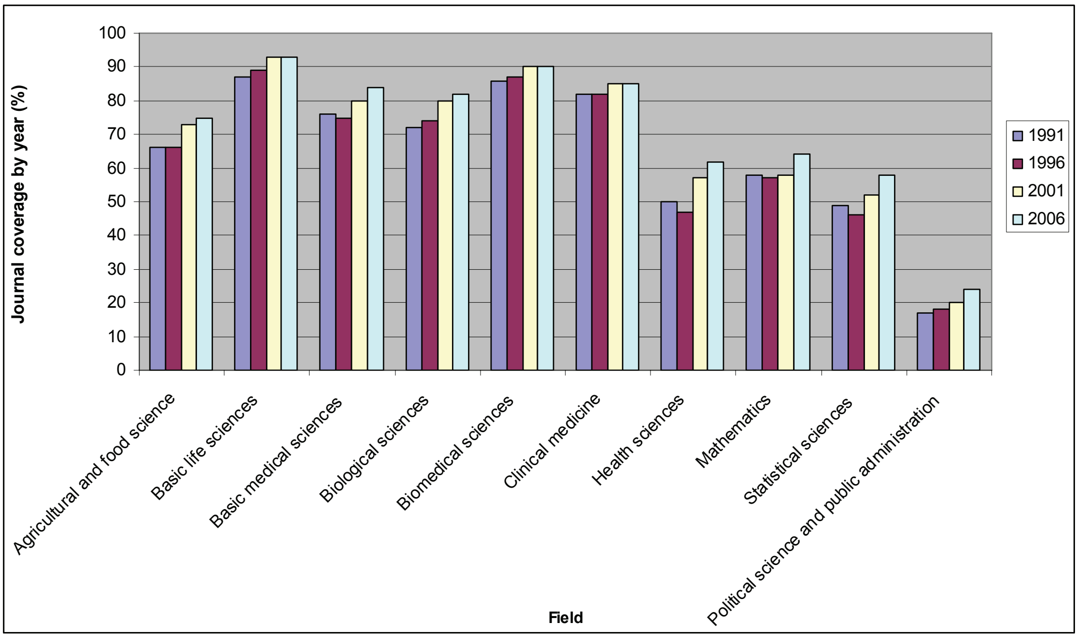

Figure 4.1: Comparison of changes in journal coverage by year, in the Thomson Reuters Scientific database

Source: Adapted from CWTS (2007)

Evidence also suggests that journal coverage is improving with time – as the data in Figure 4.1, covering a range of research fields, illustrate. Coverage in disciplines outside the core natural sciences can be quite patchy, and various studies have been undertaken over the past few years to assess quite how representative findings derived from them may be, and with variable results (consider, for example, Van Leeuwen, 2006). Specifically, a number of authors have highlighted the challenges presented by poor journal coverage in the arts and humanities relative to the natural and life sciences (Glanzel, 1996; Hicks, 1999; Van Leeuwen, 2006). In the health sciences, there are acknowledged deficiencies in coverage for health service research and publications aimed at allied health professionals. Internationally, coverage is also a particular issue because English-language journals are predominant in Thompson Reuters Scientific journal set, although Scopus specifically aims to improve international coverage.[32] However, the impact of this bias is offset to some extent by the fact that English is now broadly accepted as the international language of scientific research (King, 2004).

4.2.3 Identifying authors and their affiliations can be difficult

Particular forms of bibliometric analysis – particularly those that assess the quality of research produced by research institutions and individuals – depend on an ability to determine the identity and institutional affiliation of the author accurately. This presents challenges. First, uniquely identifying authors can be difficult where they have common names. Second, the way in which author names and affiliations are recorded can have important consequences: although journals normally note link addresses to specific

[32] There are particular issues of bias in coverage against countries such as France, China and Japan, with a high proportion of native language journals.

authors, this information is not always recorded in bibliometric databases.[33] Some disciplines follow prescribed patterns with respect to author affiliation; in others this is much less clear (Verbeek et al., 2002). Often one can access only the institutional affiliations of a first or last author. When attributing publications to a specific grant or funder, one needs to know that the institutional affiliation of an author corresponds accurately to their location at the time of a grant, or the funder of a project needs to be acknowledged – this is one issue of attribution and another key challenge in research evaluation efforts.

4.2.4 Citation windows need to be chosen carefully

Selecting citation windows for an evaluation study may have an important impact on subsequent findings. We have seen that in the natural and life sciences, journal publications are estimated to hit peak citation rate in the third or fourth year post-publication on average, but that there are important differences between fields (Van Leeuwen et al., 1999). Moreover, some have expressed concern that this approach risks overlooking 'mayflies' and 'sleeping beauties', and that in fact a longer-term approach to citation analysis may be necessary. In practice, the five-year time window represents something of a compromise between the need to set defined timescales for short-term evaluations, and the recognition that sometimes, important long-term patterns may be missed. Furthermore, studies suggest that delayed recognition (in the form of 'sleeping beauties') is a rare phenomenon: one investigation found that only 0.3% of papers published in 1980 and not cited during the period 1980–85 had received more than 15 citations in total by 2000 (Glänzel et al., 2003).

4.3 Caveats

The validity of judgements based on bibliometric information is a key issue. It is widely acknowledged that researchers cite papers for very different reasons (this is discussed in more depth in the next section). With this in mind, there are important qualifications to be borne in mind when assessing the validity of bibliometric analysis: specifically, that citation analysis measures the impact of articles on other researchers in a given field (and/or outside it), which reflects only one dimension of research quality. A robust definition of 'quality' in a research evaluation context has long eluded specialists in this area, and there is growing consensus that any judgement of quality will need to be based on a combination of qualitative and quantitative analyses – in other words, bibliometrics should be viewed as one element of a wider process of review (Moed, 2005). In addition (as will be discussed in more detail later in this section), citation behaviour is complex and not an 'ideal' monitor of scientific performance. This is particularly the case at low levels of aggregation (e.g. individual researcher level), where there may be considerable variation; one would expect some of this effect to be evened out at population level.

However, the available evidence does suggest a good correlation between performance assessments carried out on the basis of bibliometric analyses, and the results of peer-review

---

[33] This reflects an importance difference between the Web of Science and Scopus databases: the former does not formally record links between addresses and specific authors; the latter does.

processes (Rinia et al., 1998). Furthermore, the skewing effects of some of the phenomena listed above can be offset largely by selecting appropriately-sized publication samples. This section reviews some overarching issues that need to be borne in mind when drawing conclusions from bibliometric data.

4.3.1 Citation behaviour is variable, and can be explained in many different ways

Citation analysis is grounded on the premise that tracing the reference practices of scientists provides us with useful information. Unfortunately, there is no strong, *a priori* theory to explain why authors cite in the way that they do (Vinkler, 1998). A number of studies have included attempts to develop classification systems to help describe the strength of the impact that a cited work has made on a paper citing it, or to explain authors' motivations for citing papers (see for example, Hanney et al., 2005; Small, 1982). While the classification systems developed have shown promise, it is clear that further development work is needed to understand better author motivations for citing specific work. In any event, Eugene Garfield's (2006) recent study on this subject demonstrates that some motivations underpinning citation behaviour are more 'honourable' than others.

1. Paying 'homage' to pioneers.
2. Giving credit for related work.
3. Identifying methodology, equipment and so forth.
4. Providing background reading to support a particular study.
5. Correcting one's own work.
6. Correcting the work of others.
7. Criticising previous work.
8. Substantiating claims.
9. Alerting readers to forthcoming work and publications.
10. Providing leads to poorly disseminated, poorly indexed or unquoted work.
11. Identifying original publications in which a concept or idea was discussed.
12. Disclaiming work or the ideas of others.
13. Disputing the prior claims of others in previous publications.

Box: Some of the most common reasons put forward for citing published papers (adapted from Garfield, 2006)

Some of these issues pose significant challenges. First, in an age in which the use of online media is becoming increasingly widespread, citations to multiple versions of the same document may introduce skewing effects into analysis (Evidence Ltd, 2007); also, it may be that these documents are accessed preferentially because they are more readily available. Second, various studies have suggested that behavioural trends, such as the tendency to preferentially cite from a researcher's own country,[34] research group or department, or the

[34] Explanations for this tendency to cite research generated in the author's country of origin are not immediately forthcoming; this remains an area for further investigation.

tendency to self-cite, can have important effects (Evidence Ltd, 2007). Third, some papers are cited because they are wrong or their methodology has been found to be flawed – that is, as unfavourable citations – in this case, bibliometric analysis would seem to suggest that these papers are 'high impact' when in fact they may be cited as examples of poor research practice. However, on balance, the fraction of papers to which this applies is almost certainly very small (Evidence Ltd, 2007). Finally, some authors 'implicitly' cite papers or (more often) bodies of work in their own papers as part of background discussions, without ever explicitly naming them (Verbeek et al., 2002).

Some caution is advisable here. Van Raan offers a forceful defence of bibliometrics in the face of some of these criticisms, pointing out that effects arising from variable citation behaviour are very unlikely to be statistically significant:

> [N]obody can seriously maintain that the references in, for instance, this paper are totally unreasonable and completely arbitrary ….valid patterns in citations will be detected if a sufficiently large number of papers is used for analysis. Furthermore, it is statistically very improbable that all researchers in a field share the same distinct reference-biases. (1998: 134–5)

4.3.2 Fields can be difficult to define, especially for interdisciplinary research

Differences between research fields exert important influences over the kind of analysis which can be performed. In the applied and engineering sciences, for example, paper publication is not the primary means for disseminating research findings, and therefore bibliometrics is very difficult to apply. Interdisciplinary work is an enduring issue, although, as we have seen, work is underway to address this issue through the development of improved and/or specific indicators. Partly, this difficulty stems from problems over field classification for interdisciplinary research: in other words, it is not always immediately obvious under which fields interdisciplinary publications should be classified. Whatever the reason, the efficacy of bibliometric analysis in terms of assessing interdisciplinary research is in question. On the one hand, a number of studies have found significant discrepancies between the scores obtained by interdisciplinary research outputs at peer review and bibliometric analysis (Porter and Rossini, 1985; Rinia et al., 1998). On the other, more recent work has appeared to undermine this contention (Rinia et al., 2001).

4.3.3 Attribution is an enduring problem in research evaluation

The efficacy of bibliometrics as a research evaluation methodology depends to a large extent on the accuracy with which scientific publications can be attributed to particular authors, institutions, grants and so forth. Unfortunately, clearly attributing papers is often problematic (Verbeek et al., 2002). First, pieces of research (and the papers that emerge from them) are often the product of work by a number of researchers and funders. Second, researchers may hold a number of awards concurrently, making it difficult to define clearly the block of a funding to which a research paper may be attributed. Third, medical research increasingly involves collaboration between researchers and institutions across disciplines. Since papers emerging from collaborative research often feature a number of researchers funded by different bodies, it can be difficult to identify which elements of the research can be attributed to particular funders (UK Evaluation Forum, 2006). Lastly, funders are not always acknowledged on papers.

4.3.4 Researchers occasionally cite incorrectly

In some cases, authors simply cite papers or bodies of work incorrectly (Verbeek et al., 2002). In practice, the impact of this sort of error is small, especially when averaged across a field.

4.3.5 Adequate sample sizes can help to offset many of these effects

As with any method grounded in statistical analysis, due consideration for the size of the publications sample in bibliometrics is important, although often this is surprisingly neglected in the literature on bibliometrics (Yoshikane et al., 2003). In general terms, the analysis can be considered more robust, the larger the sample size. Many of the caveats described later in this chapter (such as citation of mistakes and the skewing influence of single, very highly cited papers) can be overcome by ensuring that the sample is sufficiently large.

4.4 Some broader issues

4.4.1 Impact of bibliometric evaluation processes on researcher behaviour

Sometimes, individual behaviour can introduce important biases in bibliometric analysis. Recently this issue has been raised in the context of the impact on researcher activity of the move towards bibliometrics for research evaluation. For example, it has been noted that the move towards increased used of metrics for research evaluation in some countries – the Netherlands being a notable example – has encouraged some researchers to publish more frequently and with more co-authors, but capturing smaller units of research each time, in order to help boost their output rate (and with it the potential for further citation) (Weingart, 2005).

In the UK, a longitudinal analysis of the citation patterns for a sample of researchers showed unexpected changes in behaviour in the run-up to the RAEs in 1992, 1996 and 2001. In the run-up to RAE 1992, for example, UK scientists appeared to increase their journal article output significantly in response to a new request that year for total publication counts. Similarly, the shift in the RAE 1996 away from quantity-based measures towards measures of quality seems to have resulted in an increased tendency of researchers to publish articles in higher-impact journals (Moed, 2008).

Both studies highlight the importance of taking strategic behaviour into account when conducting assessments. They also suggest that the indicators used to measure impact need to be sophisticated enough to be able to overcome some of these skewing behaviours. Indeed, in recent years this has been an important driver for the move away from pure publication counts and journal impact measures towards assessment systems based more on citation impact.

4.5 Need to focus on research quality in bibliometric analysis

One of the most trenchant critics of the rising prevalence of bibliometric analysis in research evaluation has been Peter Weingart (2005, among other publications). The core of his critique has been an accusation that the increasing demand for rankings and simple measures of research outputs and outcomes has obscured the very real problems that exist with many bibliometric indicators, suggesting a greater degree of rigour in this field than is actually the case. Weingart has suggested a refocusing of the agenda with respect to the use of bibliometric indicators towards greater emphasis on quality of publications rather than volume, and a recognition that drawing the correct judgements from quite complex forms of analysis requires a high level of expertise.

CHAPTER 5 Options for the future

Chapter summary

- Further development in some discrete areas could strengthen significantly the analytical power of *ex ante* bibliometric assessments in discrete areas.
- By investigating the linkages between publication and citation patterns and the economic impacts of individual researchers, groups or institutions, it may be possible to inform strategic funding decisions to maximise economic returns.
- New and robust indicators may be developed to identify up-and-coming researchers.
- Cross-checking systems may help to identify the small number of researchers, groups and institutions who effectively ‘play the system’.
- Further work may include developing indicators of close-to-patient work for funders looking to maximise the impact of their funding on health outcomes.

The core focus of this report has been on the *practical* uses of bibliometric techniques to support prospective R&D decision-making in the health sciences. This conclusion reviews some options for future development of bibliometric methods as an aid to prospective decision-making, focused around a series of questions which have emerged in the course of our work to date. These options do not include improvements in theoretical understanding or major changes to methodologies – for these we direct the reader to outputs from key academic centres working in this area.

5.1 How do I fund researchers to maximise economic impact?

A recently completed, major research project on the economic returns from UK-funded health and biomedical research demonstrated substantial impact in the fields of cardiovascular and mental health research (Health Economics Research Group et al., 2008). This *post hoc* evaluation examined economic impact at an aggregate level across the UK research system, and a key area of development for the future will be to identify ways in which the likely future economic impact of researchers and research projects might be anticipated, *ex ante*, perhaps by investigating the links between publication and citation patterns and economic returns.

5.2 How do I identify up-and-coming researchers?

Given concerns about the support received by early-career researchers under current funding review processes, an important area of development for the future will be to improve ways in which promising up-and-coming researchers are identified. Work conducted on behalf of NIHR by RAND Europe and CWTS on selecting individuals has included attempts to develop viable indicators of promise.

A crude measure may involve looking at publication output per year: this gives an indication of the volume published per year and success in terms of getting journal articles into print. However, as with all volume-based indicators, it can give no sense of quality, which may fluctuate from year to year and from publication to publication. Moreover, output counts are only likely to be helpful or valid *within* fields, rather than between them, unless they are appropriately normalised.

Looking at average citations per paper in a specific journal a research publishes in, and normlising that against the average citations per paper for a particular field (which the journal belongs to) may provide a good indicator for up-and-coming researchers to measure their level of ambition. Young researchers that achieve a high acceptance rate in high-impact journals are likely to have high levels of ambition and possibly high future potential. Future work in this area will need to focus on developing improved indicators of promise.

5.3 How do I identify researchers who are 'gaming' the system?

This report has noted already an emerging strand in the literature on bibliometrics highlighting the problem that research evaluation systems may help to create perverse incentives. For example, a growing tendency has been observed for researchers to split publishable work up into increasingly smaller units in a bid to boost journal article productivity (Weingart, 2005). In Australia, there is evidence to suggest that changes to funding formulas in the late 1990s and early 2000s, which were designed to reward better quality academic output, actually rewarded quantity (Butler, 2003).

Therefore, future work will need to turn some attention to improved methods for detecting researchers and institutions that 'game' the system. For publication data, this can be especially difficult, but nevertheless it may be possible to introduce cross-checks in order to minimise the impact of gaming. For example, in RAND Europe's recent work on selecting individuals for the NIHR's Faculty, it was possible to check the investigators' choice of their primary and secondary research fields against those subject fields in which their publications were classified in bibliometric databases. This helped to ensure that researchers were not selecting generally under-represented fields as their primary or secondary field as a way of securing funding from under-resourced areas. Future work will need to focus on developing and improving existing cross-checking mechanisms.

5.4 How do I identify researchers who are doing close-to-patient work?

An important area for future work will be developing new and improved methods for identifying those researchers engaged in close-to-patient work. This is a particularly important consideration, given the attention given to bolstering translational research work in the recent Cooksey Review (2006). Historically, identifying researchers working in these areas has proven to be extremely difficult from a bibliometric perspective, partly due to problems with identifying the publications that are read most frequently by those working with patients on a day-to-day basis – especially for the allied health professions.

One approach – attempted during past work by RAND Europe and CWTS – has been to try to construct a 'close-to-patient' journal set featuring those publications read most often by practitioners working with patients on a regular basis. Through contacts with colleges and professional associations, RAND Europe researchers established that practitioners were more likely to read general summaries of key research findings provided in bulletins issued by representative bodies, rather than consulting journals directly – largely because of time pressure. Exceptions to this general pattern included major practitioner-focused journals such as the *New England Journal of Medicine*, *British Medical Journal* and *The Lancet*, and RAND Europe used these journals to form a small, close-to-patient journal set as the basis for bibliometric analysis. Unfortunately, this journal set proved too small to be effective in the study in question. Future work will need to address new ways of analysing publications for evidence of close-to-patient work.

REFERENCES

References

Adams, J. (2005) 'Early citation counts correlate with accumulated impact', *Scientometrics 63*(3): 567–81.

Adams, J., Gurney, K. and Marshall, S. (2007a) 'Profiling citation impact: a new methodology', *Scientometrics 72*(2): 325–44.

Adams, J., Jackson, L. and Marshall, S. (2007b) Bibliometric analysis of interdisciplinary research. Report to the Higher Education Funding Council for England. Leeds: Evidence Ltd.

Auditor General of Canada (1994) *Report of the Auditor General. Chapter 10: science and technology; management of departmental science and technology activity*. Ottawa: Office of the Auditor General of Canada.

Bazeley, P. (2003) 'Defining "early career" in research', *Higher Education 45*(3): 257–79.

Butler, L. (2003) 'Modifying publication practices in response to funding formulas', *Research Evaluation 17*(1): 39–46.

Butler, L. (2007) 'Assessing university research: a plea for a balanced approach', *Science and Public Policy 34*(8): 565–74.

Butler, L. (2008) 'Using a balanced approach to bibliometrics: quantitative performance measures in the Australian Research Quality Framework', *Ethics in Science and Environmental Politics 8*(1): 83-92. Cole, J.R. and Cole, S. (1973) *Social stratification in science*. Chicago, IL: University of Chicago Press.

Cooksey, D. (2006) *A review of UK health research funding*. London: HMSO.

Corbyn, Z. (2008) 'Philosophers in ranking protest', *Times Higher Education*, 28 August. [Accessed 18th November 2009: http://www.timeshighereducation.co.uk/story.asp?storyCode=403321&sectioncode=26]

Centre for Science and Technology Studies, University of Leiden (CWTS) (2007) 'Scoping study on the use of bibliometric analysis to measure the quality of research in UK higher education institutions', report for the Higher Education Funding Council for England. Leiden: University of Leiden.

Council on Health Research for Development (2000) *Working group on research to action and policy. Lessons in research to action and policy: case studies from seven countries*. Geneva: Council on Health Research for Development.

Department of Health (2006) 'Best research for best health: a new national health research strategy', 25 January. [Accessed 20 July 2009: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4127127]

Donovan, C. (2007a) 'Introduction: future pathways for science policy and research assessment: metrics vs peer review, quality vs impact', *Science and Public Policy 34*(8): 538–42.

Donovan, C. (2007b) 'The qualitative future of research evaluation', *Science and Public Policy 34*(8): 585–98.

Evidence Ltd. (2007) *The use of bibliometrics to measure research quality in UK higher education institutions*. London: Universities UK.

Feller, I. (2006) 'Multiple actors, multiple settings, multiple criteria: issues in assessing interdisciplinary research', *Research Evaluation 15*(1): 5–16.

Garfield, E. (2006) 'The history and meaning of the journal impact factor', *Journal of the American Medical Association 295*(1): 90–3.

Glänzel, W. (1996) 'A bibliometric approach to social sciences: national research performances in 6 selected social science areas, 1990–1992', *Scientometrics 35*(3): 291–307.

Glänzel, W (nd) 'On the opportunities and limitations of the H-index'. [Accessed 28 October 2008: http://eprints.rclis.org/archive/00009535/01/H_Index_opprtunities.pdf]

Glänzel, W., Schlemmer, B. and Thijs, B. (2003) 'Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon', *Scientometrics 58*(3): 571–86.

Godin, B. (2002) 'Outline for a history of science measurement', *Science, Technology and Human Values 27*(1): 3–27.

Grant, J., Hanney, S. and Buxton, M. (2004) 'Academic medicine: time for reinvention. Research needs researching', *British Medical Journal 328*(7430): 48.

Hanney, S.S., Frame, I., Grant, J., Buxton, M., Young, T. and Lewison, G. (2005) 'Using categorisations of citations when assessing the outcomes from health research', *Scientometrics 65*(3): 357–79.

Harnad, S., Harnad, S., Carr, L., Brody, T. and Oppenheim, C. (2003) 'Mandated online RAE CVs linked to university e-print archives: improving the UL Research Assessment Exercise while making it easier and cheaper', *Ariadne 35*: April.

Health Economics Research Group, Office of Health Economics, RAND Europe (2008) *Medical Research: What's it worth? Estimating the economic benefits from medical research in the UK*. London: UK Evaluation Forum.

Hicks, D. (1999) 'The difficulty of achieving full coverage of international social science literature and the bibliometric consequences', *Scientometrics 44*(2): 193–215.

Hicks, D. (2009) 'Evolving regimes of multi-university research evaluation', *Higher Education 57*(4): 393–404.

HM Treasury (2004) *Science and innovation investment framework, 2004–2014*. London: HMSO.

Hirsch, J.E. (2005) 'An index to quantify an individual's scientific research output', *Proceedings of the National Academy of Science 102*(46): 16569–72.

Ismail, S., Farrands, A. and Wooding, S. (2009) *Evaluating grant peer review in the health sciences: a review of the literature*, TR-742-DH. Santa Monica, CA: RAND Corporation.

Jayasinghe, U.W., Marsh, H.W. and Bond, N. (2001) 'Peer review in the funding of research in higher education: the Australian experience', *Educational Evaluation and Policy Analysis 23*(4): 343–64.

Jayasinghe, U.W., Marsh, H.W. and Bond, N. (2003) 'A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings', *Journal of the Royal Statistical Society 166*(3): 279–300.

King, D. (2004) 'The scientific impact of nations: what different countries get for their research spending', *Nature 430*(6997): 311–16.

Lewison, G. and Devey, M.E. (1999) 'Bibliometric methods for the evaluation of arthritis research', *Rheumatology 38*(1): 13–20.

MacLeod, D. (2005) 'Labour turns spotlight on science investment', *The Guardian*, 7 March. [Accessed 18th November 2009: http://www.guardian.co.uk/education/2005/mar/07/highereducation.uk]

Martin, B.R. (1996) 'The use of multiple indicators in the assessment of basic research', *Scientometrics 36*(3): 343–62.

Martin, B.R. and Irvine, J. (1983) 'Assessing basic research: dome partial indicators of scientific progress in radio astronomy', *Research Policy 12*(2): 61–90.

Moed, H.F. (2002) 'The impact–factors debate: the ISI's uses and limits', *Nature 415*(6873): 731–2.

Moed, H.F. (2005) *Citation analysis in research evaluation*. Dordrecht: Springer.

Moed, H.F. (2007) 'The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review', *Science and Public Policy 34*(8): 575–83.

Moed, H.F. (2008) 'UK Research Assessment Exercises: informed judgements on research quality or quantity?', *Scientometrics 74*(1), 141–9.

Moed, H.F., Burger, W.J.M., Frankfort, J.G. and Van Raan, A.F.J. (1985) 'The use of bibliometric data for the measurement of university research performance', *Research Policy 14*(3), 131–49.

Narin, F., Olivastro, D. and Stevens, K.A. (1994) 'Bibliometrics/theory, practice and problems', *Evaluation Review 18*(1): 65–76.

National Audit Office (2003) *Getting the evidence: using research in policymaking*, HC 586-1, session 2002–3. London: National Audit Office.

National Institute of Health Research (NIHR) (2006) 'Biomedical research centres'. [Accessed 18th November 2009: http://www.nihr.ac.uk/infrastructure/Pages/infrastructure_biomedical_research_centres.aspx]

Noyons, E.C.M. and Moed, H.F. (1999) 'Combining mapping and citation analysis for evaluative bibliometric purposes: a bibliometric study', *Journal of the American Society for Information Science 50*(2): 115–31.

Noyons, E.C.M., Moed, H.F. and Van Raan, A.F.J. (1999) 'Integrating research performance analysis and science mapping', *Scientometrics 46*(3): 591–604.

Noyons, E.C.M., Buter, R.K. and Van Raan, A.F.J. (2003) 'Mapping excellence in science and technology across Europe', report to the European Commission by the Centre for Science and Technology Studies, Leiden University, and Frauenhofer Institute for Systems and Innovation Research, Karlsruhe.

Organisation for Economic Co-operation and Development (OECD) (2002) *Frascati Manual: proposed standard practice for surveys on research and experimental development.* Paris: OECD.

Oppenheim, C. (1997) 'The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology', *Journal of Documentation 53*(5): 477–87.

Porter, A.L. and Rossini, F.A. (1985) 'Peer review of interdisciplinary research proposals', *Science, Technology & Human Values 10*(3): 33–8.

Pritchard, A. (1969) 'Statistical bibliography, or bibliometrics?', *Journal of Documentation 25*(4): 348–9.

Research Evaluation and Policy Project (2005) *Quantitative indicators for research assessment – a literature review: discussion paper 05/1*. Canberra: Research School of Social Sciences, Australian National University.

Rinia, E.J., Van Leeuwen, T.N., Van Vuren, H.G. and Van Raan, A.F.J. (1998) 'Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands', *Research Policy 27*(1): 95–107.

Rinia, E.J., Van Leeuwen, T.N., Van Vuren, H.G. and Van Raan, A.F.J. (2001) 'Influence of interdisciplinarity on peer-review and bibliometric evaluations of physics research', *Research Policy 30*(3): 357–61.

Scopus (2008) 'Scopus in detail: what does it cover?' [Accessed 20 July 2009: http://info.scopus.com/detail/what/]

Small, H.G. (1982) 'Citation context and content analysis', in B. Dervin and M. Voight (eds) *Progress in Communication Sciences*, pp. 287–310. Norwood, NJ: Ablex.

Smith, R. (2001) 'Measuring the social impact of research', *British Medical Journal 323*(7312): 528.

Testa, J. (2009) 'Essay: The Thomson Reuters Journal Selection Process'. [Accessed 20 July 2009: http://isiwebofknowledge.com/benefits/essays/journalselection/]

Tijssen, R.J.W. (2003) 'Scoreboards of research excellence', *Research Evaluation 12*(2): 91–103.

Times Higher Education (2008) '2008 World University Ranking'. [Accessed 7 August 2009: http://www.timeshighereducation.co.uk/hybrid.asp?typeCode=430&pubCode=1&navcode=105]

UK Evaluation Forum (2006) *Medical research: assessing the benefits to society*. London: Academy of Medical Sciences.

Van Leeuwen, T. (2006) 'The application of bibliometric analyses in the evaluation of social science research: who benefits from it and why it is still feasible', *Scientometrics 66*(1): 133–54.

Van Leeuwen, T. and Moed, H.F. (2002) 'Development and application of journal impact measures in the Dutch science system', *Scientometics 53*(2): 249–66.

Van Leeuwen, T., Moed, H.F. and Reedijk, J. (1999) 'Critical comments on Institute for Scientific Information impact factors: a sample of inorganic molecular chemistry journals', *Journal of Information Science 25*(6): 489–98.

Van Leeuwen, T., Visser, M.S., Moed, H.F., Nederhof, T.J. and Van Raan, A.F.J. (2003) 'The Holy Grail of science policy: exploring and combining bibliometric tools in search of scientific excellence', *Scientometrics 57*(2): 257–80.

Van Leeuwen, T., Hoorens, S. and Grant, J. (2009) 'Using bibliometrics to support the procurement of NIHR biomedical research centres in England', *Research Evaluation 18*(1): 71–82.

Van Raan, A.F. (1998) 'In matters of quantitative studies of science the fault of theorists is offering too little and asking too much', *Scientometrics 43*(1): 129–39.

Van Raan, A.F. (2001) 'Bibliometrics and internet: some observations and expectations', *Scientometrics 50*(1): 59–63.

Verbeek, A., Debackere, K., Luwel, M. and Zimmermann, E. (2002) 'Measuring progress and evolution in science and technology – I: the multiple uses of bibliometric indicators', *International Journal of Management Reviews 4*(2): 179–211.

Vinkler, P. (1998) 'Comparative investigation of frequency and strength of motives toward referencing, the reference threshold model: comments on theories of citation?', *Scientometrics 43*(1): 107–27.

Warner, J. (2000) 'A critical review of the application of citation studies to the Research Assessment Exercises', *Journal of Information Science 26*(6): 453–60.

Webster, B.M., Lewison, G. and Rowlands, I. (2004) *Mapping the landscape*. London: City University.

Weingart, P. (2005) 'Impact of bibliometrics upon the science system: inadvertent consequences?', *Scientometrics 62*(1): 117–31.

Wellcome Trust (1998) *Mapping the landscape. National biomedical research outputs 1988–1995*. London: Wellcome Trust.

Yoshikane, F., Kageura, K. and Tsuji, K. (2003) 'A method for the comparative analysis of concentration of author productivity, giving consideration to the effect of sample size dependency of statistical measures', *Journal of the American Society for Information Science and Technology 54*(6): 521–8.

APPENDICES

Appendix A: Understanding 'impact', 'quality' and 'excellence' in research evaluation

The development of new structures and processes for supporting decision-making in science and technology policy has taken place despite ongoing disagreement about what might be considered the 'terms of the debate'. Partly, this is because evaluators seek evidence of sometimes quite different things, depending on their position. For example, a research institution head is likely to be more interested in details of the scientific impact of the organisation's activities rather than the kind of social or economic impact information that a national-level research funder may seek to collect. However, in broad terms, often the use of evaluation techniques is premised on the notional goal of improving the 'quality' and/or 'impact' of scientific research.

These terms 'quality' and 'impact' are notoriously difficult to define. Possible dimensions of quality could include (after Moed et al., 1985):

- 'cognitive' quality – in terms of the formulation of the research question;
- methodological quality – in terms of the rigour of the research approach;
- aesthetic quality – which the authors do not fully explain; and
- 'scientific quality' – however defined.

There is little consensus on these issues and various other approaches have been suggested, with some emphasising the relative and socially-constructed nature of quality, and others favouring apparently more objective, absolute definitions (for example, Cole and Cole, 1973).

However, there is greater agreement on the definitions of 'impact'. Martin and Irvine define impact as:

> the actual influence on surrounding research activities at a given time. While this will depend partly on the importance of the research, it may also be affected by such factors as the location of the author, and the prestige, language and availability of the publishing journal. (1983: 70)

This approach is broadly accepted. Therefore, bibliometric measures of impact tend to focus on aspects of the reception of scientific work by others working in the same or similar fields, and the actual importance of work as judged by the research community.

Importantly, for our purposes here, there has been an increasing trend in science and technology policy recently to dissociate 'impact' – i.e. scientific publications – from other forms of output or outcome such as translation into medicines, social and economic outcomes and so forth. In this sense, the key proxy for 'quality' in research remains measures of publication output and impact, as opposed to downstream outputs and outcomes that reflect the 'utility' of particular streams of scientific research. While this approach has its merits for short to medium-term performance analysis exercises, it hinges on a rather restrictive view of quality on the terms described above.

In recent years there has been a renewal of interest in the idea of research 'excellence', reinforced in Europe by its inclusion as a leading theme in recent European Commission framework programmes (Tijssen, 2003). Although the pursuit of 'excellence' is highlighted frequently as a central aim by both research funders and academic institutions, few have tackled the meaning of this term directly. Tijssen (2003) is a rare exception: in a recent article on the subject, he argues that excellence is defined by two key features: (1) "superior quality", and (2) the extent to which it "goes beyond a standard" (although, unfortunately, he does not elaborate in great detail on the meaning of these terms). Some of the indicators that have been developed to support this kind of analysis are reviewed in the main body of this report.

Appendix B: Applying bibliometrics in other contexts – two examples

Using bibliometrics as part of a composite approach to R&D activity assessment: the Frascati Manual

For some years the OECD has been one of the most highly regarded sources of information on international R&D expenditure, and has issued updated figures regularly on the basis of an agreed method, as outlined in the Frascati Manual (OECD, 2002). The most recent version is the sixth edition of a document first published in 1963.

At the time of the inception of the Frascati approach, there was some scepticism that collecting this kind of information in a meaningful way was possible, for several reasons. First, the information collected was used primarily for management, and therefore could not allow for exhaustive examination of the key underpinning factors for specific types of investments. Second, the basic nature of university R&D meant that outputs were difficult to identify (Godin, 2002). The Frascati Manual helped to overcome these doubts by putting forward a composite approach to data-gathering on R&D, with bibliometric methods forming one part of the vast array of methods included in the manual to produce overall measures of science and technology output. These measures include assessments of inputs into the R&D system (whether in terms of personnel or expenditure), and surveys of key actors in the science and technology field.

Using bibliometrics to rank universities

There is growing interest in the potential of bibliometrics as a tool which can be used to help rank higher education establishments both within countries and internationally. Two leading examples of this are the Academic Ranking of World Universities,[35] produced by the University of Shanghai, and the annual rankings issued by *Times Higher Education* in the UK (see Times Higher Education, 2008). Both aim to provide an overall sense of the quality of higher education institutions, including teaching as well as research quality and impacts.

[35] See: www.arwu.org (as of 18th November 2009).

Both of these ranking systems include assessments of research outputs as part of their ranking protocol; crucially, however, they form only one indicator among a suite of several – some quantitative, others qualitative – all of which are combined to derive the final institutional ranking. The table below compares the Shanghai approach with that used by *Times Higher Education*, demonstrating in particular the weight attached to peer review by the British model.

Table: List of key measures included in the University of Shanghai and *Times Higher Education* world university ranking systems, with weightings in each case

University of Shanghai ranking criterion	Weighting	*Times Higher Education* criterion	Weighting
Quality of education, measured by number of alumni winning Nobel Prizes/Fields Medals	10%	Peer review by academics in similar fields	40%
Quality of faculty, measured by number of staff winning Nobel Prizes/Fields Medals	20%	Recruiter review – by graduate recruiters, companies and so forth	10%
Research output – highly cited researchers in 20 fields	20%	Faculty-student ratio	20%
Research output – articles published in *Nature* and *Science*	20%	Research output – citations per faculty member	20%
Research output – articles in the Science Citation Index	10%	International faculty score – percentage of overseas faculty on staff	5%
		International student score – percentage of overseas students	5%