September 2009/39
Issues paper

This report is for information

This report presents the work undertaken to develop the bibliometrics element of the Research Excellence Framework, the key part of which has been a pilot exercise to develop bibliometric indicators.

Report on the pilot exercise to develop bibliometric indicators for the Research Excellence Framework

HIGHER EDUCATION FUNDING COUNCIL FOR ENGLAND

Contents

The annexes are available to download at www.hefce.ac.uk with this document under Publications.

Report on the pilot exercise to develop bibliometric indicators for the Research Excellence Framework

To	Heads of HEFCE-funded institutions Heads of universities in Northern Ireland
Of interest to those responsible for	Research, Planning
Reference	2009/39
Date	September 2009
Enquiries to	Pam Macpherson Barrett tel 0117 9317471 e-mail p.macpherson-barrett@hefce.ac.uk

Executive summary

Purpose

1. This report presents the work undertaken to develop the bibliometrics element of the Research Excellence Framework (REF), the key part of which has been a pilot exercise to develop bibliometric indicators.

2. The report forms part of the evidence base for our consultation on the REF (HEFCE 2009/38).

Content

3. The report outlines the aims and process of the pilot and summarises the outcomes and feedback received from participating HEIs and our REF Expert Advisory Groups. We also present the results of further bibliometric analysis that arose out of the pilot exercise and discuss analysis planned for the future.

Coverage

4. The pilot exercise covered 35 Units of Assessment (UOAs) from the 2008 Research Assessment Exercise (RAE 2008) where coverage of research outputs was 40 per cent or greater in the commercial bibliometric databases. Twenty-two higher education institutions (HEIs) were selected to participate in the exercise. They covered a range of HEI types and sizes.

5. Publications between 2001 and 2006 inclusive were analysed; including citations up to 2007. We used the two main commercially available citation databases for the pilot exercise: Web of Science (WoS) and Scopus.

Main models

6. We tested three models in the pilot exercise:

a. **Model 1, based on institutional address.** Data were harvested from the citation databases from publications recorded between 2001 and 2006 according to the inclusion of an address associated with a participating HEI. These data were then mapped to UOAs.

b. **Model 2, based on staff, all papers.** Data sets were constructed to include all publications produced by staff submitted by the HEI to RAE 2008 between 2001 and 2006.

c. **Model 3, based on staff, selected papers.** Data sets were constructed to include the six most highly cited papers of staff submitted by HEIs to RAE 2008.

Outcomes

7. The main outcomes are presented graphically and in tabular form in the annexes to this report, which may be downloaded alongside this document at www.hefce.ac.uk under Publications. We took extensive advice on the interpretation of the outcomes from the REF Expert Advisory Groups and participating pilot institutions.

Key points

8. Bibliometrics are not sufficiently robust at this stage to be used formulaically or to replace expert review in the REF. However there is considerable scope for citation information to be used to inform expert review.

9. The robustness of the bibliometrics varies across the fields of research covered by the pilot, lower levels of coverage decreasing the representativeness of the citation information. In areas where publication in journals is the main method of scholarly communication, bibliometrics are more representative of the research undertaken.

Action required

10. No action is required in response to this document.

Introduction

11. We are consulting on proposals for the Research Excellence Framework (REF), the first assessment of which will run in 2013 (HEFCE 2009/38). The REF will consist of a single, unified framework for the funding and assessment of research across all subjects. It will make greater use of quantitative indicators than the current method, the Research Assessment Exercise (RAE), while taking account of key differences between the different disciplines.

12. Bibliometric information will form part of the new quantitative indicators for the REF. In order to assess bibliometrics potential and how they might be used, HEFCE has been carrying out developmental work since 2007. The pilot exercise was an important strand in this work, helping to develop and test potential models, assess the process and implications of data collection for bibliometric analysis, and how they could be used within the REF.

13. In May 2009, we published an interim report of the pilot exercise[1].

14. This final report of the pilot exercise forms part of the evidence base for the consultation. It provides the main outcomes from the pilot exercise, although we are undertaking some continued work on bibliometrics that will be reported on in due course.

Structure of the report

15. This report is an overview of the pilot process and provides further detailed information in a series of annexes. Some information from related reports is also summarised here but for greater detail, please refer to the original reports.

16. The main report covers the following areas:

- aims and scope of the pilot, giving a general overview of the pilot and its processes; we note how the pilot institutions were selected and present the models developed for producing bibliometrics
- compilation of the pilot databases including data collection for each of the main models and a description of the databases
- the process of normalising citation counts to give meaningful indicators
- the analysis performed to produce the indicators and outcomes for the bibliometrics pilot exercise; graphs and tables containing the pilot outcomes are referenced in annexes to this section
- the interpretation and feedback provided to HEFCE by the Expert Advisory Groups and the pilot institutions
- further analysis on developing bibliometrics for use in the REF including areas highlighted through feedback
- areas where work continues but which will feed into the continued development of the REF.

[1] Interim report of the REF bibliometrics pilot exercise. A report by HEFCE http://www.hefce.ac.uk/pubs/rdreports/200 9/rd13_09/

Aims and scope of the pilot

17. The pilot exercise aimed to:

- explore which subjects should use bibliometric indicators under the new framework
- assess which categories of staff and publications should be included in citation indicators for the REF
- test the main sources of citation data (the Web of Science (WoS) and Scopus)
- develop the process for collecting and managing bibliographic data
- develop and test methods for analysing citations and benchmarking against international norms
- identify our preferred means of constructing the indicator in the form of a citation profile
- develop proposals for how citation indicators should be used to assess research quality within the REF
- explore what supplementary information the process can usefully generate.

Units of Assessment included in the REF bibliometrics pilot

18. The pilot exercise was not intended to be a 'dry run' for using bibliometrics in the REF but to explore and test the various options available to us across a range of subjects. We chose to include RAE 2008 Units of Assessment (UOAs) that had coverage of 40 per cent or greater in either the WoS or Scopus databases, as determined by work undertaken by CWTS Leiden[2]. This was in order to determine where bibliometrics could provide reliable, useful information.

UOA	UOA name
1	Cardiovascular Medicine
2	Cancer Studies
3	Infection and Immunology
4	Other Hospital Based Clinical Subjects
5	Other Laboratory Based Clinical Subjects
6	Epidemiology and Public Health
7	Health Services Research
8	Primary Care and Other Community Based Clinical Subjects
9	Psychiatry, Neuroscience and Clinical Psychology
10	Dentistry
11	Nursing and Midwifery

[2] 'Appraisal of citation data sources: a report to HEFCE by the Centre for Science and Technology Studies, Leiden University' (September 2008) may be read in full at www.hefce.ac.uk under Publications/Research & evaluation.

UOA	UOA name
12	Allied Health Professions and Studies
13	Pharmacy
14	Biological Sciences
15	Pre-clinical and Human Biological Sciences
16	Agriculture, Veterinary and Food Sciences
17	Earth Systems and Environmental Sciences
18	Chemistry
19	Physics
20	Pure Mathematics
21	Applied Mathematics
22	Statistics and Operational Research
23	Computer Science and Informatics
24	Electrical and Electronic Engineering
25	General Engineering and Mineral & Mining Engineering
26	Chemical Engineering
27	Civil Engineering
28	Mechanical, Aeronautical and Manufacturing Engineering
29	Metallurgy and Materials
32	Geography and Environmental Studies
34	Economics and Econometrics
40	Social Work and Social Policy & Administration
43	Development Studies
44	Psychology
46	Sports-Related Studies

Selection of pilot institutions

19. A key part of the pilot exercise was the participation of higher education institutions (HEIs). The aim was to develop and test bibliometric indicators for a sector-wide assessment process and so we wished to recruit a spread of HEIs. Institutions were invited to express an interest in principle in participating in the pilot as part of the November 2007 consultation[3].

[3] 'Research Excellence Framework: Consultation on the assessment and funding of higher education research post-2008' (HEFCE 2007/34).

20. Sixty-three UK institutions initially expressed an interest. These HEIs were sent a survey giving further details of the pilot exercise and asked to respond to a set of questions:

a. Would the institution wish to take part in the pilot if selected?

b. To what extent would the institution be able to rely on existing sources of data for data collection?

c. Which UOAs would the institution be able to provide data for?

d. Did the institution foresee any difficulties with particular categories of staff or papers?

e. Would the institution have any particular concerns with our plans to publish the results of the pilot exercise?

f. Did the institution have any preferences about the level and format of data it would receive at the end of the pilot?

g. If selected, would the institution have any particular concerns about the timetable and/or workload involved in the pilot?

21. On return of the survey, 46 institutions indicated that they would be prepared to participate in the pilot exercise and were therefore put forward for selection.

22. Institutions were selected on the following criteria:

- to ensure that a minimum of three institutions covered any one UOA within the pilot exercise
- to cover the range of institution types, weighted towards the more research-intensive
- to cover the range of research information management systems, from sophisticated to basic
- to ensure a geographical spread across the UK.

23. Following these criteria we selected 22 institutions to participate in the pilot exercise.

HEIs participating in the REF bibliometrics pilot exercise

Bangor University

University of Bath

University of Birmingham

Bournemouth University

University of Cambridge

University of Durham

University of East Anglia (UEA)

University of Glasgow

Imperial College London

Institute of Cancer Research (ICR)

University of Leeds

London School of Hygiene and Tropical Medicine (LSHTM)

University of Nottingham

University of Plymouth

University of Portsmouth

Queens University, Belfast

Robert Gordon University

Royal Veterinary College

University of Southampton

University of Stirling

University of Sussex

University College London (UCL)

Models for producing bibliometric indicators

24. Through the bibliometrics pilot we have identified and tested three main approaches or models for producing bibliometric indicators in the REF; within these are a number of sub-variants, discussed further below. In each of these models we have mapped work to RAE 2008 UOAs. This was the most straightforward method of grouping work into cognate subject areas for the purpose of the pilot, as we discuss below. As can be seen from the accompanying consultation, we propose to reconfigure UOAs in the forthcoming REF exercise.

Model 1 – Based on institutional address

25. In this model we took the papers associated with each HEI directly from either WoS or Scopus, based on address data within the databases. Papers were assigned to HEIs based on the addresses of their authors, and they were assigned to a subject category (or multiple subject categories) depending on the journal in which they were published[4].

26. This model is potentially a low-cost approach because – in principle – citation indicators for each discipline at each HEI could be produced without input from institutions. In principle, it provides a comprehensive picture of all an institution's outputs (within WoS or Scopus). However, papers are not linked to specific members of staff, and they are assigned subject areas on the basis of the journals in which they are published, which does not map well onto HEI organisational structures.

Model 2 – Based on authors; all papers

27. In this model, we attempted to identify all the papers published by specified groups of staff within each UOA. For the purposes of the pilot process, we included all staff that were selected for RAE 2008, in relevant UOAs (although variations to this were also tested).

28. This model provides a better link between papers, staff and UOAs. However, considerable effort was required to collect the data.

Model 3 – Based on authors; selected papers

29. In this model, we looked only at the most highly cited papers by staff that were selected for RAE 2008. (In the REF, we propose that institutions should select the papers to be assessed; for the purpose of the pilot we automated the selection process based purely on citation data).

Model variants

30. We looked at a range of model variants for the author-based models. These included:

- limiting the papers included to those that were written while at least one of the authors was at the HEI
- looking at papers associated with a wider spectrum of researchers than those they submitted to the RAE.

31. We found that the outcomes for these models tend to be quite similar to model 2 ('based on authors; all papers'). Unfortunately many institutions found it difficult to produce the additional

[4] Note that these subject categories are essentially clusters of journals, as defined within each of the two citation databases.

data required to construct these variants, making comparisons between them difficult. Therefore the report focuses on the three main models.

Data collected from institutions

32. Because we were testing a range of options and models for constructing bibliometric indicators, we asked for a wide-ranging data set from participating institutions. We appointed Evidence Ltd to run aspects of the pilot process, and developed with Evidence a specification for the data required from institutions; Evidence then managed the process of collecting the data from institutions and reconciling it with the WoS data.

33. Our minimum data requirement was all staff submitted to RAE 2008, and all available outputs by these staff published between 2001 and 2007. We asked pilot HEIs to go beyond this where possible and return details of additional staff and outputs.

34. The citation analysis in the bibliometrics pilot considers only journal articles and review papers. However, we requested data on all types of research outputs from HEIs in order to assess the levels of coverage of the citation databases. Conference proceedings are not included in the pilot process but are increasingly covered within the databases and we will conduct further analysis including conference proceedings in due course.

35. We found that HEIs took different approaches to extending the data beyond staff selected for RAE 2008.

36. Although we collected outputs published between 2001 and 2007, for the purposes of analysis we have only included papers from 2001 to 2006. This was to enable time for new outputs to become cited.

37. A full specification for pilot data collection is at Annex A.

38. The pilot process sought to use readily available data where possible, and in particular to build on HEIs' existing RAE databases. It therefore used similar definitions and timeframes. We have analysed each of the three main models using proxies based on such definitions. These are described below.

Citation databases

39. We are using data from both the WoS and Scopus in the bibliometrics pilot. Both databases are under continued expansion and development and the data that we have taken from them represent a snapshot at the date on which we took it. Evidence was contracted to work with WoS data, HEFCE worked with Scopus data.

Pilot process

40. The pilot process involved the following stages:

a. **Data collection**. Data about institutional papers were harvested directly from WoS/Scopus for the address model, and data about staff and papers were collected from the pilot HEIs and matched to WoS/Scopus for the author-based models. This part of the project was carried out principally by Evidence.

b. **Citation counts and normalisation**. The number of citations to each paper was counted, and this count was normalised by field, year and document type (in the same way

for all papers in the author and address models). Evidence calculated the normalisation factors for WoS; HEFCE for Scopus.

c. **Analysis**. We have produced citation indicators for the papers associated with each UOA at each pilot HEI, using each model (and their sub-variants). Advice on the interpretation of the outcomes was sought from the Expert Advisory Groups and pilot institutions. HEFCE and Evidence have continued to conduct further analysis on key issues raised by these groups; some of which is presented in this report.

Compilation of databases

41. The compilation of the databases used to produce the REF bibliometrics pilot outcomes involved the following main stages:

a. Collection of data from the pilot institutions. This stage was managed by Evidence in conjunction with HEFCE.

b. Creating an address-based database for pilot institutions from WoS based on outputs found using Evidence's reconciled address database.

c. Matching institutional data to WoS and identifying possible additional papers associated with institutions. This process was undertaken by Evidence.

d. Evidence and its subcontractor Symplectic then created data sets of all identified publications for verification by each pilot institution.

e. Normalisation factors were calculated by Evidence for WoS and HEFCE for Scopus.

f. Symplectic made the databases available to HEFCE to calculate outcomes for each model and each institution.

42. The methodology for the bibliometrics pilot was developed with the advice of Evidence. It is described in detail in 'Development of a bibliographic database'[5]. The aim was to develop a wide range of models, described in the section below, for producing bibliometric indicators and assessing their suitability for the REF. Part of the pilot was to test the process of collecting and managing data, so we asked participating HEIs to provide as comprehensive a set of data relating to past and present staff and their publications as possible. This was then extended with additional publications data found by Evidence.

43. We also wished to test the suitability of readily available data for a low-burden approach to producing bibliometric indicators. For this, publication data were pulled directly from WoS and Scopus, where there was an article or review containing the address of a pilot HEI. This led to the creation of the databases subsequently used to conduct analysis of the address-based model.

[5] 'Pilot study of bibliometric indicators of research quality: Development of a bibliographic database. Report to UK HE funding bodies by Evidence' (July 2009) is available at www.hefce.ac.uk under Publications/Research & evaluation.

44. The experiences of pilot institutions in collecting assembling and verifying data were reported by Technopolis in May 2009[6]. There was a consensus that gathering extensive data on all publications associated with staff at an institution would be extremely difficult for many HEIs. There were difficulties for some in drawing together data from disparate sources, some of which were not held electronically. Many institutions also found it difficult to comprehensively identify all published research produced by their staff. This was exacerbated where institutions were searching for publications data for past members of staff. When HEIs were asked to verify additional publications data, further issues arose due to the need to disambiguate papers written by different people with the same name. This was a particular problem for large institutions where the bigger cohort of staff significantly increased the likelihood of this occurring.

45. Pilot HEIs were asked to return data on all of their publication outputs, not just journal articles. This was to help our understanding of publication practices in the different subject areas. This is discussed more fully in Evidence's report 'Construction of a bibliographic database[7]. We only included outputs flagged as RAE type D ('Journal article') in the analyses.

46. In this report, we discuss the general methods that we followed, and the analyses we conducted. Analyses using Scopus were performed in-house at HEFCE. The work on WoS was performed by Evidence. Detailed discussion of these analyses, written by Evidence and HEFCE, is presented in the annexes to this report.

Data collection

Institutional address model

47. For the institutional address model Evidence and HEFCE used data from WoS and Scopus that could be linked to a pilot HEI by its address. For Scopus, HEFCE used the 'affiliation ID' field to extract records associated with each pilot HEI. Where there was ambiguity about whether activity at associated medical schools and the like should be included in the extract, we took an inclusive approach.

48. For WoS, Evidence used address-mapping between UK addresses in the WoS data and identifiable organisations. This mapping has been developed over several years and is most detailed for UK HEIs. Where there is doubt, the reconciliation is verified by checking addresses via researcher web-pages. Evidence used this mapping to extract records associated with each pilot HEI. Where there was ambiguity about whether activity at, for example, medical schools associated with HEIs should be included in the extract, Evidence took an inclusive approach.

Author models

49. For these models, the pilot HEIs provided Evidence and HEFCE with information on their staff and publications within the census period of the pilot process. These were supplemented by additional papers found by Evidence and verified by the pilot HEIs. Evidence and HEFCE then

---

[6] 'Identification and dissemination of lessons learned by institutions participating in the Research Excellence Framework (REF) bibliometrics pilot: Results of the Round One consultation. A report to HEFCE by Technopolis' (May 2009) is available at www.hefce.ac.uk under Publications/Research & evaluation.

[7] See footnote 5

matched these data to WoS and Scopus respectively. These data were aggregated into UOAs by linking each paper to members of staff, and hence UOA.

50. Further discussion of data collection for Scopus is at Annex B.

Description of the databases

51. The number of outputs included in each model is summarised in table F1 of Annex F.

What is not included in the databases?

52. The pilot databases contain only outputs identified as RAE 2008 'type D': journal articles that were subsequently matched to WoS journal articles and reviews by Evidence and Symplectic and those additional articles and reviews identified by Evidence.

53. Evidence carried out some analysis on the balance of outputs types, discussed in chapter 8 of the 'Development of a bibliographic database' report[8]. The balance of output types submitted by pilot institutions is shown in Table 1 Note that this does not fully reflect the publishing patterns of the pilot institutions, rather those publications in the pilot UOAs for which they were easily able to find data.

Table 1 Output balance (total count and percentage by type, ranked by percentage of journal articles) in REF pilot HEI submissions

	Count	Books	Articles	Proceedings	Patents	Reports	Other
Robert Gordon	2,019	4.3	47.2	0.0	1.7	3.4	43.4
Bournemouth	1,803	9.5	59.4	24.3	0.0	4.8	2.0
Southampton	36,873	12.7	60.0	20.8	0.0	0.0	6.4
Nottingham	37,944	13.6	66.8	15.4	0.2	2.2	1.8
Stirling	4,652	12.8	67.1	10.4	0.0	5.5	4.2
Bangor	4,584	9.9	68.5	17.0	0.4	4.3	0.0
UEA	3,068	9.0	71.1	19.6	0.0	0.0	0.4
Bath	18,545	7.9	74.1	16.6	0.0	1.2	0.1
Sussex	6,605	10.4	74.4	11.8	0.4	0.5	2.4
Imperial	54,389	5.0	76.9	16.3	0.7	0.8	0.2
UCL	47,685	8.1	77.1	10.6	0.2	1.1	2.9
Durham	9,427	8.7	77.8	7.5	0.1	0.8	5.0
Leeds	8,534	2.8	81.0	11.8	0.2	0.0	4.1
LSHTM	9,117	8.4	81.8	5.7	0.0	1.7	2.4
Glasgow	21,520	5.0	82.7	11.6	0.0	0.1	0.5
ICR	3,928	4.7	83.3	3.9	0.0	0.0	8.0

[8] See footnote 5.

	Count	Books	Articles	Proceedings	Patents	Reports	Other
Queens	7,471	0.0	84.8	15.2	0.0	0.0	0.0
Plymouth	6,232	9.7	86.6	3.3	0.0	0.3	0.2
Portsmouth	494	0.8	91.3	4.5	0.0	0.0	3.4
Cambridge	24,652	2.3	95.9	0.0	0.6	1.3	0.0
Royal Vet	1,873	0.6	98.7	0.6	0.0	0.0	0.1
Birmingham	16,549	0.0	100.0	0.0	0.0	0.0	0.0
TOTAL	327,964	7.5	76.9	12.1	0.2	1.0	2.2

Normalisation

54. The number of times that papers are cited is not in itself an informative indicator; citation counts need to be benchmarked or normalised against similar research. In particular: citations accumulate over time, so the year of publication needs to be taken into account; citation patterns differ greatly in different disciplines, so the field of research needs to be taken into account; and citations to review papers tend to be higher than for articles[9] and this also needs to be taken into account.

55. We can normalise a citation count by dividing it by the average number of citations obtained by all items included in the bibliometric database that were published in the same year, of the same document type and in the same field as the item under assessment. We discuss how we define a field in paragraphs 57 to 60, below.

56. The approach taken to normalising for the bibliometrics pilot was developed with Evidence. Normalisation for the WoS database was carried out by Evidence as described in Annex C and by HEFCE for Scopus as described in Annex B

What is a field?

57. Fields are used by the providers of the commercial databases (WoS and Scopus) to organise data in a user-friendly, discipline-based format. Journals, and the papers in them, are assigned to one or more fields according to their content. There are different numbers and levels of field categories in WoS and Scopus: at the level used for normalisation in the pilot there are 247 WoS journal categories and 334 Scopus categories.

58. Both databases contain a category described as “multidisciplinary science”. The scope of this category is very broad; it contains journals whose remit is the whole of science, such as Nature, Science and the Proceedings of the National Academy of Sciences (PNAS). So that we can more easily compare like with like, we reassign papers in the multidisciplinary science category to more appropriate categories. This is achieved by using the bibliographies of the multidisciplinary items to infer the subject area they are a part of. The process is the same for both WoS and Scopus, and is described in Annexes C and B respectively.

[9] Note that for the pilot analysis only documents classified as ‘article’ or ‘review’ by the appropriate bibliometric database are included in the analysis.

59. In the pilot, this procedure was only applied to journals in the 'multidisciplinary science' category. The scope of some of the other categories in both databases is quite broad, and we intend to investigate whether these might also benefit from a similar process. We also briefly discuss alternative approaches to normalisation in paragraphs 120 to 135.

60. A list of the normalisation factors we used in the pilot are included in Annex K. They are shown separately for each year, document type (article or review) and subject category.

Indicators and outcomes

Analysis for the address-based models

61. As noted in the model description in paragraphs 25 to 26, harvesting citation data directly from WoS or Scopus enables us to find out about the publication and citation patterns within institutions. It is not possibly to relate this information directly to individuals or UOAs.

62. Both Evidence and HEFCE performed mapping exercises to take information about the pilot institutions' publications and citations and align it to the pilot UOA structure. It should be noted that the field categories in WoS and Scopus do not map directly on to the RAE 2008 UOA structure, or onto HEI management structures. This means that data in this model align more closely with the research published in a whole institution than within a particular department. This is because a researcher may publish across more than one field, not necessarily closely related to the particular fields of their 'home' department.

63. The mapping process undertaken by Evidence for the address-based model is described in the 'Development of a bibliographic database' report[10]. We discuss the mapping process that we followed in Annex B. This is the only point in the entire pilot process where data from the RAE 2008 database was used.

64. One of the main limitations of this method is that it assumes that the RAE 2008 UOAs are mutually exclusive. In practice, this is not the case; there are areas of research that fall within the remit of several UOAs. The choice of where to submit this work in the RAE was an institutional one; we cannot capture the range of institutional choices in this methodology.

65. Furthermore, the mapping process we present here can only take account of the journal in which the work was published. We know that some journals cover a broad remit, which is wider than a single UOA. In these cases, one clearly cannot perform a very accurate mapping of work to the most appropriate UOA. Additionally, researchers can often publish in journals outside their 'home' discipline. Interdisciplinary work is an obvious example of this; collaborations between, say, applied mathematicians and other science disciplines are another. The mapping process cannot take account of this.

66. The address-based models do not link papers to people. Assuming that groups of people remain the most fundamental unit in the REF (for example people's postgraduate research students, people's research grants), there would be substantial challenges in integrating any assessment based on this methodology with other parts of the framework.

[10] See footnote 5.

67. In summary, a submission in the address based model consists of all outputs that we were able to associate with the HEI under evaluation, based on the addresses associated with each output in Web of Science or Scopus. These were assigned to UOAs on the basis of the journals in which they were published.

Staff-based models

68. In this section we explain the process we undertook to generate the staff-based models used in the REF pilot. These models make use of the Symplectic data that were supplied to us by Evidence, to assign outputs to RAE 2008 UOAs. This is done using the UOAs assigned to the staff who wrote them. In contrast to the address-based model, this provides a much 'tighter' link between outputs and the UOA(s) to which they are assigned. These models can reflect the institutional decisions that were made in deciding where to submit staff, and can handle journals whose remit spans several UOAs.

69. Symplectic provided three tables of data: a table of staff, which contains information about RAE submission status, the UOA associated with the individual, and so on; a table of outputs; and a table of links. The table of links associates outputs to individuals. This is a 'many to many' mapping; people have generally written more than one paper, and papers are frequently written by more than one person.

70. We adopted the same procedure to generate the staff-based models for both WoS and Scopus data. Evidence provided us with a list of output identifiers from Symplectic's output table, and the associated bibliometric information for them. We constructed a similar table by matching Symplectic's output table to Scopus, using the procedure described in Annex C (practical approaches section).

71. The supplied table of links included approved links (where the HEI had accepted the proposed link between a member of staff and an output) and declined links (where they had rejected the proposed link). We only included the accepted links in our models.

72. For each of the models we constructed a list of individuals who met the criteria for inclusion and the outputs associated with them. Since each member of staff is linked to a UOA, we can then associate outputs with one (or more) UOAs. We took the approach of only allowing each output to occur at most once per UOA. For example, if an output was linked to two biologists and a chemist, it would appear once in the biology submission for an institution, and once in the chemistry submission. We did not fractionally assign papers between UOAs and/or institutions.

73. Each output should have been de-duplicated in the outputs table, so that it appeared once, regardless of the number of HEIs and individuals associated with it. In practice, we found that there were some instances where the same paper appeared more than once in the outputs table (and had been matched to the same record in WoS or Scopus). To deal with these cases, we filtered outputs down to one instance per submission by the bibliographic databases' unique identifiers ('UT' for WoS and 'keyid' for Scopus).

74. In all models, with the exception of the short time window model, we included only outputs published between 2001 and 2006. We did this to allow a clear year for publications to acquire citations, as discussed in Annex C.

75. We looked at a selective model, in which we ranked each individual's included outputs by normalised citation score, and kept the six papers with the largest normalised citation score. We

then de-duplicated these at the submission level to limit each paper's inclusion to once per submission.

76. We adopted this approach for reasons of computational simplicity, rather than optimising the selection of papers so that each member of staff's best six papers were selected without duplication with other people's papers from the submission. As such, we present this as an example of a selective model, rather than a suggestion of how such a model would actually be implemented in the REF.

77. We believe that any selective model would require the judgement of institutions in making the selection, because the feedback we have received indicates that bibliometrics should be used in conjunction with expert review, not to replace it. The approach we took means that there will typically be fewer than six papers per member of staff. This is because co-authorship 'wastes' some of the papers because we limit each output to once per submission.

78. In summary, the all outputs model contains all articles and reviews that were indexed by the Web of Science or Scopus that were written by RAE submitted individuals. We assign outputs to UOAs on the basis of the UOAs the outputs' authors were submitted to. The selected papers model only includes an output in the model if it was any of the authors' best[11] six outputs.

Other variants of staff-based models

79. We looked at a range of author-based models. Work using some of the author metadata continues; for example, we plan to look at equalities and early career researcher issues using the data in the staff table. These are discussed in Work in progress from paragraph 118 below.

Calculation of indicators

80. Having constructed sets of papers, we then calculated a series of indicators to summarise the scores of the papers included in the submission. We give the outcomes for each submission with the following indicators:

- mean normalised citation score
- median normalised citation score
- proportion of outputs in submission greater than twice the world average (that is to say, with a normalised citation score greater than two)
- proportion of outputs in submission greater than four times the world average
- proportion of outputs not yet cited.

81. We selected the above indicators to look at a range of properties of the distribution of citation scores within each submission. The average normalised citation score is given by the mean; however because the distribution of citations scores is skewed, we also include the median normalised citation score to indicate the centre of the distribution. To give some idea of the amount of work of different quality levels, we also calculated proportions or outputs at twice and four times the world average. Taken together, these indicators give an indication of the overall make-up of a submission.

[11] We measure "best" by including the outputs with the six largest normalised citation scores.

82. Evidence has done further work on the choice of indicators suggesting that centiles, deciles and quartiles might provide more or complementary information. See Annex I for further details.

Sources of citations

83. We were interested in examining the issue of self-citation in the pilot. Unfortunately, it is difficult to rigorously define and measure author self-citation; for example, how does one handle co-authorship or synonyms? The citation counts we calculated, do not take account of (author) self-citation. Instead, we decided to provide additional submission-level information to the Expert Groups regarding institutional and international citation rates. We provided these as a possible 'contextual' indicator, reported alongside the other results in Annex F.

84. We looked at institutional citations, which we defined as citations coming from the same institution as the output that was being analysed. This means that each output has a separate institutional citation count for each of the institutions who were linked to the output. In defining institution, we used Scopus' 'affiliation ID' field, which consolidates address variants, to determine whether a citation came from the same institution as the one who wrote it.

85. We also looked at the number of international citations an output received. We defined this as the number of citations from papers with at least one author associated with a non-UK address.

86. These two definitions are not mutually exclusive; a paper could be co-authored between the institution under analysis and an institution abroad. If this case, the citation would be both international and institutional.

87. Evidence defined these measures in the same way for its WoS analyses, using the database of address variants in place of Scopus' affiliation ID field.

88. We only returned institutional and international citation figures at the level of the submission.

89. We also computed the proportion of outputs in a submission that were the result of an international collaboration. We considered papers in a submission to be the result of an international collaboration if they had a non-UK address on them.

The centile approach

90. When outcomes were shown to members of the Expert Advisory Groups (discussed in paragraphs 97 to 101), they indicated that it would be helpful to see data presented in centiles. Evidence has therefore conducted some exploratory analysis towards presenting bibliometric data in centiles and this is discussed in Annex I. HEFCE is conducting further analysis internally and will publish this in due course.

Outcomes

Main outcomes

91. Graphs showing the outcomes for the three models considered in this report can be found at Annex E. These models are:

a. **The address-based model.** Note that some UOAs do not have any outputs associated with them in the address-based model, owing to the processes used to map database subject categories to UOAs.

b. **The (baseline) submitted staff model.** This includes all outputs linked to RAE submitted members of staff in each UOA

c. **The selected outputs model.** This takes the staff included in the submitted staff model, and limits the papers included to each individual's six papers with the highest normalised citation score.

92. To aid interpretation of these data, we include indicators from the RAE 2008 output sub-profile (this sub-profile is more relevant than the overall quality profiles which include environment and esteem factors). This is in order to provide an initial 'sense check' of the outcomes of the pilot process against the only other available quality indicators (the RAE). This is not intended as a direct comparison between the RAE and the bibliometrics pilot outcomes, because there are a number of differences between the scope, coverage, assessment criteria and methods used for each exercise. For these reasons, the preferred bibliometric model should not necessarily be the one that provides the closest fit with RAE outcomes.

93. For the purposes of this 'sense check' against the pilot outcomes, we present the 'percentage above twice the world average' bibliometric indicator alongside the proportion of outputs rated 3* or above in the RAE in the top graphs. We present the 'percentage above four times the world average' bibliometric indicator alongside the proportion of outputs rated 4* in the RAE in the bottom graphs. This is for illustrative purposes and does not mean that we regard 'twice the world average' as equivalent to 3* or 'four times the world average' as equivalent to 4*.

94. In reporting the bibliometric outcomes, we have applied a size threshold to the summary graphs where submissions with fewer than 50 matched papers in either the WoS or Scopus submitted staff model have not been included.

Detailed outcomes

95. Outcomes tables containing detailed information for each of the main models can be found in Annex F. For each model, we show the following indicators:

- the number of outputs included in the model
- the mean normalised citation score for these outputs
- the median citation score for these outputs
- the proportion of the outputs greater than twice world average
- the proportion of the outputs greater than four times world average
- the proportion of the outputs that are (as yet) uncited
- the proportion of citations to the outputs that are from the same institution
- the proportion of citations to the outputs that are from overseas
- the proportion of outputs that are an international collaboration.

96. We also show the number of individuals associated with the submitted staff model. We have not included the results of a model where it contains fewer than 50 matched outputs or fewer than 10 staff.

Interpretation

Expert Advisory Group feedback

97. The members of the Expert Advisory Group were recruited to provide expert feedback to HEFCE on the development of the REF. At meetings during April and May 2009 we sought their advice on interpretation of the outcomes and the options for using bibliometrics in the REF. Paragraphs 98 to 100 are the key points of their discussions at a series of break-out groups.

98. There was a strong consensus that bibliometrics are not sufficiently mature to be used formulaically or to replace expert review, but there is considerable scope for citation indicators to inform expert review in the REF.

99. There was widespread agreement that the most appropriate approach is to focus on citation indicators for selected papers by the staff in each submission, rather than attempt to capture all papers.

100. There are a number ways in which bibliometrics can be used to inform expert review to enhance the reliability of the process and in some cases reduce panel workloads; the particular ways in which the data are used could vary between panels.

101. A fuller summary of the discussion of bibliometrics is at Annex G.

Pilot HEI feedback

102. Feedback from the pilot institutions was gathered in two stages by Technopolis. The first stage, published in May 2009 focused on the data collection and verification aspects of the pilot exercise[12]. Pilot institutions found the data requirements challenging where they did not already have sophisticated research information management systems.

103. The report states that a requirement on institutions to collect publications information for an 'authors; all papers' model would be burdensome where there were no well-developed publications systems in place. Institutions also felt that even though an address-based model would be less burdensome in principle, they would still wish to verify papers associated with their institution.

104. The second round of feedback was gathered after the pilot institutions had received their outcomes and sought to gather views on the accuracy of the bibliometrics results, reasons for any outliers and implications for institutional research management.

[12] See footnote 6.

105. The report[13] presenting the results of Technopolis' second consultation reveals that institutions generally felt that a selective staff-based model best reflected their perception of quality as it was the closest match to the RAE results. Institutions felt that the selective model, and that based on HEI address, were potentially the lowest-burden approaches requiring the least deviation from current research practices.

Further analysis

106. Since receiving feedback from the Expert Advisory Group and bibliometrics pilot institutions we have been conducting further analysis to address some of the issues raised. There are a series of issues that go further than the pilot outcomes in working out how we can use bibliometrics in the REF. This work is presented in this section and the related annexes. Work on some aspects continues, for example on equalities and testing the centiles approach and will be published in due course.

Time windows

107. Evidence has conducted analysis to determine what timeframe would give a reasonable balance between allowing enough time for citation counts to grow and being unduly retrospective. It suggests that a four-year publication window for citation analysis provides a measurable benefit over a two-year window in producing results that are not significantly different from those that would be obtained by using a six-year window (see the section on time windows in Annex I).

Stability of indicators

108. We have been working to assess the stability of the indicators we produced and to relate this to the quality and coverage of the citation databases available commercially. The work completed so far relates to the comprehensiveness of coverage of the pilot data submissions (this is discussed in Annex J). Further work will focus on the effect of journals ceasing to be covered.

Coverage of databases

109. Evidence has done some work illustrating the coverage of WoS by database subject area, which can be found in Annex H. This is done by looking at the proportion of citations from each subject area that are linked to other records in WoS. This can be done by counting the proportion of citations that have been given a 'reference item ID'. This is used to link the citation in the bibliography of the citing item to the record corresponding to the item that is being cited. If such a link is not made then no reference item ID is assigned. Even if the record being cited is outside the period for which data are held, it is possible to see whether the link to another WoS record has been made by the existence of the reference item ID.

[13] 'Identification and dissemination of lessons learned by institutions participating in the Research Excellence Framework (REF) bibliometrics pilot: Results of the Round Two Consultation' (September 2009) is available at www.hefce.ac.uk under Publications/Research & evaluation.

110. We cannot perform a similar analysis on the Scopus data that we hold, because they handle citations to items they do not cover differently. Rather than leave an unresolved link, they create a 'dummy' record containing the citation's information. We do not hold these dummy records, and cannot distinguish between them and normal records on the basis of the item identifier.

111. We are currently working on alternative methods of measuring the coverage in Scopus, and will report on this in due course.

Field differences

112. There are differences in the density of citations across subject fields in both databases. This has been investigated by Evidence for WoS (see Annex H, coverage of cited items) who found that coverage is greatest for the natural sciences in WoS but that coverage decreases in Engineering and Social Science.

113. We have produced a table of normalisation factors for WoS and Scopus, presented in Annex K. This shows the average number of citations per paper, per field, per year. Where the normalisation factors are low, the effect of a single citation is much greater and the information is less robust and therefore less useful. We expect that the data will inform decisions about which disciplines will use citation data in the REF.

Work in progress

114. There are a number of areas where work is still in progress. These will be reported on when completed in the autumn. This section identifies the areas where HEFCE is still performing analyses and presents any early information.

Database properties

115. We are carrying out some comparative analysis on the differences in citations in WoS and Scopus. It has so far been noted that where there were differences in the citation counts, some were fairly significant. The analysis on database properties will be published when complete.

Citation link accuracy

116. We know that a small proportion of citation links in both databases are likely to have been made incorrectly (for example the bibliography item of the citing paper will have been linked to an incorrect target item), or not be made. It is also likely that errors in the published bibliographies of citing documents will cause some citation links to be missed. We will undertake further work on this to examine how frequent incorrect or missing links are, by comparing the two databases.

117. It is very challenging to determine whether occasional citation links to a paper have been missed by one or other of the database providers. This is because we cannot detect their absence except by comparison with the other database (assuming both the citing and cited journals are covered in both databases). We plan to look at journals that are covered in both databases in order to investigate this further.

Equalities citation analysis

118. We are currently carrying out analysis on the pilot data set with regard to equalities factors. This will focus principally on early career researchers because this is where the data set is richest. The principal question we are asking is: 'are early career researchers as likely to produce highly cited research papers as other researchers?'

119. The methodology used will be similar to that used for 'Selection of staff for inclusion in RAE2001' (HEFCE 2006/32).

Other methods for constructing normalisation factors

120. As discussed in paragraphs 54 to 60, we compare the number of citations of each output against the average number of citations for similar papers. We define similar papers as being those published in the same year, of the same document type (article or review) and in the same subject area.

121. It is clear that we must take the length of time elapsed since publication into account when defining similar papers; papers are, when first published, uncited and accumulate citations over time. This is illustrated in Evidence's analysis accompanying this report at Annex C.

122. The work by Evidence at Annex C shows that articles and reviews attract citations at different rates (controlling for publication year and subject area), with reviews typically attracting several times the number of citations as an equivalent 'normal' article. Additionally, there remain some questions over the research content of some reviews, and whether such items are always appropriately classified by the database providers.

123. In defining the subject areas against which we wish to normalise, both we and Evidence use the subject categories that are assigned to each journal covered by each of the databases. There are 247 subject categories used in WoS and 334 in Scopus. The detail of how we construct normalisation factors from these categories is given in Annex B for Scopus and at Annex C for WoS.

124. The academic 'breadth' of the subject categories within each database varies: some are very specific while others are more general. At one extreme, the 'multidisciplinary science' categories in each database cover the whole spectrum of scientific activity, with journals such as Nature and PNAS assigned to it. Both HEFCE and Evidence reclassified work in this category, on a per paper basis, to more appropriate ones, by looking that the work that each item assigned to this category had cited. This is described in paragraph 58.

125. Journals with a broader remit than a single database subject category are generally assigned to several categories by the data providers, reflecting the scope of the journal's coverage. In the analysis we present in the main part of the report, we dealt with this by assigning each item a composite normalisation factor that was the mean of the normalisation factors corresponding to the subject categories to which the journal had been assigned (for the appropriate document type and publication year).

126. Evidence looked the range of subject categorisations available in Thompson-Reuters' bibliographic products. These range from the fine-grained WoS classifications to the much coarser 'essential science indicators' categorisation. Their work, which is shown in Annex C, suggests that the level of normalisation chosen does not, in general, strongly affect the outcomes at a submission level.

127. Although the subject categorisation used for normalisation is unlikely to have a large effect on the outcomes of a reasonably sized submission level bibliometric assessment (as illustrated by Evidence's work in Annex C), its importance is likely to be far greater when small submissions are assessed and/or paper level bibliometric indicators are given to expert panels. Additionally, an inappropriate selection of normalisation category is likely to have a distorting effect on publication and/or research practices.

128. We commissioned some work from Linda Butler, of RAND Europe, to discuss some of the issues surrounding subject categorisations. This draws on her experiences with the Australian Research Council, and their approach to defining sets of journals for subject categories. This work, 'Options for Defining Normalisation Fields: A report to HEFCE by RAND Europe' can be found at www.hefce.ac.uk under Research/Research Excellence Framework/Resources.

129. In paragraphs 133 to 136, we consider some of the other approaches that we might adopt in constructing groups of cognate work, which may be suitable for normalisation. It must be emphasised that this work is preliminary and speculative.

130. Almost all work in bibliometrics has, to date, looked at classifying work at the level of the journal[14] – all work within a given issue of a journal (of the same document type) will be treated identically for normalisation purposes. It is not clear whether this is always appropriate (in the case of journals such as Science and Nature, which cover large swathes of academic activity, it is not, which is why we reclassified papers in these journals to more appropriate subject categories). However, classifying each output individually is clearly a much more challenging proposition than if we treat all items of the same document type within a journal issue identically. Such a methodology would clearly need to be largely data-driven in its approach if it were to effect a reduction on panel burden.

131. It seems likely that there is no single 'best' way of grouping the subject areas covered by the REF into cognate groups for normalisation; if we adopt too fine a granularity, the process will be unduly self-referential (consider an area of research that only a single research group is working on in the world; the process would merely compare the work of the group with itself). If we adopt too coarse a level of granularity, then any variation in citation practices between areas of research that have been grouped together will create undesirable behavioural consequences.

132. In order to classify papers into cognate groups, we need some method of quantifying the similarity between any two papers. Measures that may provide useful measures of similarity include:

a. **Co-citation links**. Outputs which cite many of the same papers in the literature are likely, themselves, to be similar.

b. **Co-authorship**. People tend to collaborate with others doing similar work. There are, however, difficulties in disambiguating authors with the same name if we adopt this approach.

c. **Text-based similarity**. The databases include keyword information with each output, and the abstract. It may be possible to use these to generate a measure of similarity between outputs.

[14] With the exception of multidisciplinary science, as discussed in paragraph 58.

133. Having determined a way of quantifying the similarity between papers, we then need to group papers into cognate groups. There are several ways we could approach this problem, but it should be noted that the volume of data is likely to impose limits on the techniques we use to do this. One approach may be to treat each year of data independently to reduce the number of papers we are dealing with. This approach would also allow us to look at the stability of groupings with time but could present problems for areas of work where comparatively few research outputs are produced per year.

134. Suitable approaches may include clustering techniques and network-based approaches. In the latter, we can construct a network where similar papers are joined by 'edges', and use 'community detection' algorithms to look for areas of the network that represent similar work.

135. We note that there have been some recent data-driven approaches to 'map' science, including work using online journal usage data[15] and 'concept' mapping[16]. These offer useful insights into approaches that we may take.

Coverage of conference proceedings

136. HEFCE has started work to understand better how conference proceedings behave in Scopus. This has so far focused on conference proceedings in standard journals. Preliminary information is presented in Annex L. Evidence has provided some preliminary thoughts on conference proceedings in Annex K.

[15] Bollen J, Van de Sompel H, Hagberg A, Bettencourt L, Chute R, et al. (2009), 'Clickstream Data Yields High-Resolution Maps of Science'. PLoS ONE 4(3): e4803. doi:10.1371/journal.pone.0004803

[16] 'Scoping study on the use of bibliometric analysis to measure the quality of research in UK higher education institutions: Report to HEFCE by the Centre for Science and Technology Studies, Leiden University', chapter 5 (November 2007). Available at www.hefce.ac.uk under Publications/Research & evaluation.

List of abbreviations

ANZSRC	Australia and New Zealand Standard Research Classification
ARC	Australian Research Council
ASJC	All science journal classification
CC	Thomson Reuters Current Contents
DOI	Digital Object Identifier
ERA	Excellence in Research for Australia (evaluation system)
ESI	Thomson Reuters Essential Science Indicator
HEFCE	Higher Education Funding Council for England
HEI	Higher education institution
ICR	Institute of Cancer Research
IDG	(ERA) Indicator Development Group
ISSN	International Standard Serial Number
LSHTM	London School of Hygiene and Tropical Medicine
O/E	Ratio between observed/expected citation count
PNAS	Proceedings of the National Academy of Sciences
RAE	Research Assessment Exercise
RBI	Rebased impact
REF	Research Excellence Framework
UCL	University College London
UEA	University of East Anglia
UOA	Unit of Assessment
WoS	Web of Science