EPSRC Engineering and Physical Sciences Research Council

evidence

EPSRC CITATIONS STUDY 2009

October 2009

Contact details

Evidence, Thomson Reuters	103 Clarendon Road, Leeds LS2 9DF, UK
t/	+44(0)113 384 5680
f/	+44(0)113 384 5874
e/	scientific.enquiries.evidence@thomson.com

Contents

1 Executive Summary

- This is an initial report of bibliometric analyses of research funded by the Engineering and Physical Sciences Research Council (EPSRC). The analyses have been based on journal articles listed on final reports received during the period 2003–2005.
- The range of journal categories in which EPSRC-funded research has published is diverse, with no one category predominating. However, the principal categories represent the broad research areas covered by EPSRC including Chemistry, Physics and Engineering and Materials Science with specialist areas within these like Optics and Polymer Science.
- Journals used most frequently by EPSRC-funded researchers are those which are leading journals in their respective research fields. EPSRC-funded researchers have published a similar proportion of papers in the elite, multidisciplinary journals of Nature and Science as compared to the UK overall (0.44% compared to 0.42% papers published between 2002 and 2004).
- EPSRC-funded research papers have a good citation impact compared to UK and world baselines. The overall rebased impact (RBI) of all of the EPSRC-funded research papers is 1.63 (where world average is 1.0). For comparison, the UK's average rebased impact relative to world baseline for the 5-year period 2003-2008 inclusive in physical sciences was 1.39 and in engineering was 1.05. More detailed analyses have shown that the average rebased impact of EPSRC-funded research is above world average in all fields.
- Research papers published in Condensed Matter Physics journals and in journals assigned to Multidisciplinary Physics have particularly good citation impact, being more than twice the world average in each category.
- The Impact Profile® for EPSRC-funded research benchmarked against physical sciences and engineering research in the UK shows a clear and consistent profile: fewer uncited papers and more well- and highly-cited research than other, comparable UK research with almost one-tenth of EPSRC-funded research being internationally significant.
- There is no evidence for any significant difference in citation performance for papers arising from 'Responsive' and ''Targeted' funding modes.
- Papers identified as arising from larger grants have, however, had greater citation impact than those from other grants.

2 Introduction

2.1 EPSRC-funded research publications – outline of analyses and report

The Engineering and Physical Sciences Research Council (EPSRC) has provided a bibliography of research papers associated with projects that it has funded. The bibliography has been collated from the journal articles listed on final reports received during the period 2003–2005. It principally covers the years 1999 to 2008, though a few papers outside these years were included.

EPSRC has commissioned *Evidence* to link these publication records of EPSRC-funded research with bibliometric data. The principal aim of the project is to study the quality of outputs from responsive and managed research using citation analysis.

This report provides background bibliometric analyses and UK benchmarks to set the research performance of EPSRC-funded research for a 10-year period between 1999 and 2008 into context as well as the specific comparisons between responsive and managed research and research associated with larger grants.

For the analyses, bibliometric data was sourced from Thomson Reuters and *Evidence* databases. World average citation impact data are sourced from the Thomson Reuters National Science Indicators 2008.

The UK benchmark data cover all sectors of the research community including higher education institutions (HEIs), hospitals and NHS trusts, companies, public sector research organisations and charities. *Evidence* has performed a complete address reconciliation for the Thomson Reuters databases for the UK. This enables an accurate association of journal articles with institutions, using a combination of the name, address components and post-codes.

2.2 Bibliometrics and citation analysis

Bibliometrics are about publications and their citations. The field has emerged from 'information science' and refers to analyses and methods used to study and index texts and information.

Publications cite and are cited by other publications. This provides linkages and networks. Many links are likely to be related to significance or impact. Meaning is determined from keywords and content. Citation analysis and content analysis are therefore commonly used bibliometric methods. Historically, bibliometric methods had been used to trace relationships amongst academic journal citations. Bibliometrics now are increasingly important in indexing research performance. Bibliometric data have particular characteristics of which the user should be aware, and these are considered here.

Journal papers (publications, sources) report research work. Papers refer to or 'cite' earlier work relevant to the material being reported. New papers are cited in their turn. Papers that accumulate more citations are thought of as having greater 'impact', interpreted as significance or influence in their field. Citation counts are therefore recognised as a measure of impact, which can be used to index the excellence of the research from a particular group, institution or country.

The origins of citation analysis as a widespread tool of research performance can be traced to the mid-1950s, when Eugene Garfield proposed the concept of citation indexing and introduced the Science Citation Index, the Social Sciences Citation Index and the Arts & Humanities Citation Index, produced by the Institute of Scientific Information (currently the Science business of Thomson Reuters).

Most impact measures use average citation counts from groups of papers, because some individual papers may have unusual or misleading citation profiles. These outliers are diluted in larger samples.

Data source

The data used by *Evidence* are derived from Thomson Reuters databases, including the Web

of Science, a single source collated to the same standard and therefore providing a level of comparability not found in other databases. These data are also valuable because they can readily be disaggregated by field, by year, by country and by institution. The Web of Science is part of a larger entity, the Web of Knowledge, focussing on research published in journals and conferences in science, medicine, arts, humanities and social sciences. The Web of Science was primarily regarded as an awareness and information retrieval tool but has an increasingly important secondary use for citation analysis and bibliometrics for research evaluation. Coverage is both current and retrospective in the sciences, social sciences, arts and humanities, in some cases back to 1900. Within the research community these data are often still referred to by the acronym 'ISI'.

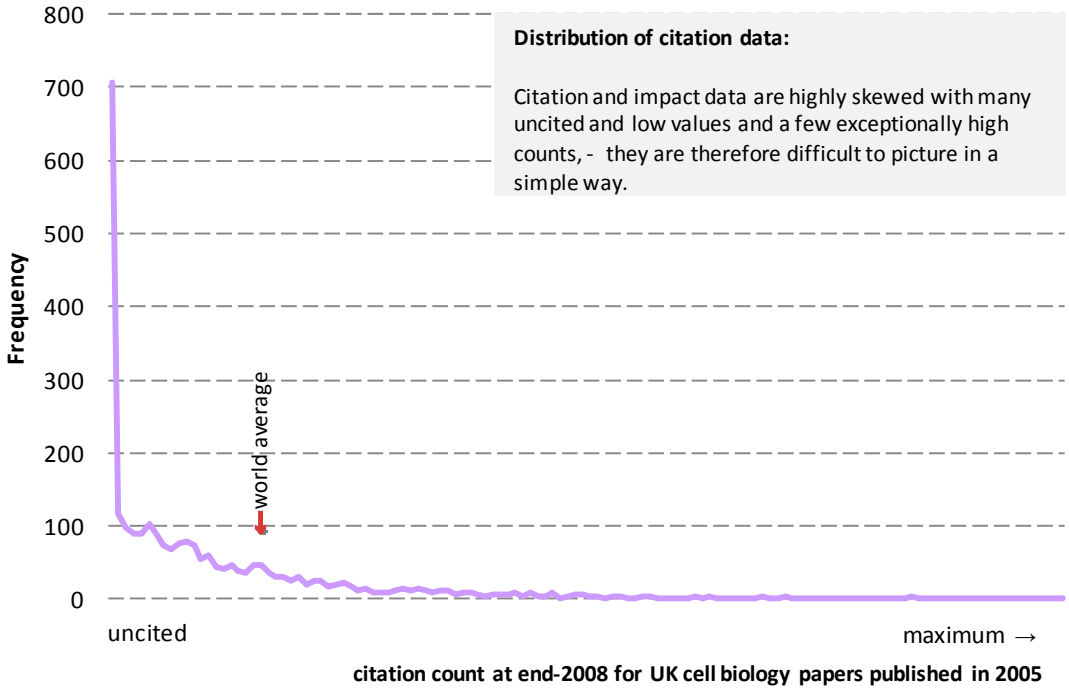
Unlike other databases, the Thomson Reuters Web of Science and underlying databases are selective, that is, the journals abstracted are selected using rigorous editorial and quality criteria. The authoritative, multidisciplinary content covers over 10,000 of the highest impact journals worldwide, including Open Access journals and over 110,000 conference proceedings. The abstracted journals actually encompass the majority of significant scientific reports and, more importantly, an even greater proportion of the scientific research output which is cited. This selective process ensures that the citation counts remain relatively stable in given research fields and do not fluctuate widely from year to year, which increases the usability of such data for performance evaluation.

Evidence, a business of Thomson Reuters, has extensive experience with databases on research inputs, activity and outputs and has developed innovative analytical approaches for benchmarking and interpreting international, national and institutional research impact.

Citation counts

A paper accumulates citation counts when it is referred to by more recent papers. Some papers get cited frequently and many get cited rarely or never, so the distribution of citations is highly skewed.

Why are so many papers never cited? Certainly, some papers remain uncited because their content is of little or no impact, but that is not the only reason. It might be because they have been published in a journal not read by researchers to whom the paper might be interesting. It might be that they represent important but 'negative' work reporting a blind alley to be avoided by others. The publication may be a commentary in an editorial, rather than a normal journal article and thus of



general rather than research interest. Or it might be that the work is a 'sleeping beauty' that has yet to be recognised for its significance.

The figure shows the skewed distribution of more or less frequently cited papers from a sample of UK authored papers in cell biology. The skew in the distribution varies from field to field. It is to compensate for such factors that actual citation counts must be normalised, or rebased, against a world baseline.

Other papers can be very highly cited: hundreds, even thousands of times. Again, there are multiple reasons for this. Most frequently cited work is being recognised for its innovative significance and impact on the research field of which it speaks. Impact here is a good reflection of quality: it is an indicator of excellence. But there are other papers which are frequently cited because their significance is slightly different: they describe key methodology; they are a thoughtful and wide-ranging review of a field; or they represent contentious views which others seek to refute.

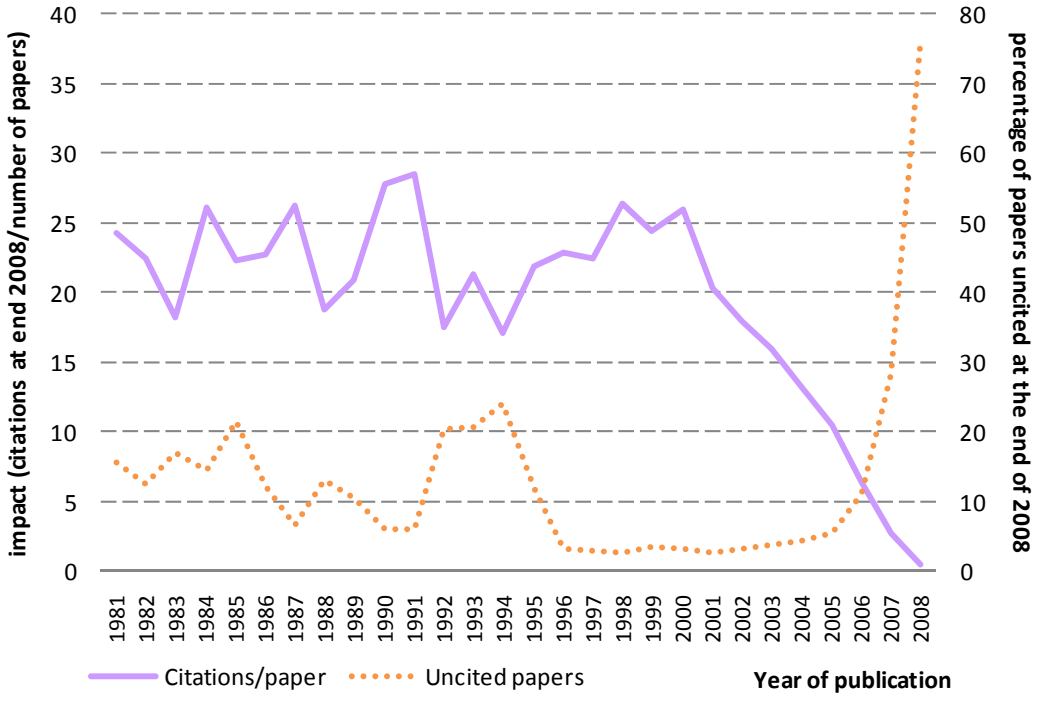
Citation analysis cannot make value judgments about why an article is uncited nor about why it is highly cited. The analysis can only report the citation impact that the paper has achieved. We normally assume, based on many other studies which have linked bibliometric and peer judgments, that high citation counts correlate on average with the quality of the research.

Time factors

Citations accumulate over time. Older papers therefore have, on average, more citations than more recent work. The graph below shows the pattern of citation accumulation for a set of 33 journals in the journal category ***Materials science, Biomaterials***. Papers less than eight years old are, on average, still accumulating additional citations. The citation count goes on to reach a plateau for older sources.

The graph shows that the percentage of papers that have never been cited drops over about five years. Beyond five years, between 5% and 10% or more of papers remain uncited.

Account must be taken of these time factors in comparing current research with historical patterns. For these reasons, it is sometimes more appropriate to use a fixed five-year window of papers and citations to compare two periods than to look at the longer term profile of citations and of uncitedness for a recent year and an historical year.



Discipline factors

Citation rates vary between disciplines and fields. For the UK science base as a whole, ten years produces a general plateau beyond which few additional citations would be expected. On the whole, citations accumulate more rapidly and plateau at a higher level in biological sciences than physical sciences, and natural sciences generally cite at a higher rate than social sciences.

Papers are assigned to disciplines (journal categories or research fields) by Thomson Reuters, bringing cognate research areas together. The journal category classification scheme has been recently revised and updated. Before 2007, journals were assigned to the older, well established Current Contents categories which were informed by extensive work by Thomson and with the research community since the early 1960s, this scheme has been superseded by the 251 Web of Science journal categories which allow for greater disaggregation for the growing volume of research which is published and abstracted.

Papers are allocated according to the journal in which the paper is published. Some journals may be considered to be part of the publication record for more than one research field. As the example below illustrates, the journal *Acta Biomaterialia* is assigned to two journal categories: ***Materials science, Biomaterials*** and ***Engineering, Biomedical***.

Very few papers are not assigned to any research field and as such will not be included in specific analyses using rebased impact data. The journals included in the Thomson Reuters databases and how they are selected are detailed here

http://scientific.thomsonreuters.com/mjl/.

Some journals with a very diverse content, including the prestigious journals *Nature* and *Science* were classified as ***Multidisciplinary*** before 2007. The papers from these ***Multidisciplinary*** journals are now re-assigned to more specific research fields using an algorithm based on the research area(s) of the references cited by the article.

From the publication data, bibliometric analyses have been carried out for all the papers identified as funded by EPSRC and for those papers falling within specific fields. It is rarely possible to make sensible analyses on individual papers and gross averages are unlikely to be informative for management purposes. Mapping publications to research fields satisfactory for appropriate management comparisons is problematic. We use the Thomson Reuters journal categories because these are well-established and are informed by extensive Thomson Reuters work with the research community over the last twenty years.

The research fields used in this report were selected from the EPSRC-funded dataset as the ten most-frequently used from the 251 Web of Science journal categorisation scheme to which papers are assigned by Thomson Reuters. This selection coincided with fields with an output of more than 150 papers over the 10-year period. The categories, ranked by most-used to less-frequently used, were:

- Engineering, Electrical & Electronic Engineering
- Chemistry, Multidisciplinary
- Physics, Condensed Matter
- Physics, Applied
- Materials Science, Multidisciplinary
- Physics, Multidisciplinary
- Optics
- Chemistry, Physical
- Chemistry, Organic
- Engineering, Mechanical

Journal lists and Thomson Reuters definitions for the ten most frequently used research fields are given in Annex 1.

Rebased impact

For the reasons given above, all analyses must take both field and year into account. In other words, because the absolute citation count for a specific article is influenced by its field and by the year it was published, we can only make comparisons of indexed data after normalising with reference to these two variables. In addition, the type of publication will influence the citation count, for example, a review will typically be cited more frequently than an article, and both of these types will tend to be cited more than editorials or meeting abstracts.

Consequently, only citation counts from reviews and articles are used in calculations of impact. The most common normalisation factors are the average citations per paper for the year and either the field or journal in which the paper was published. This normalisation is also referred to as 'rebasing' the citation count.

Impact is therefore most commonly analysed in terms of 'rebased impact', or RBI. The following schematic illustrates how the rebased impact is calculated at paper level and journal category level.

This article in the journal *Acta Biomaterialia,* with 28 cites to the end December 2008*,* is assigned to two journal categories: ***Materials science, Biomaterials*** and ***Engineering, Biomedical***. The world average baselines for, as an example, ***Materials science, Biomaterials*** are calculated by summing the citations to all the articles and reviews published worldwide in the journal *Acta Biomaterialia* and the other 32 journals assigned to this category for each year and dividing this by the total number of articles and reviews published in the journal category. This gives the category-specific rebased impact (in the above example the category-specific RBI for ***Materials Science, Biomaterials*** is 2.66 and the category-specific RBI for ***Engineering, Biomedical*** is higher at 3.63. Most papers (nearly two-thirds) are assigned to a single journal category whilst a minority are assigned to more than 5.

Average rebased impact

As noted above, the distribution of citations amongst papers is highly skewed, many papers are uncited and a very few papers accumulate extensive citation counts. Historically, research performance has been indexed using average citation impact (rebased as described to a world average that accounts for time and discipline).

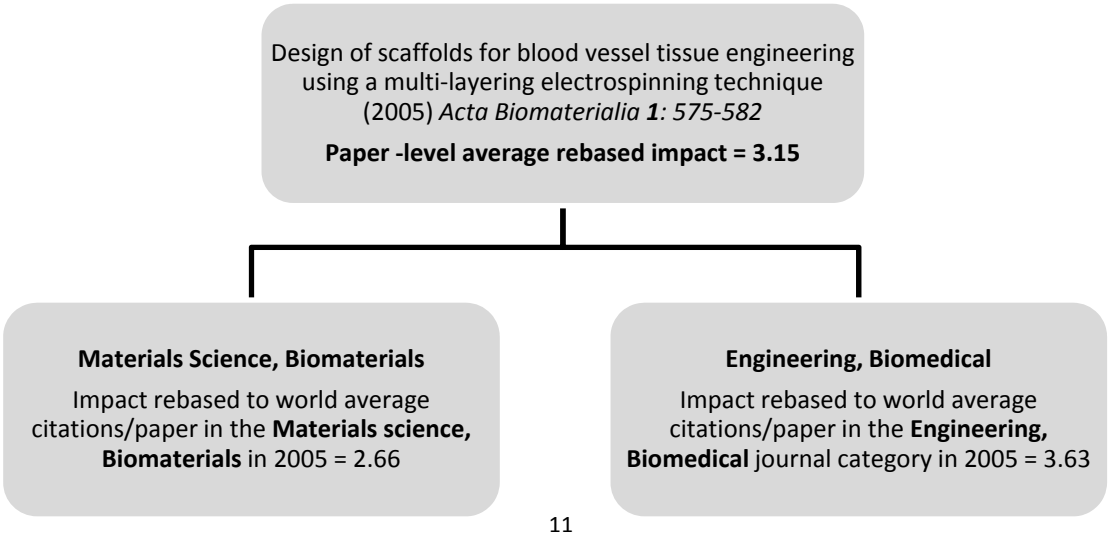
An average may be misleading, however, if assumptions are made about the distribution of the data beneath it. Almost all research activity metrics are skewed: many low performance values and a few exceptionally high values. In reality, therefore, the average impact tends to be significantly different from either the median or mode in the underlying distribution.

The average (rebased) impact can be calculated at an individual paper level where it can be associated with more than one journal category. It can also be calculated for a set of papers at any level from a single country to an individual researcher's output.

Thus, in the example below, the average RBI of the *Acta Biomaterialia* paper can be given as 3.15 (average of the two category-specific RBIs, 2.66 and 3.63)

Impact Profiles®

Evidence has developed a bibliometric methodology which shows the proportion of papers that are uncited and the proportion that lie in each of eight categories of relative citation rates, normalised (rebased) to world average. An Impact Profile® enables an examination and analysis of the strengths and weaknesses of published outputs relative to world average and relative to a reference profile. This provides much more information about the basis and structure of research performance than conventionally reported averages in citation indices.

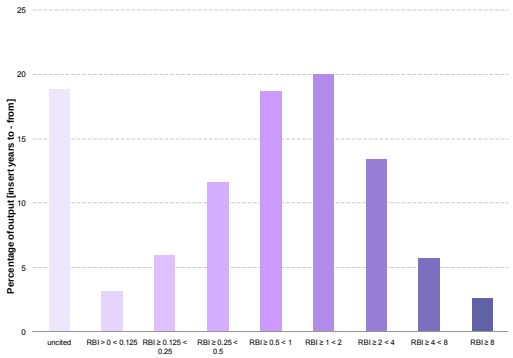


Papers which are "highly-cited" are defined as those with an average rebased impact (RBI) greater than or equal to 4.0 i.e. those papers which have received greater than or equal to four times the world average number of citations for papers in that subject published in that year.

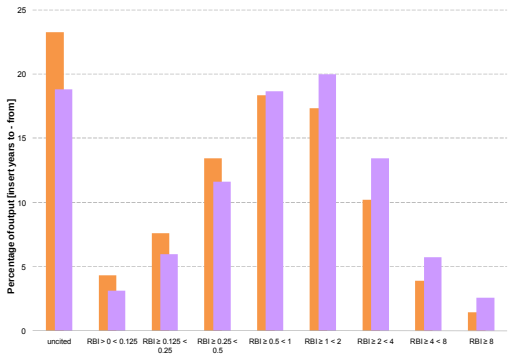
The proportion of uncited papers in a dataset can be compared to the benchmark for the UK, the USA or any other country. Overall, in a typical ten-year sample, around one-quarter of papers have not been cited within the 10-year period, the majority of these, of course, are those that are most recently published.

The Impact Profile® histogram can be presented in a number of ways which are illustrated below.

A



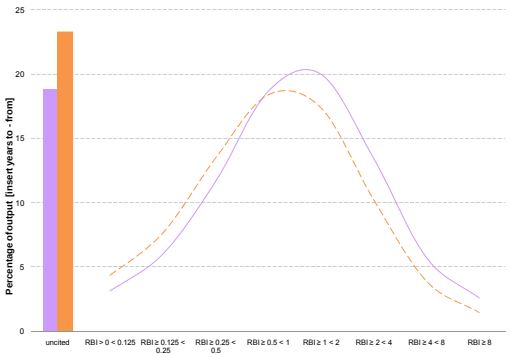
B



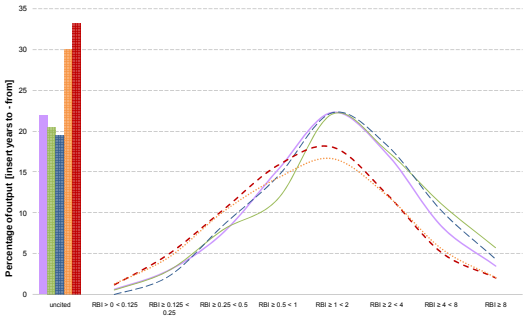
A: is used to represent the total output of an individual country, institution or researcher with no benchmark data, visually it highlights the numbers of uncited papers (weaknesses) and highly cited papers (strengths).

B & **C**: are used to represent the total output of an individual country, institution or researcher (**client**) against an appropriate benchmark dataset (**benchmark**). The data are displayed as either histograms (B) or a combination of histogram and profile (C). Version C prevents the 'travel' which occurs in histograms where the eye is drawn to the data most offset to the right, but can be less easy to interpret as categorical data.

C



D



D: illustrates the complexity of data which can be displayed using an Impact Profile®. These data show research output in defined journal categories against appropriate benchmarks: client, research field X; client, research field Y; client, research field Z; benchmark, research field X+Y; benchmark, research field, Z.

2.3 Data description

Dataset size affects year-to-year variability in impact (citations/paper), so samples with greater numbers of papers typically show less fluctuation. Citation data also tend to be highly skewed, with many zero-citation values and a few very high values. To address this, the report also uses the Impact Profile® methodology to examine the data using impact categories. If considered alongside the overall average rebased impact, the proportion of papers in each impact category can indicate whether the overall average is highly-slewed due to a few exceptionally-cited papers or whether the overall average is supported by a substantial body of well-cited papers.

Citation data provided by Thomson Reuters are assigned on an annual census date referred to as the Article Time Period. For the majority of papers the Article Time Period is the same as the year of publication but for a few papers (especially those published at the end of the calendar year in less main-stream journals) the Article Time Period may vary from the actual year of publication.

2.4 Data definitions

Publications: Thomson Reuters abstracted publications include editorials, meeting abstracts, book reviews as well as full journal articles.

Papers: is used in the context of this report to refer to publication types which are used in analyses these are substantive journal articles and reviews and exclude editorials, meeting abstracts and conference papers.

Citations: the citation count is the number of times that a given paper has been cited since it was published up to a given census date, in these analyses this date is December 2008.

Impact: impact is calculated by dividing the sum of citations by the sum of papers in the dataset (which for a single paper is its citation count). This can be done for papers within a specific research field such as Plant Sciences, or for a specific institution or group of institutions. Citation impact inevitably declines in the most recent years of any time-period as papers have less time to accumulate citations (papers published in 1996 will typically have more citations than papers published in 2001). Not all publication types are used in the calculation of impact: substantive journal articles and reviews are normally included but meeting abstracts and editorials are excluded.

Rebased Impact (RBI): rebased impact is raw impact normalised to the world average. For example, rebased impact for a sample in a given research field is the impact of the sampled papers in that field and year of publication divided by the impact for all world papers across that research field in the same year of publication.

Adams J, Gurney K and Marshall S (2007) Profiling citation impact: a new methodology. Scientometrics, 72 (2), 325-344.

Adams J (2005) Early citation counts correlate with accumulated impact. Scientometrics, 65 (3), 567-581.

Adams J and Smith D. (2002) Maintaining Research Excellence and Volume. A report to the Higher Education Funding Councils. http://www.hefce.ac.uk/pubs/rdreports/2002/rd08_02/

3 EPSRC-funded research publications

The publications for this report and bibliometric analysis were provided to *Evidence* as an Excel spreadsheet with the associated grant data received as an update to the original data. A total of 12,205 publications, collated from final reports to EPSRC grants received in 2003-2005, were listed in the dataset.

The publications data principally cover the years 1999 to 2008 though some publications were outside these years.

Individual publications were assigned, via the grant reference number; to funding mode, either responsive or targeted; and to grant type, either larger grants (defined as those with a value greater than £350K and Announced length greater than 36 months) or other grants.

3.1 EPSRC-funded papers for bibliometric analyses

The publications data were cleaned and matched, where possible, to Thomson Reuters citation databases. Table 3.1.1 details the outcome of this matching process. No attempt was made to manually match publications flagged by EPSRC as 'Not refereed'. Matching is typically done by using some of the bibliographic information such as journal title, volume and/or page numbers to citation databases. However, in this dataset almost one-third had to be matched by more manual means as the data quality was variable and the 'not found' numbers are relatively high.

Table 3.1.1: Outcome of matching EPSRC-funded publications to Thomson Reuters citation data

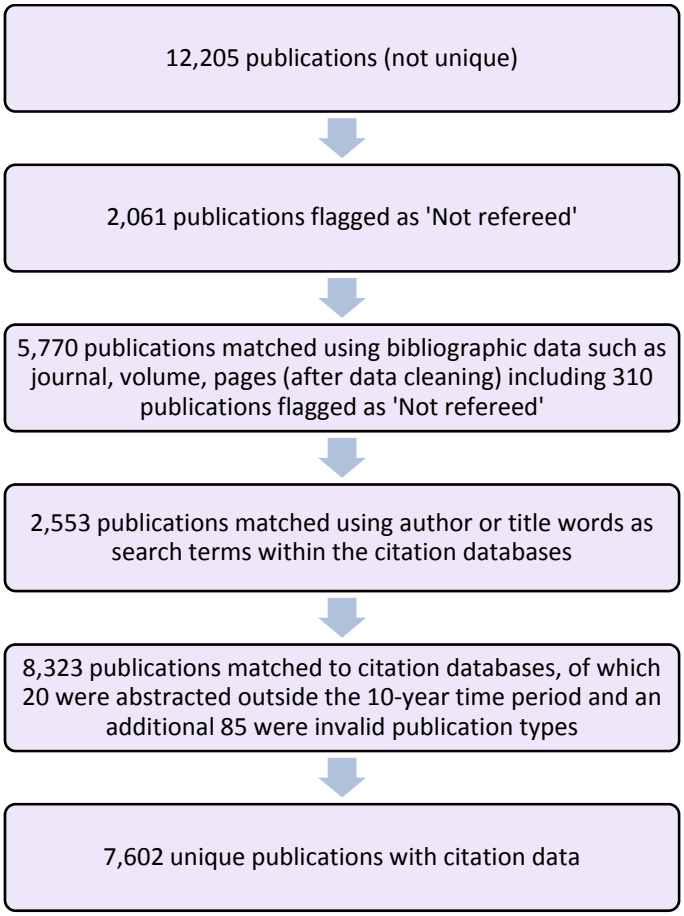
Outcome		Total number of papers
matched	Linked to Thomson Reuters citation data: some of these will be duplicates, invalid publication types or have been published outside the 10-year census period.	8,323
not abstracted	These journal titles are not abstracted by Thomson Reuters in the years of publication, also includes 4 books.	143
not found	These publications, despite being reported as in abstracted journals, could not be identified with a Thomson Reuters publication from the given information.	1,410
Invalid publication type	Includes those publications flagged as 'Not refereed' by EPSRC and others found during the manual matching process (for example editorials, conference proceedings papers).	2,329
Total		**12,205**

All papers from the EPSRC publications dataset were accounted for in the matching process.

The schematic diagram in Figure 3.1 outlines the process for matching the publications data from EPSRC to Thomson Reuters citation data. Initial matching was done using journal name, volume and pages; followed by a number of steps using various truncated forms of the article title along with some of the bibliographic data (where available). Finally, the databases held by *Evidence* were searched using article

title words in combination rather than taking the article title as submitted by EPSRC. For example, a paper titled "Strain stability in bioreactors degrading recalcitrant organic compounds" might be identified using search terms "stability" and "recalcitrant". The success of this final approach highlighted the numbers of papers with draft or inaccurate article titles in the publications dataset. The bibliometric analyses presented in this report will not cover conference proceedings, meeting abstracts, books, chapters in books or grey literature such as reports. It therefore captures only a specific part of the total output funded by EPSRC over the period, but this part is usually recognised as describing the most direct contribution to the research base.

Figure 3.1: Schematic for process of assigning Thomson Reuters citation data to EPSRC-funded publications



3.2 Description of the dataset to be used in bibliometric analyses

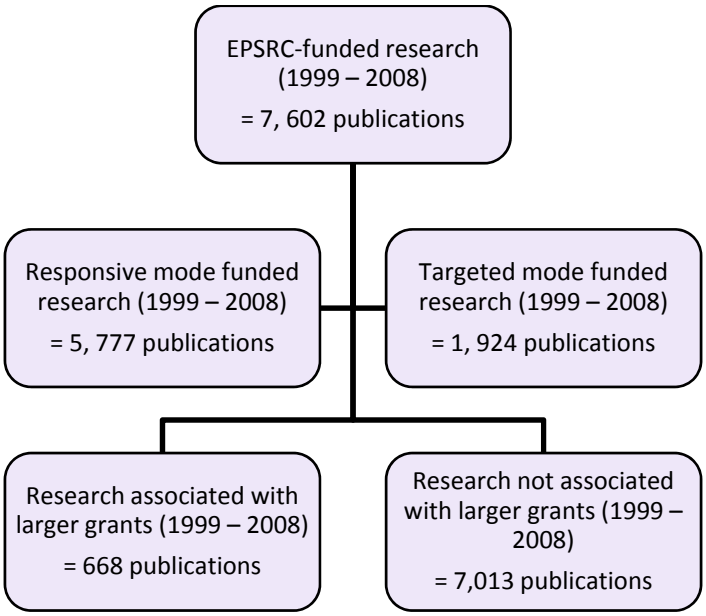
The structure of the data is such that there are multiple occurrences of papers within the collated database. In all, 571 papers were recorded more than once; most of these were duplicates but at the extreme end there was one paper with 5 occurrences.

As an example, there are four occurrences of a paper by Grattan KTV & Sun T (2002) *MRS Bulletin* **27**: 389-395; all from the Research Grants funding mechanism but associated with four different grants, twice for Targeted mode and twice for Responsive mode. For bibliometric analyses, this paper (and the citations to it) would be counted once for the Targeted funding Mode and once for the Responsive funding Mode. When considering the EPSRC-funded papers overall it would be counted once only.

A further example from the larger grant data would be the paper by Abou-Chakra H *et al*. (2003) Advanced Powder Technology **14**: 167-176 which has been associated with three different grants, two classified as larger, and one not. For bibliometric analyses, this paper (and the citations to it) would be counted once within the larger grant funded dataset and once for the other grant dataset. Again, when considering the EPSRC-funded papers overall it would be counted once only.

The datasets for the bibliometric analyses of EPSRC-funded research will use citation data from each journal article or review only once, both overall and within funding modes. The rationale behind this methodology is that the output from the Responsive funding mode is a dataset of *N* papers even if co-authored by more than one researcher supported by Responsive funding.

For reference the unique paper counts for each funding mode, by grant type, and for EPSRC-funded research overall are given below:



The matching rates between the Responsive and Targeted funding modes were very similar (around 75%) though the matching success of publications associated with larger grants was rather better than for publications arising from other grants (89% compared to 75%).

4 EPSRC-funded research – bibliometric indicators

Section 4 describes the basic characteristics of the EPSRC-funded publications dataset and compares it to those of the UK as a whole.

- The EPSRC-funded dataset used in this report comprises a total of 7,602 journal articles and reviews from 1,404 journals for the 10-year period between 1999 and 2008 (Section 4.1).
- Journals used most frequently by EPSRC-funded researchers are those which are leading journals in their respective research fields (Section 4.2). EPSRC-funded researchers have published a similar proportion of papers in the elite, multidisciplinary journals of Nature and Science as compared to the UK overall (0.44% compared to 0.42% papers published between 2002 and 2004).
- The range of journal categories in which EPSRC-funded research has published is diverse, with no one category predominating. However, the principal categories represent the broad research areas covered by EPSRC including Chemistry, Physics and Engineering and Materials Science with specialist areas within these like Optics and Polymer Science (Section 4.3).
- EPSRC-funded research papers have a good citation impact compared to UK and world baselines. Research papers published in Condensed Matter Physics journals and in journals assigned to Multidisciplinary Physics have particularly good citation impact, being more than twice the world average in each category (Section 4.5).

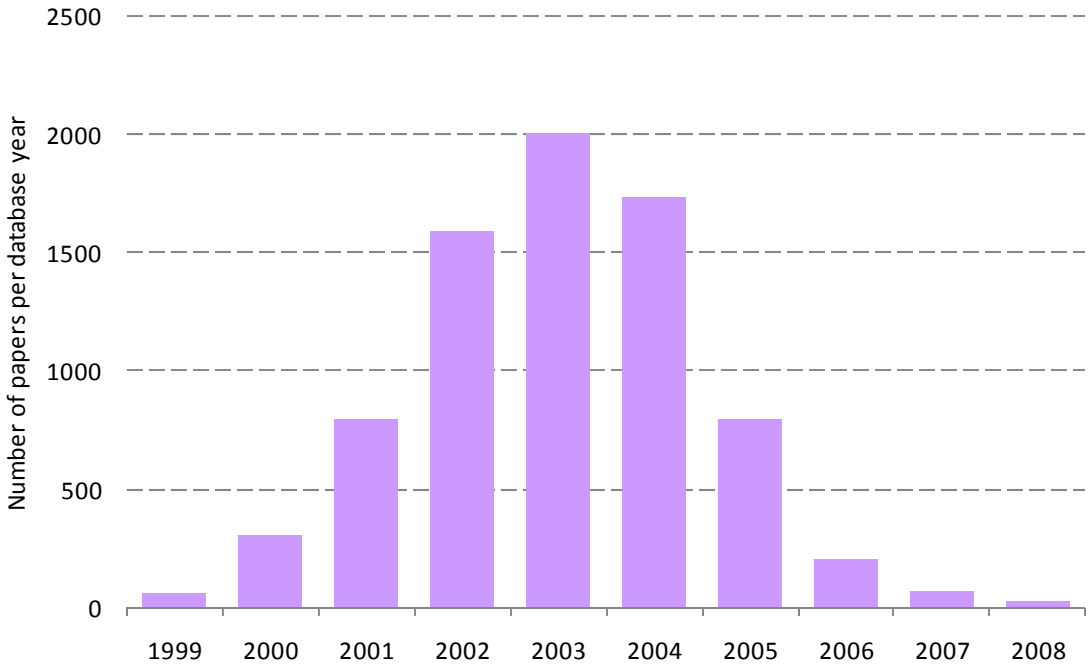
The indicators described in this section suggest that EPSRC-funded papers are, on average, of higher quality than those from the rest of the UK research base. Within this dataset, bibliometric indicators are presented for two funding modes, Responsive and Targeted, and for research associated with larger grants against that which is not. The Impact Profiles® of these papers is analysed further in Section 5.

- There is no evidence for any significant difference in research performance from the summary bibliometric indicators in Section 4.5 or the Impact Profiles® (Section 5.2) for papers arising from 'responsive' and ''targeted' funding modes.
- The citation impact of research arising from larger grants is much higher than any other EPRSC-funded research. This is shown by the summary bibliometric indicators in Section 4.5 and confirmed by the Impact Profiles® (Section 5.3). These indicators reflect a consistent excellent performance not resting on a few extremely highly-cited papers.

  This effect may have been exacerbated by a change in the data returned to EPSRC in final reports after 2003 whereby final reports before this date recorded all published research associated with a grant and in 2004 onwards, five selected publications were returned. Typically, however, such initiatives tend to aid discrimination between research performance rather than hinder it and the citation analysis of selected papers is the preferred model for the UK Research Excellence Framework (HEFCE publications 2009/38 Research Excellence Framework: Second consultation on the assessment and funding or research[1]; 2009/39 Report of the pilot exercise to develop bibliometric indicators for the Research Excellence Framework[2]).

[1] http://www.hefce.ac.uk/pubs/hefce/2009/09_38/
[2] http://www.hefce.ac.uk/pubs/hefce/2009/09_39/

4.1 Annual paper output



The figure indicates the annual paper output from EPSRC-funded research included in this analysis. This shows that the dataset is not spread evenly over the 10-year period, though the distribution is entirely appropriate for publications collated from final reports received in 2003-2005 as much of the research will have taken place in the years 2001-2005.

However, this distribution indicates that it would be inappropriate to use a 10-year benchmark from the UK data so we have created new customised benchmark data for the UK using 2002-2004 data.

4.2 Journal usage

The twenty journals used most frequently by EPSRC-funded researchers are listed in Table 4.2a.

Papers in these more commonly-used journals total just over 2,000 papers or one-quarter of the total output. The majority of these journals are ranked in the 'top' 25% (by Journal Impact Factor) of journals in their specific research fields (where the 2008 Impact Factor is in bold). This indicates that the journals used by EPSRC-funded researchers contain papers that are well-regarded amongst their peers.

EPSRC-funded researchers have published a similar proportion of papers in the elite, multidisciplinary journals of Nature and Science as compared to the UK overall (0.44% compared to 0.42% papers published between 2002 and 2004).

Table 4.2a: Journals in which EPSRC-funded researchers have published most frequently

Journal Title (ranked by total papers 1999-2008)	Number of papers	Impact Factor 2008
Physical Review Letters	221	**7.180**
Physical Review B	216	**3.322**
Chemical Communications	195	**5.340**
Applied Physics Letters	177	**3.726**
Journal of Applied Physics	118	**2.201**
Journal of Chemical Physics	111	**3.149**
Journal of the American Chemical Society	102	**8.091**
Journal of Materials Chemistry	90	**4.646**
Electronics Letters	88	1.140
Journal of Physics-Condensed Matter	75	1.900
Angewandte Chemie-International Edition	74	**10.879**
Tetrahedron Letters	69	2.538
Langmuir	68	**4.097**
Dalton Transactions	66	**3.580**
Journal of Physical Chemistry B	62	**4.189**
Macromolecules	60	**4.407**
Physical Chemistry Chemical Physics	57	**4.064**
Physical Review A	56	**2.908**
Physical Review C	55	**3.124**
Optics Letters	53	**3.772**

The journal impact factor is calculated by Thomson Reuters as the average number of times articles from the journal published in the past two years were cited in 2008. Thus, a journal impact factor of 2.0 means that, on average, the articles published in 2006 or 2007 have been cited twice. Citing articles may be from the same journal; most citing articles are from different journals.

Table 4.2b lists the twenty journals with the highest journal impact factors used more than once by EPSRC-funded researchers. The list is dominated by 'review' journals as articles in such journals tend to be cited more frequently than articles in research journals. It also includes journals which may not reflect 'core' research activity for EPSRC-funded researchers, for example, medical journals such as Arthritis and Rheumatism.

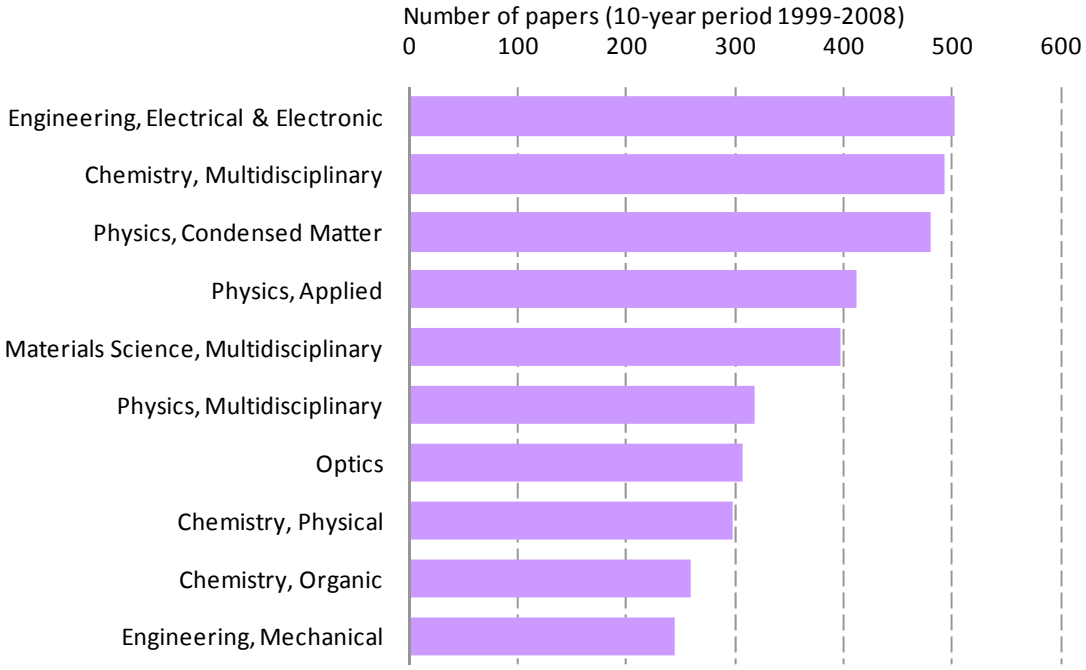
Table 4.2b: High impact factor journals used by EPSRC-funded researchers

Journal Title (ranked by journal impact factor for 2008)	Number of papers	Impact Factor 2008
Nature	21	31.434
Science	13	28.103
Chemical Reviews	5	23.592
Nature Materials	8	23.132
Progress in Materials Science	2	18.132
Chemical Society Reviews	7	17.419
Circulation	2	14.595
Angewandte Chemie-International Edition	74	10.879
Current Biology	2	10.777
Coordination Chemistry Reviews	3	10.566
Nano Letters	12	10.371
Circulation Research	2	9.989
Proceedings of the National Academy of Sciences USA	5	9.380
Advanced Materials	26	8.191
Journal of the American Chemical Society	102	8.091
Physical Review Letters	221	7.180
Nucleic Acids Research	3	6.878
Advanced Functional Materials	10	6.808
Arthritis and Rheumatism	2	6.787
Biomaterials	12	6.646

4.3 Most frequently used research fields

Papers are allocated by Thomson Reuters to one or more research fields according to which journal the paper is published in (Section 2.2). Research published by EPSRC-funded researchers has been assigned to more than 150 of the 251 Web of Science journal categories or research fields used by Thomson Reuters.

The chart below shows the 10 most frequently used journal categories with paper numbers for the 10-year period.



These journal categories represent the broad research areas covered by funding from EPSRC of Chemistry, Physics and Engineering and Materials Science with specialist areas within these like Optics and Polymer Science.

The chart above is based on one, unique journal category for each paper – this category is selected as the one in which the paper performs best when ranked by the citation count within the category. However, bibliometric analyses using journal categories (as in Section 4.4) will have used all papers which can be assigned to the specific journal category – this will be higher than the numbers above due to assignment of some papers to more than one category. For example, there are 446 papers which have been published in Optics journals – of these 306 have been assigned to Optics as their unique 'best' journal category, the others have been assigned to unique categories such as Telecommunications.

4.4 Citation impact of EPSRC-funded research

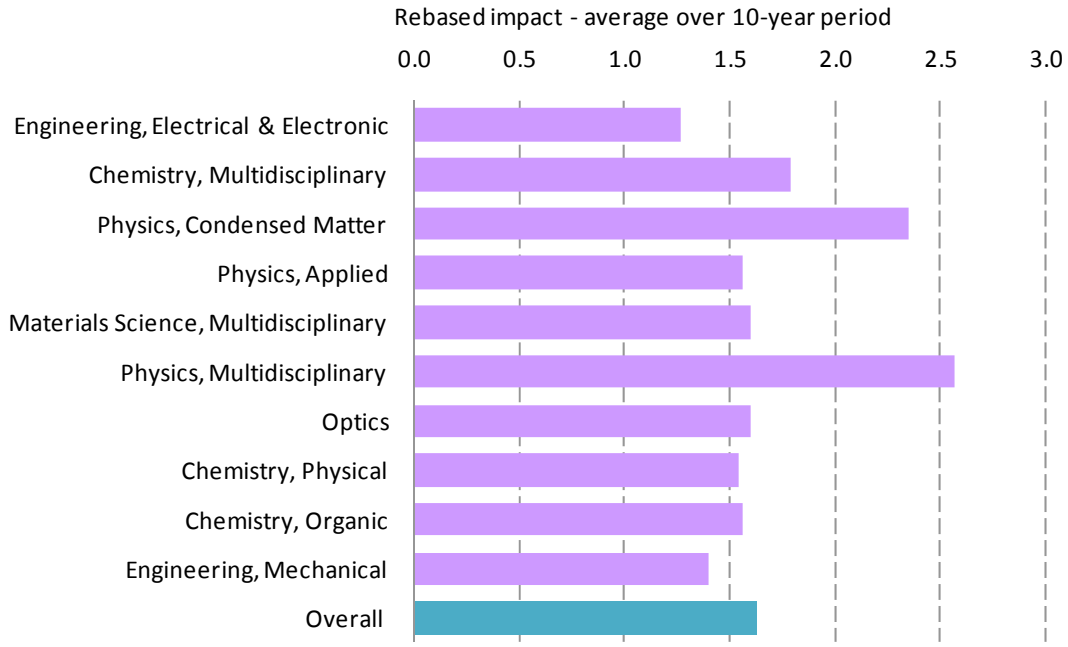
The relative citation impact of EPSRC-funded research papers is good.

Impact of research, an index linked to citation accumulation, is field-dependent. All data presented in this report are therefore impact normalised, or rebased, to the relevant field world average. This then allows appropriate comparison between years and between fields.

For example, the rebased impact (RBI) for EPSRC-funded research in Optics is the raw impact (citations/paper) of EPSRC papers in Optics for any specific year of publication divided by the raw impact (citations/paper) for all world papers in Optics and in the same year of publication.

The overall rebased impact (RBI) of all of the EPSRC-funded research papers is 1.63 (where world average is 1.0). For comparison, the UK's average rebased impact relative to world baseline for the 5-year period 2003-2008 inclusive in physical sciences was 1.39 and in engineering was 1.05 (*Evidence* report for the Department of Business, Innovation, and Skills, **International comparative performance of the UK research base**, indicators 1.10.06 and 1.10.07).[3]

The chart below shows the rebased impact (average for the 10-year period between 1999 and 2008) for the EPSRC-funded papers in their most frequently used journal categories.



The figure shows that the average rebased impact of EPSRC-funded research is above world average in all fields.

The impact of papers published in Multidisciplinary Physics and Condensed Matter Physics journals is more than twice the world average.

[3] http://www.dius.gov.uk/science/science_funding/science_budget/~/media/publications/I/ICPRUK09v1_4

4.5 Bibliometric indicators for EPSRC-funded research by funding mode and by grant type

This section provides summary bibliometric indicators for EPSRC-funded research benchmarked against a customised benchmark of research papers in similar research fields (physical sciences and engineering) for the UK, excluding those papers identified as being funded by the EPSRC, for the years 2002-2004.

The following indicators have been calculated:

- Percentage of papers which are uncited;
- Percentage of papers cited more than expected for the journal/year of publication;
- Overall rebased citation impact of research papers;

	Number of papers in analysis	Percentage of papers which are uncited	Percentage of papers which are cited more than expected	Overall average rebased impact 2001-2008
EPSRC-funded research, overall	7,602	7.2%	43.1%	1.63
EPSRC-funded research, Responsive	5,777	6.9%	42.4%	1.62
EPSRC-funded research, Targeted	1,924	8.1%	44.9%	1.65
EPSRC-funded research, larger grants	668	5.4%	41.5%	2.11
EPSRC-funded research, other grants	7,013	7.3%	43.3%	1.58
UK physical sciences and engineering, 2002-2004 excluding EPSRC-funded papers	66,567	16.2%	35.1%	1.19

Percentage of papers which are uncited

EPSRC-funded research, overall or by funding mode or grant type, is more likely to be cited than UK physical sciences and engineering research. EPSRC-funded research has less than half as many uncited papers as the rest of the UK.

Percentage of papers which are cited more than expected

This indicator is journal-specific and compares the citation count for a paper with the average expected for the type of paper, journal and year (or volume in some cases). A high value in this indicator in combination with a low citation impact value can indicate that papers are being submitted to journals which are either not typical of the journal category or are not well-regarded by the peer group. However, a high value indicates that the papers are being comparatively well-cited.

Overall average rebased impact

EPSRC-funded research, overall or by funding mode or grant type, has a higher citation impact than similar UK research in physical sciences and engineering. Within the funding mode and grant analyses, the citation impact of research arising from larger grants is much higher than any other EPRSC-funded

research. This is confirmed by the Impact Profile® (Section 5.3) and reflects a consistent excellent performance not resting on a few extremely highly-cited papers.

Research activity and performance data are almost invariably highly skewed. That is to say, whether we look at income per researcher, PhDs awarded per unit or per staff, publications per grant-holder or citations per paper, there are relatively many incidences of data points of low value (little relative income, few research students, few publications, uncited papers) and relatively few incidences of high value.

The consequence is that the data must be transformed in order to get an accessible picture that can be more readily interpreted and compared with related distributions. For citation impact data, this can be done with the Impact Profile® methodology.

For statistical analyses we are presented with a related problem. The citation impact data are not normally distributed, that is to say: they do not without transformation fit a bell-shaped curve. The mid-point of the distribution (the median) is nowhere near the average value; indeed the average is usually much greater than the median. Because of this, *Evidence* (in common with most other scientometric analysts) has not usually applied standard parametric statistics and does not seek to express the impact for any sample in terms of its difference to a benchmark (e.g. analysing whether the performance of a named institution is statistically significantly different from the national average).

That said, and with caution, non-parametric statistics can be used to compare samples (e.g. Mann-Whitney U) or with very large samples we can assume an approximation that allows us to use standard parametric comparisons (e.g. t-test). Applying this to the EPSRC data samples, we can acquire some information regarding the differences between the average rebased impact associated with papers produced by different funding modes and grant types.

		Funding mode		**Grant type**	
		Targeted	**Responsive**	**Larger grants**	**Other grants**
Articles	Number in sample	1,924	5,777	668	7,013
Citations - rebased impact	Average	1.65	1.62	2.11	1.58
	Median	0.92	0.96	1.09	0.94
	Inter-quartile mean	1.09	1.15	1.27	1.12
	Significance of difference	0.73, n.s.		P < 0.001	

The tests applied here confirm that there is no significant difference in citation impact between the publications produced from responsive and targeted mode grants but that larger grants are associated with papers that are on average of higher impact. There is a highly significant difference between the impact distribution for such grants and the rest of the data.

In recent years, investigators have been required to submit only a selection of their higher impact publications to EPSRC. It would be expected that a larger, longer project grant would produce relatively more publications than other projects. This would create a greater pool from which selection might be made. This is, of course, also the case for the more research-active and productive departments that submit to the Funding Councils' research assessment processes. *Evidence* has shown elsewhere that selective submission of publications tends to increase the differentiation between higher and lower performing units. When total output portfolios are examined then excellence may be diluted by the mediocre. When selected publications are examined then it becomes evident that some units are able to produce a sample of exceptional and consistent quality. EPSRC's current policy of selective submission would therefore lend itself to enabling this differentiation rather than obscuring differences.

What if we consider the broader picture, comparing the papers associated with EPSRC funding with the rest of the UK's output in physical sciences and engineering? There is a substantial volume of such publications not included in the dataset evaluated here.

The 7,602 EPSRC papers were compared with the UK's non-EPSRC output in physical sciences and engineering for the three years 2002-2004, i.e. the centre of the EPSRC time distribution, giving a sample of 66,567 papers.

The paper-level average rebased citation impact (RBI) was calculated for each paper. The RBI distributions are extremely positively skewed. A log transform was applied to the data, from which transformed data-set confidence limits were calculated. With such a large reference sample some slightly greater level of confidence in the statistical outcomes might then be felt.

For the EPSRC *vs* UK comparison the statistical difference, using such transformed data, appears to be very highly significant ($P < 0.001$) whichever single reference year is considered, let alone the dataset as a whole. The untransformed average impact of the EPSRC data was 1.63 whereas the UK average is 1.23, 1.14 and 1.19 in the three relevant years.

Evidence remains reluctant to place reliance on statistical tests as part of the evaluation of these research performance data. One reason is the belief that if illustrations of the data do not provide sufficiently convincing information to allow an informed and expert group to draw a conclusion as to whether two data samples are 'different' for research management and policy purposes then they are probably not different. A statistical test would not increase the panel's confidence if they could not already see such a difference emerging.

A further concern would be that legitimisation of statistical differentiation at a high, policy level would percolate down. In due course, individual institutions would start to apply such tests to the performance of a research group compared to an institutional or national average. This reductionist approach would not only be ill-advised but almost certainly lead to erroneous outcomes.

5 Impact Profiles® of EPSRC-funded papers

Impact Profiles® enable an examination and analysis of the balance of published outputs relative to world average and relative to a reference profile. This provides much more information about the basis and structure of research performance than conventionally reported averages in citation indices.

An Impact Profile® shows what proportion of papers are uncited and what proportion are in each of eight categories of relative citation rates, normalised (rebased) to world average (which becomes 1.0 in this graph). Rebased citation rates above 1.0 indicate papers cited more often than world average for the field in which that journal is categorised and in their year of publication.

Attention should be paid to:

- The proportion of uncited papers on the left of the chart
- The proportion of cited papers either side of world average (1.0)
- The location of the most common (modal) group near the centre
- The proportion of papers in the most highly-cited categories to the right, (≥4 x world, ≥8 x world).

What are uncited papers?

It may be a surprise that some journal articles are never subsequently cited after publication, even by their authors. This accounts for about half the total global output and almost one quarter of UK output. We cannot tell why papers are not cited. It is likely that a significant proportion of papers remain uncited because they are reporting negative results which are an essential matter of record in their field but make the content less likely to be referenced in other papers. Inevitably, other papers are uncited because their content is trivial or marginal to the mainstream or plain wrong. It should not be assumed that this is the case for all such papers.

There is variation in non-citation between countries and between fields. On the whole, relatively more engineering papers tend to remain uncited than papers in other sciences, indicative of a disciplinary factor as well as a quality/significance factor. There is also an obvious increase in the likelihood of citation over time but most papers that are going to be cited will be cited within a few years of publication.

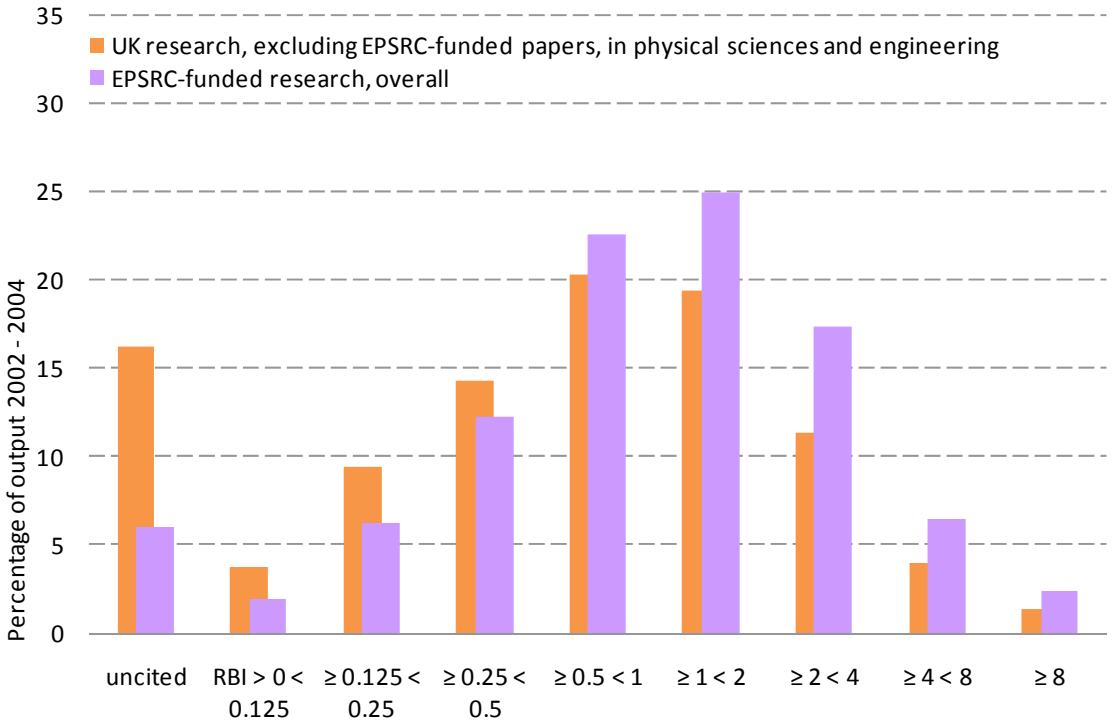
We work on the assumption that relative non-citation rates within a field are one of the indicators of the extent to which a body of work is regarded by others in the same field to be of greater or lesser significance to their subsequent work.

What is the threshold for 'highly cited'?

Thomson Reuters has traditionally used the term 'Highly Cited Paper' to refer to the world's 1% of most frequently cited papers, taking into account year of publication and field. In rough terms, UK papers cited more than 8 times as often as relevant world average would fall into the Thomson Highly Cited category. About 1-2% of papers (all papers, cited or uncited) typically pass this hurdle. Such a threshold certainly delimits exceptional papers for international comparisons but, in practice, is an onerous marker for more general management purposes.

After reviewing the outcomes of a number of analyses, we have chosen a more relaxed definition for our descriptive and analytical work. We deem papers that are cited more often than 4 times the relevant world average to be relatively highly-cited for national comparisons. This covers the two most highly-cited categories in our graphical analyses. About 5% of total UK papers typically pass this hurdle.

5.1 Impact Profile® for EPSRC-funded research benchmarked against UK physical sciences and engineering

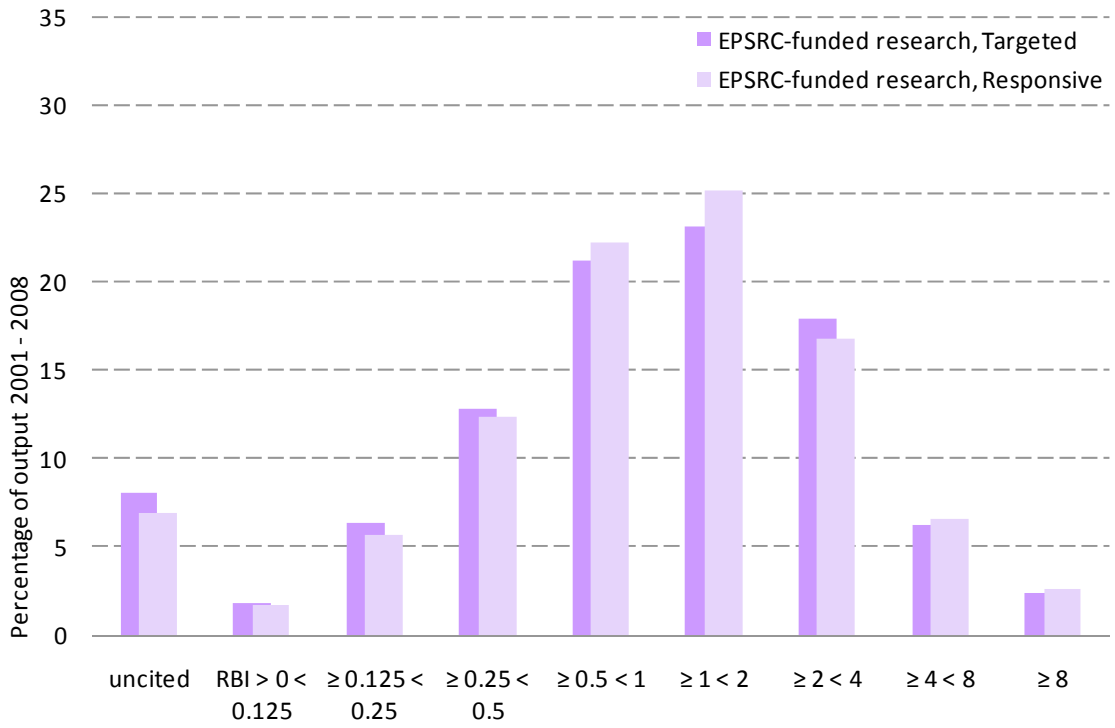


- EPSRC-funded research is internationally competitive - almost one-tenth (8.8%) of papers are highly-cited (cited more than 4 times the relevant world average).
- The modal group for EPSRC-funded research is above world average (1.0) at RBI 1–2 whereas it is below world average for UK physical sciences and engineering research.
- In all impact categories above world average (to the right of the figure), EPSRC-funded papers perform better than the rest of the UK, indicating that research funded by EPSRC is much more frequently cited than similar papers from the rest of the UK.
- Relatively few of the EPSRC-funded papers are uncited. (The EPSRC data used in the Impact Profile has been limited to 2002-2004 to allow ready comparison with the UK background).

	EPSRC-funded	UK
Number of papers in analysis	5,263	66,657
Percentage of papers which are uncited	6.0%	16.2%
Percentage of papers which are highly cited (RBI ≥4.0)	8.8%	5.4%
Percentage of papers cited more than world average	51.0%	36.1%
Overall average rebased impact 2002-2004	1.62	1.19

The following Impact Profiles® show EPSRC-funded research over the 10-year period by funding mode and by grant type.

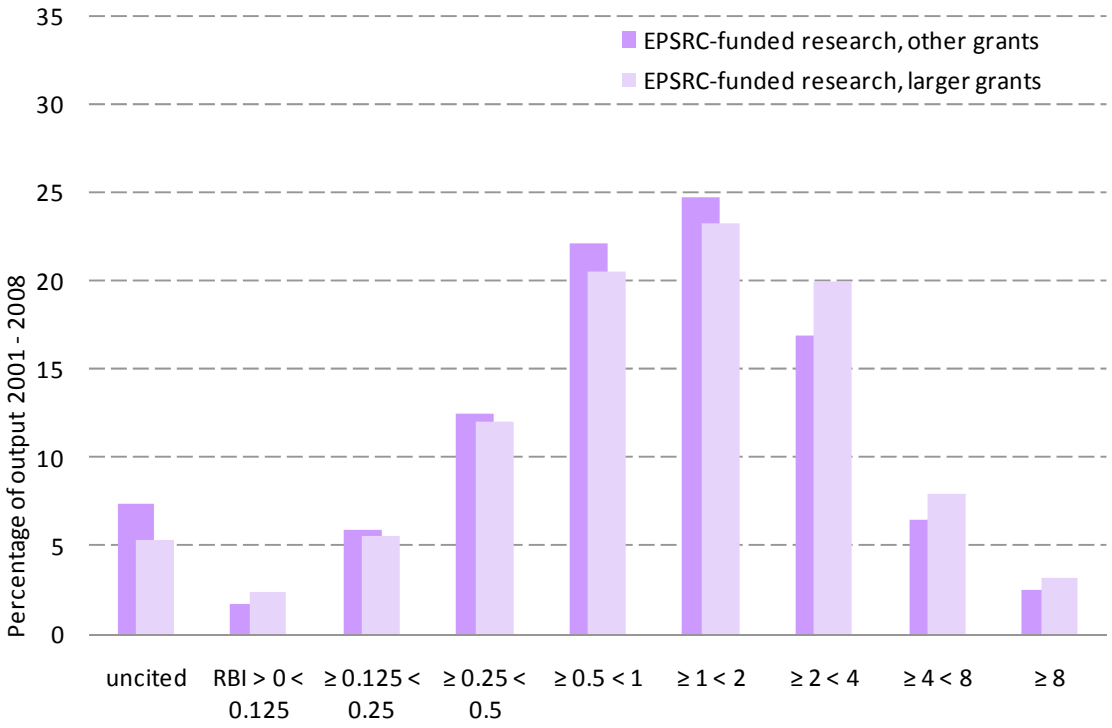
5.2 Impact Profiles® for EPSRC-funded research: Responsive and Targeted funding modes



- Research published by EPSRC-funded researchers receiving funding through the Targeted mode has had similar citation impact to research published through the Responsive mode.
- 8.6% of papers arising from Targeted mode grants were highly-cited (cited ≥ 4 times the relevant world average) compared to 9.2% for Responsive mode grants.
- The modal group for both funding modes was above world average (1.0) at RBI 1–2. The proportion of cited papers above the world average (1.0) differs by only 1% (55% for Responsive mode and 54% for Targeted mode).
- 8.1% of papers arising from Targeted mode grants were not cited (by the end 2008) compared to 6.9% for Responsive mode grants.

	EPSRC-funded research	
	Targeted	**Responsive**
Number of papers in analysis	1,924	5,774
Percentage of papers which are uncited	8.1%	6.9%
Percentage of papers which are highly cited (RBI ≥4.0)	8.6%	9.2%
Percentage of papers cited more than world average	49.7%	51.2%
Overall average rebased impact 2001-2008	1.65	1.62

5.3 Impact Profiles® for EPSRC-funded research: grant type



- Research published by EPSRC-funded researchers receiving larger grants has had greater citation impact than similar research published from other grants – the Impact Profile® above shows a clear and consistent shift to the higher impact categories to the right for larger grants compared to other grants.
- Just over 11% of papers arising from larger grants were highly-cited (cited ≥ 4 times the relevant world average) compared to 8.9% for other grants (and compared to 8.8% overall).
- The modal group for both groups was above world average (1.0) at RBI 1–2. The proportion of cited papers above the world average (1.0) is higher for larger grant funded research than other research (57.3% compared with 54.5%).
- There is a lower percentage of uncited larger grant funded research than for other grant funded research (5.4% compared with 7.3%).

	EPSRC-funded research	
	Larger grants	**Other grants**
Number of papers in analysis	668	7,013
Percentage of papers which are uncited	5.4%	7.3%
Percentage of papers which are highly cited (RBI ≥4.0)	11.1%	8.9%
Percentage of papers cited more than world average	54.2%	50.5%
Overall average rebased impact 2001-2008	2.11	1.58

Annex 1: Thomson Reuters definitions for the ten most frequently used journal categories for EPSRC-funded research

A.1 Engineering, Electrical & Electronic

Journals in the Engineering, Electrical & Electronic category deal with the applications of electricity, generally those involving current flows through conductors, as in motors and generators. This category also includes journals that cover the conduction of electricity through gases or a vacuum as well as through semiconducting and superconducting materials. Other relevant topics in this category include image and signal processing, electromagnetics, electronic components and materials, microwave technology, and microelectronics.

A.2 Chemistry, Multidisciplinary

The Chemistry, Multidisciplinary category includes journals having a general or interdisciplinary approach to the chemical sciences. Special topic chemistry journals that have relevance to many areas of chemistry are also included in this category. Journals having a primary focus on analytical, inorganic and nuclear, organic, physical, or polymer chemistry are placed in their own categories.

A.3 Physics, Condensed Matter

The category of Physics, Fluids & Plasmas covers journals on the kinetic and transport theory of fluids, the physical properties of gases, and the physics of plasmas and electric discharges. This category may include journals on nuclear fusion.

A.4 Physics, Applied

The journals in the Physics, Applied category deal with the applications of condensed matter, optics, vacuum science, lasers, electronics, cryogenics, magnets and magnetism, acoustical physics, and mechanics. This category also may include journals on physics applications to other sciences, engineering, and industry.

A.5 Materials Science, Multidisciplinary

The category of Materials Science, Multidisciplinary covers journals that have a general or multidisciplinary approach to the study of the nature, behaviour, and use of materials. Relevant topics include ceramics, composites, alloys, metals and metallurgy, nanotechnology, nuclear materials, and adhesion and adhesives.

A.6 Physics, Multidisciplinary

The Physics, Multidisciplinary category covers journals that have a general or interdisciplinary approach to physics. This category also includes theoretical and experimental physics as well as special topics that have relevance to many areas of physics.

A.7 Optics

Journals in the Optics category deal with the genesis and propagation of light, the changes that it undergoes and produces, and other phenomena closely associated with it. This category covers subject areas such as lasers and laser technology, infrared physics and technology, microwave technology, quantum optics, lightwave technology, fibre optics, opto-electronics, and photonics. Journals on photometry and luminescence are also included in this category.

A.8 Chemistry, Physical

The category Chemistry, Physical includes journals on photochemistry, solid state chemistry, kinetics, catalysis, quantum chemistry, surface chemistry, electrochemistry, chemical thermodynamics, thermophysics, colloids, fullerenes, and zeolites.

A.9 Chemistry, Organic

The Chemistry, Organic category includes journals that focus on synthetic and natural organic compounds their synthesis, structure, properties, and reactivity. Research on hydrocarbons, a major area of organic chemistry, is included in this category.

A.10 Engineering, Mechanical

The category Engineering, Mechanical includes journals on the generation, transmission, and use of heat and mechanical power, as well as with the production and operation of tools, machinery, and their products. Topics in this category include heat transfer and thermodynamics, fatigue and fracture, wear, tribology, energy conversion, hydraulics, pneumatics, microelectronics, plasticity, strain analysis, and aerosol technology.