Report on the citation database for the Human Frontier Science Program

Bibliometric baselines

October 2009

Contact details

The reporting organisation is

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 on the citation database provided for the Human Frontier Science Program (HFSP), covering research publications by researchers funded by the HFSP over a ten-year period between 1999 and 2008.
- Most frequently used journals by HFSP-supported researchers are those which are leading journals in their respective research fields.
- Nearly one-third of HFSP-supported research between 1999 and 2008 was published in journals assigned to the Biochemistry & Molecular Biology research field. Overall, HFSP-supported research was more likely to have been cited than comparable research published in Biochemistry & Molecular Biology journals worldwide.
- HFSP-supported research publications have an extremely good impact compared to world baselines. The average rebased impact of all HFSP-supported research publications = 3.03 (world average = 1.0, world average for Biochemistry & Molecular Biology = 1.38. More detailed analyses have shown that the average rebased impact of HFSP-supported research is above world average in all fields. In particular, the impact of publications in Plant Sciences, Genetics & Heredity and Immunology are outstanding.
- Throughout the 10-year period covered by this report the average impact of HFSP-supported research has been consistently well above the world average.
- The overall Impact Profile® confirms these baseline bibliometric indicators showing a strong shift towards the high impact categories consistent with performance achieved by world-class institutions.
- The Impact Profile® methodology has been used to benchmark HFSP-supported research in Biochemistry & Molecular Biology against similar research published by co-authors in the UK and US, which is typically very highly cited. This analysis has confirmed that HFSP-supported research is exceptionally well regarded worldwide in this field.

2 Introduction

2.1 The Human Frontier Science Program

The Human Frontier Science Program (HFSP) supports novel, innovative and interdisciplinary basic research focused on the complex mechanisms of living organisms; topics range from molecular and cellular approaches to systems and cognitive neuroscience. A clear emphasis is placed on novel collaborations that bring biologists together with scientists from fields such as physics, mathematics, chemistry, computer science and engineering to focus on problems at the frontier of the life sciences.

Research Grants are awarded for novel collaborations between teams of scientists from different countries to enable them to combine their expertise to approach questions that could not be answered by individual laboratories.

Emphasis is placed on novel collaborations that bring together scientist from different disciplines (e.g. from chemistry, physics, computer science, engineering) to focus on problems in the life sciences.

Two types of Research Grant are currently available: Young Investigators' Grants and Program Grants.

Young Investigators' Grants are awarded to teams of researchers, all of whom are within the first five years after obtaining an independent laboratory (e.g. Assistant Professor, Lecturer or equivalent).

Program Grants are awarded to teams of independent researchers at any stage of their careers. The research team is expected to develop new lines of research through the collaboration.

In addition, researchers may be supported individually with Fellowships: Long-term Fellowships provide young scientists with up to three years of postdoctoral research training in an outstanding laboratory in another country. Long-Term Fellows who return to their home country at the end of their Fellowship are eligible to apply for a Career Development Award.

The citations database covers research publications arising from both research grant and Fellowship funding, principally for the 10-year period between 1999 and 2008.

The HFSP Trust has commissioned *Evidence*, subcontracted to the Manchester Institute of Innovation Research (MIoIR), to provide bibliometric data linked to the publications database created by HFSP for this project. Summary bibliometric analyses will also be provided.

This report provides background bibliometric indicators for the publications associated with HFSP funding in the 10-year period between 1999 and 2008.

2.2 HFSP research publications database

The HFSP research publications database was created specifically for this project (thanks to Martin Reddington, Director of Scientific Affairs and Communications, HFSP). In summary, research publications were initially identified using Google Scholar for the years 1998 to 2009 using the search terms: HFSP OR HFSPO OR "Human Frontier" OR "Human Frontiers". In addition, research publications were identified from Web of Science by searching for names of grant, Fellowship and CDA awardees.

The resulting database was then visually screened for overlapping results to identify research publications by HFSP-funded researchers which acknowledged support from the HFSP. Each identified 'overlapped' publication was flagged in Endnote to ensure that the ISI ID numbers and the HFSP reference numbers in the LABEL field were maintained.

"HFSP" was added to the CALL NUMBER field to identify those publications which contained a reference to HFSP in the full text of the article. Some papers may have been missed if Google Scholar did not have full coverage of all the articles in the journals found in Web of Science.

Duplications may occur where a laboratory with a grant also has a Fellow. In addition, in

the Fellowship data there may be apparent duplications if several Fellows were co-authors. An example would be the laboratory of Elaine Fuchs at Rockefeller, who sponsors a successful candidate each year who are encouraged to collaborate. Consequently, the same record may occur several times with a different reference number in the LABEL field.

The program-specific EndNote databases detailed below were extracted from the above file:

RG_Collab_journals_final: contains the Web of Science records, with grant reference number and, where relevant, "HFSP" in the CALL NUMBER field. The searches were based on co-authorships among team members. The search was performed on all years since 1991 to enable estimation of the degree of previous collaboration before the award was made. This resulted in 3642 records, 767 of which overlapped with Google Scholar. Note that the early years use RG in the ref numbers, later years either RGY (for young investigator grants) or RGP (for program grants). For analysis over the whole period, RG should be considered equivalent to RGP.

LT_journals_final: contains the Web of Science records with the Fellowship reference number and, where relevant, "HFSP" in the CALL NUMBER field. The homonym problem were minimised by searching on the combination of au=Fellow AND Host supervisor. Articles published by the Fellow alone will have been missed but these are expected to be a rarity during the postdoctoral period. The database contains 4514 records, 1960 of which overlapped with Google Scholar. It will contain all collaborative papers, even those that were sustained beyond the Fellowship. As mentioned above the same article may occur several times if more than one Fellow is a co-author.

CDA_journals_final: contains the Web of Science records with the CDA reference number and, where relevant, "HFSP" in the CALL NUMBER field. As CDAs are all former Fellows, Web of Science was searched for publications WITHOUT the host supervisor(s) and from the year of the CDA award. The homonym problem was only manageable due to the lower numbers of CDAs. A question mark at the end of the reference number indicates some uncertainty over whether the author is, in fact, a CDA holder.

A note on Fellowship reference numbers: since award year 2005 the suffix "-C" has referred to the cross-disciplinary Fellowships and "-L" to the long-term Fellowships. The suffix "-C" was used previously as "combined" after the neuroscience ("-B") and molecular biology ("-M") fields were merged. Consequently, for analysis all awards made in 2004 and previously can be considered equivalent to "-L".

2.3 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'.

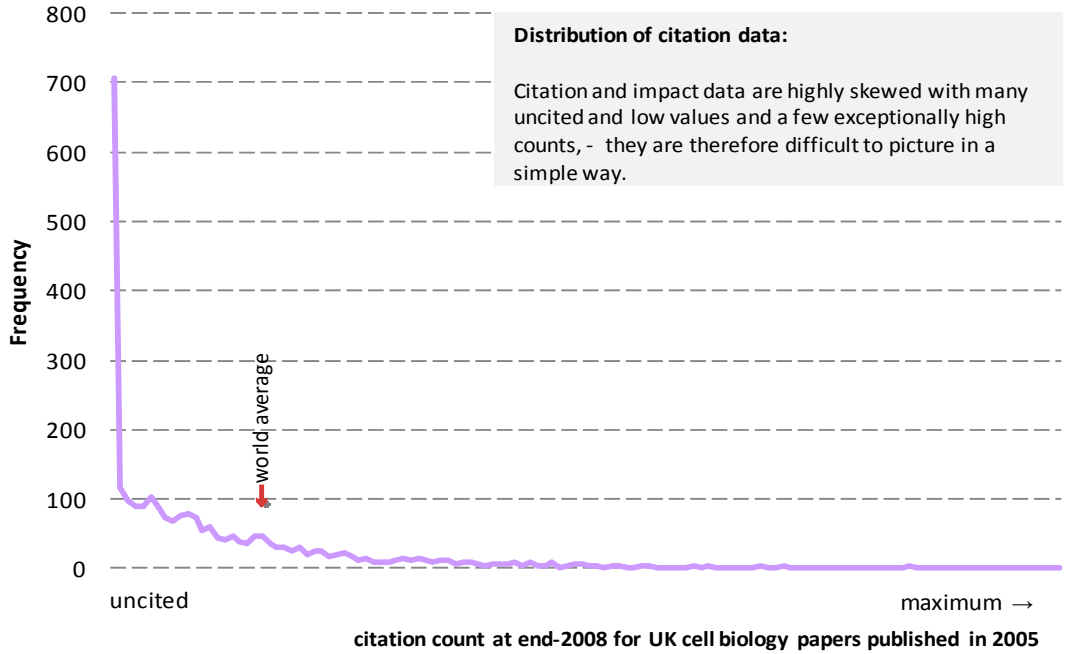
Unlike other databases, the Thomson Reuters Web of Science and underlying databases are selective, that is, the journals abstracted are selected using rigorous editorial and quality criteria. The authoritative, multidisciplinary content covers over 10,000 of the highest impact journals worldwide, including 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



citation count at end-2008 for UK cell biology papers published in 2005

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 above 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.

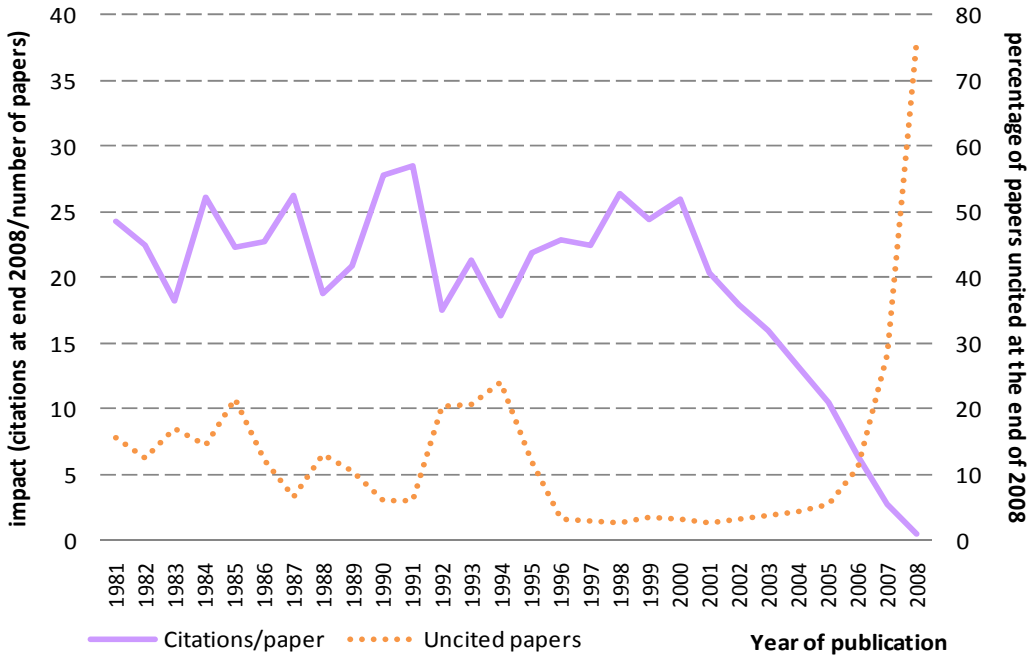
Citation analysis cannot make value judgments about why an article is uncited nor about why it is highly cited. The analysis can only report the citation impact that the paper has achieved. We normally assume, based on many other studies linking bibliometric and peer judgments, that high citation counts 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 below 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 on the next page 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 ten most frequently research fields in the HFSP-supported dataset of the 251 Web of Science journal categories to which publications are assigned by Thomson Reuters are listed below. The categories, ranked by most-used to less-frequently used, were:

- Biochemistry & Molecular Biology
- Neurosciences
- Cell Biology
- Developmental Biology
- Genetics & Heredity
- Immunology
- Biophysics
- Plant Sciences
- Physiology
- Microbiology

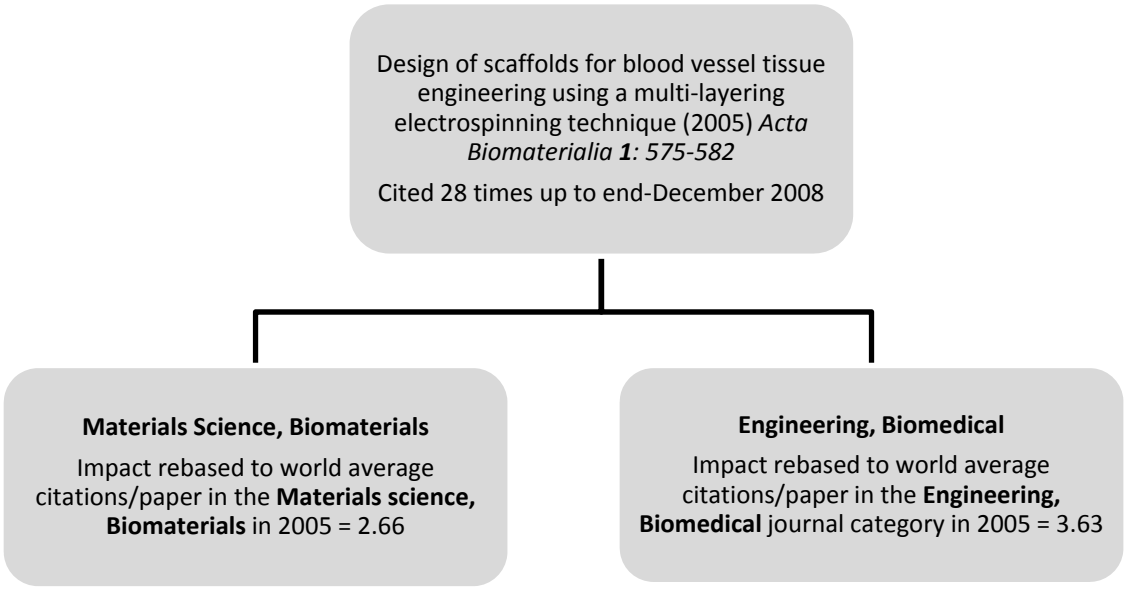
Thomson Reuters definitions for these most frequently used journal categories are given in Annex A.

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* 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 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 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 above, the average RBI of the *Acta Biomaterialia* paper can be given as 3.15.

Impact Profiles®

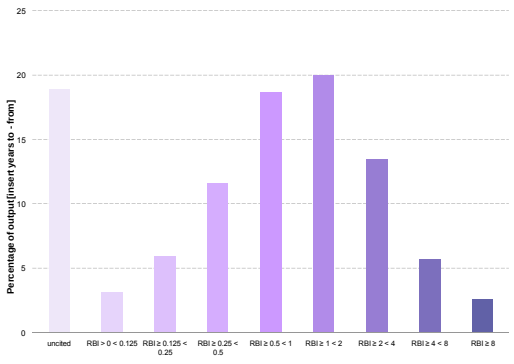
Evidence has developed a bibliometric methodology which shows the proportion of papers that are uncited and the proportion that lie in each of eight categories of relative citation rates, normalised (rebased) to world average. An Impact Profile® enables an examination and analysis of the strengths and 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.

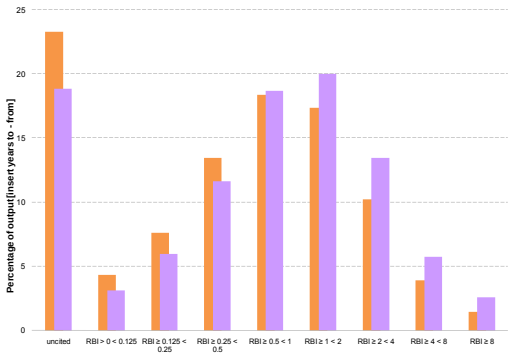
The proportion of uncited papers in a dataset can be compared to the benchmark for the UK, the USA or any other country. Overall, in a typical ten-year sample, around one-quarter of papers have not been cited within the 10-year period, the majority of these, of course, are those that are most recently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

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

A



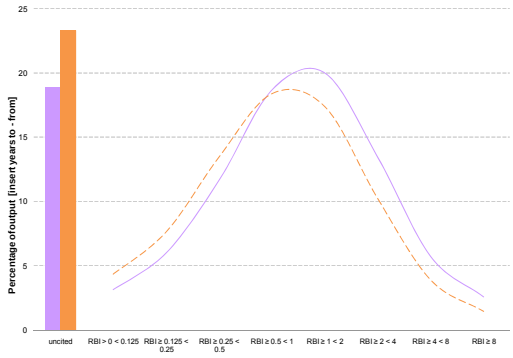
B



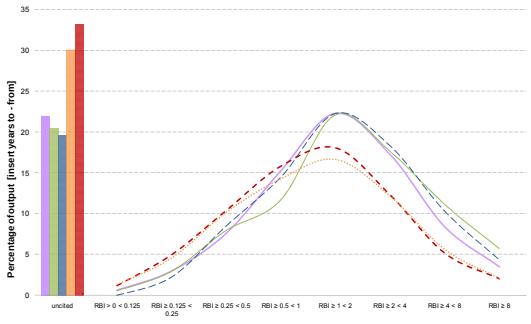
A: is used to represent the total output of an individual country, institution or researcher with no benchmark data, visually it highlights the numbers of uncited papers (weaknesses) and highly cited papers (strengths).

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.

2.4 Outline of analyses

The summary bibliometric analyses in this report cover publications acknowledging HFSP support for the most recent 10-year period from 1999 to 2008 inclusive.

The background data cover all sectors of the research community including higher education institutions (HEIs), hospitals and NHS trusts, companies, public sector research organisations and charities.

World average impact data have been sourced from the Thomson Reuters Web of Science baseline data for 2008.

Using Thomson Reuters' data, the following bibliometric information has been collated and assigned by *Evidence* to each publication record:

- Observed citation count for each paper, i.e. the number of times the paper has been cited from publication to end-2008.
- Expected citation count for each paper, i.e. the average number of citations to a paper in the same journal and published in that year.
- Annual citation counts, i.e. the number of times the paper has been cited in any year.
- The Category-specific rebased impact for each paper normalised, or rebased, to the world average for year and research field (may be more than one value for an individual paper)
- The Paper-level rebased impact – the average of the Category-specific rebased impact(s) for each individual paper.

2.5 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.

2.6 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 HFSP research publications database

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 supported by the HFSP over the period, but this part is usually recognised as describing the most direct contribution to the research base.

The publications for bibliometric analyses in this report were collated from the previously described EndNote databases (Section 2.2). In brief, with some details below – three databases and two EndNote 'groups' of data from a fourth database were imported into a single Access database for analysis:

RG_Collab_journals_final: 3,642 publications imported of which 767 explicitly acknowledge support from the HFSP, identified by CALL NUMBER field = HFSP.

LT_journals_final: 4,514 publications imported of which 1,960 explicitly acknowledge support from the HFSP, identified by CALL NUMBER field = HFSP.

CDA_journals_final: 505 publications imported of which 167 explicitly acknowledge support from the HFSP, identified by CALL NUMBER field = HFSP.

EndNote group **RGP Collab**: 1160 publications imported of which 318 explicitly acknowledge support from the HFSP, identified by CALL NUMBER field = HFSP.

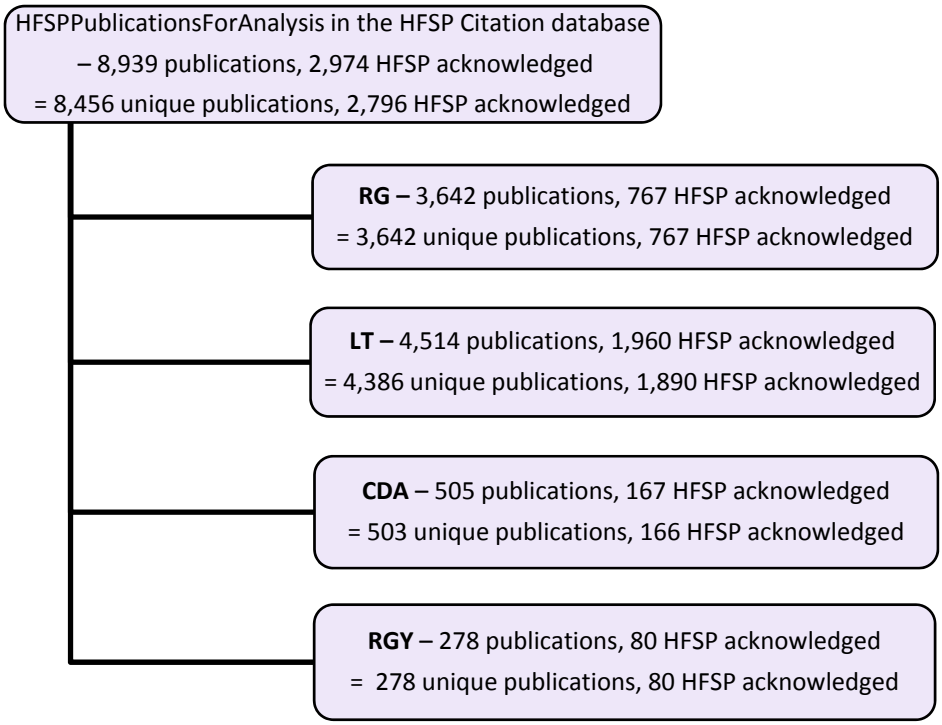
EndNote group **RGY Collab**: 278 publications imported of which 80 explicitly acknowledge support from the HFSP, identified by CALL NUMBER field = HFSP.

The Access table HFSPJournalsFinal contains these data. The two EndNote groups duplicate some of the data in **RG_Collab_journals_final** but with more current grant information [early years use RG in the ref numbers, later years either RGY (for young investigator grants) or RGP (for program grants)]. For analysis over the whole period, RG should be considered equivalent to RGP. An Access table HFSPPublicationsForAnalysis was generated from HFSPJournalsFinal to enable analysis using RG over the whole period. This table contains 8,939 publications.

The schematic diagram in Figure 3.1 outlines some characteristics of this dataset before bibliometric analyses were performed.

Thomson-designated data types for publications from the HFSP citation database (based on 2,796 unique publications with HFSP acknowledgement)	**Number of publications of this type**
Articles	2,562
Reviews	209
Meeting abstracts	9
Editorial material	8
Correction	4
Letters	4

Figure 3.1: Characteristics of the dataset for bibliometric evaluation of HFSP supported research publications



RG figures do not include RGY publications.

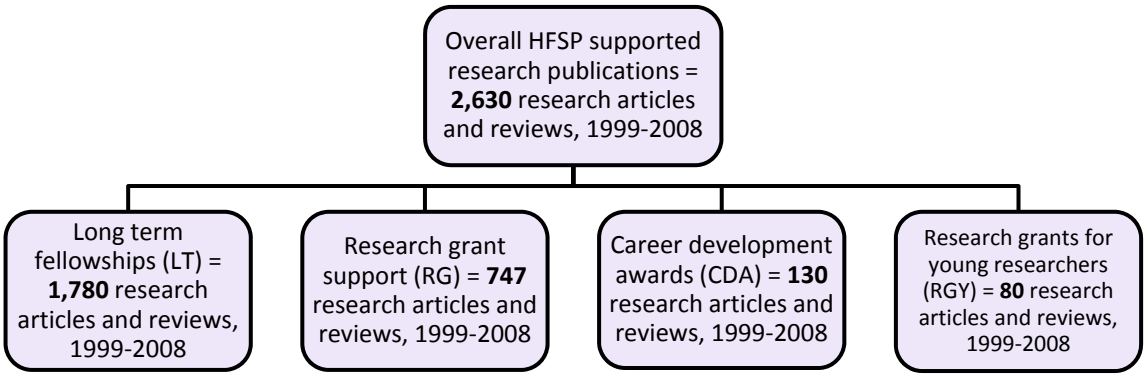
Table 3.1a HFSP supported research publications, all data types, by database year

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
LT	3	36	118	158	204	196	211	267	175	189	238	95
RG	9	25	45	60	87	89	108	104	89	71	80	
CDA						1	5	9	24	33	60	34
RGY		1	1		4	5	15	14	13	14	13	
Total	12	61	162	216	290	284	320	378	284	287	373	129

Table 3.1b Research articles and reviews within 10-year time period by HFSP support system

	LT	RG	CDA	RGY	Total 10 yrs
Total in database	1,890	767	166	80	
Articles	1,653	697	108	72	2,432
Reviews	127	50	22	7	198
Total – 'valid' data type and within 10-year time period	1,780	747	130	79	2,630

For bibliometric analyses, duplicated publications (and the citations to them) within an analysis dataset are counted once only. The total unique paper counts used in the bibliometric analyses are given below.



4 HFSP-supported research – summary bibliometric indicators

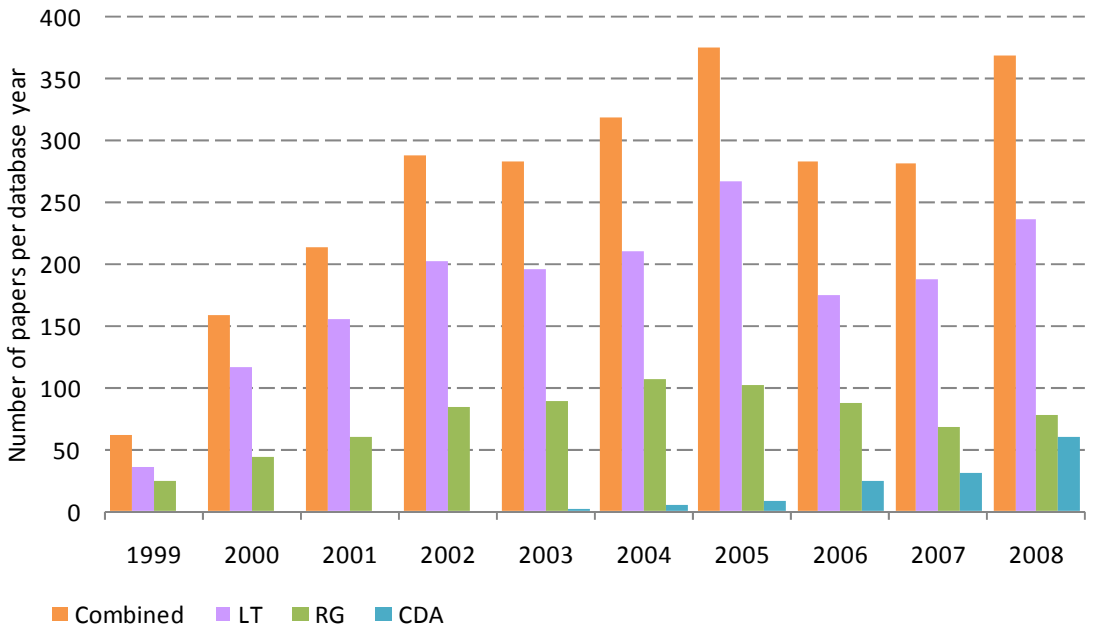
Section 4 describes the basic characteristics of the HFSP-supported publications dataset. In summary:

- The HFSP-supported dataset used for bibliometric analyses in this report comprises a total of 2,630 journal articles and reviews from over 380 journals for the 10-year period between 1999 and 2008 (Section 4.1).
- Most frequently used journals by HFSP-supported researchers are those which are leading journals in their respective research fields (Section 4.2).
- Nearly one-third of HFSP-supported research between 1999 and 2008 was published in journals assigned to the Biochemistry & Molecular Biology research field (Section 4.3).
- Overall, HFSP-supported research was more likely to have been cited than comparable research published in Biochemistry & Molecular Biology journals worldwide (Section 4.4).
- HFSP-supported research publications have an extremely good impact compared to world baselines. In particular, the impact of publications in Plant Sciences, Genetics & Heredity and Immunology are outstanding (Section 4.5).

The indicators described in this section suggest that HFSP-supported publications are exceptionally well-cited in many research fields. The categorised impact profiles of these publications is analysed further in Section 5.

4.1 Annual publication output

The figure indicates the annual publication output from HFSP-supported research for the 10-year period between 1999 and 2008.



Data coverage in the first three years of the 10-year period is rather sparse, possibly reflecting the lag period between grant award and publication of resulting research.

4.2 Journal usage

The 20 journals used most frequently by HFSP-supported researchers are listed in Table 4.2a.

Publications in these more commonly-used journals total over 1,450 papers, more than 55% of the total output. All these journals are ranked in the 'top' 25% (by Journal Impact Factor) in their specific research fields. This indicates that the journals used extensively by HFSP-supported researchers contain papers that are well-regarded amongst their peers.

Table 4.2a: Journals in which HFSP-supported researchers have published most frequently

Journal Title (ranked by total papers in database)	**Number of papers**	**Impact Factor 2008**
Proceedings of the National Academy of Sciences USA	201	9.380
Journal of Biological Chemistry	151	5.520
Cell	108	31.235
Journal of Neuroscience	96	7.452
EMBO Journal	94	8.295
Neuron	89	14.170
Science	81	28.103
Nature	80	31.434
Current Biology	71	10.777
Development	58	6.812
Genes and Development	58	13.623
Molecular Cell	57	12.903
Journal of Neurophysiology	52	3.648
Journal of Molecular Biology	50	4.416
Developmental Cell	43	12.882
Journal of Cell Biology	42	9.120
Molecular and Cellular Biology	37	5.942
Journal of Immunology	33	6.000
PLoS Biology	31	12.683
Nature Neuroscience	28	14.164

The 2008 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, an 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 20 journals with the highest journal impact factors used more than once by HFSP-supported researchers. The list is dominated by 'review' journals as articles in such journals tend to be cited more frequently than articles in research journals.

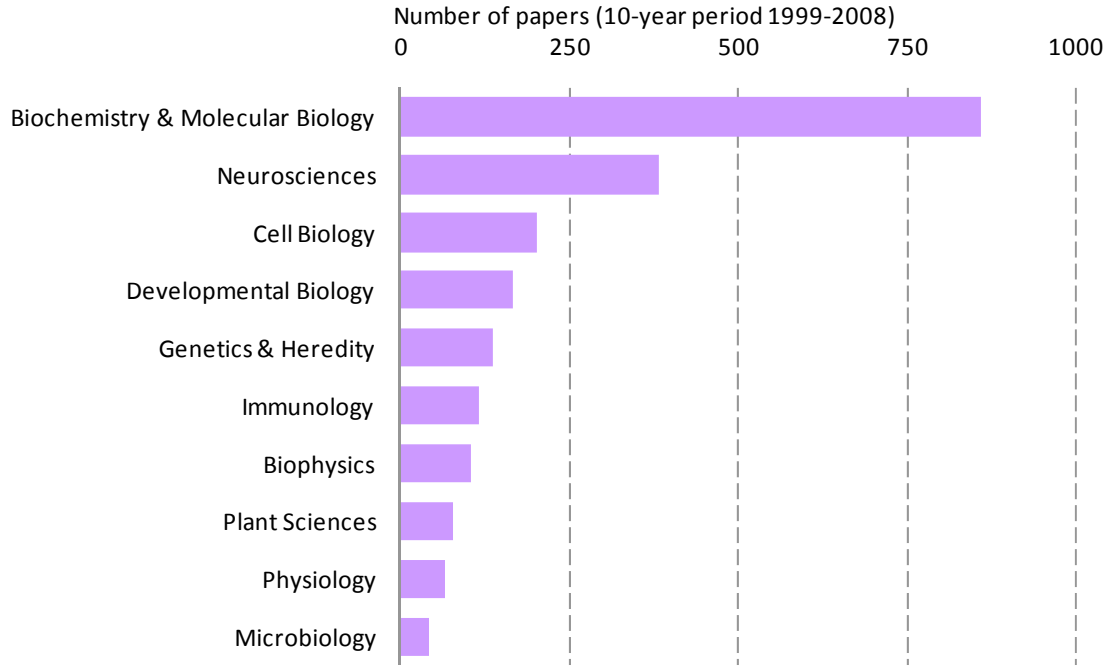
Table 4.2b: High Impact Factor journals used by HFSP-supported researchers

Journal Title (ranked by journal impact factor for 2008)	**Impact Factor 2008**	**Number of papers**
Nature Reviews Molecular Cell Biology	35.423	4
Physiological Reviews	35.000	1
Nature	31.434	80
Cell	31.253	108
Nature Reviews Cancer	30.762	2
Nature Genetics	30.259	19
Annual Review of Biochemistry	30.016	4
Science	28.103	81
Nature Medicine	27.553	3
Annual Review of Neuroscience	26.405	3
Nature Reviews Neuroscience	25.940	6
Nature Immunology	25.113	11
Cancer Cell	24.962	3
Nature Reviews Genetics	24.185	3
Nature Biotechnology	22.297	3
Immunity	20.579	26
Nature Cell Biology	17.774	26
Microbiology and Molecular Biology Reviews	16.950	2
Cell Stem Cell	16.826	4
Journal of Clinical Investigation	16.559	11
Cell Metabolism	16.107	3
Journal of Experimental Medicine	15.219	24
Nature Chemical Biology	14.612	3
Neuron	14.170	89

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 0). Research published by HFSP-supported researchers has been assigned to 137 of the 251 Web of Science journal categories or research fields used by Thomson Reuters.

Nearly one-third of HFSP-supported research between 1999 and 2008 was published in journals assigned to the Biochemistry & Molecular Biology research field. Two-thirds of HFSP-supported research in this time period is covered by the top five most frequently used research fields including Neurosciences, Cell Biology, Developmental Biology and Genetics & Heredity.

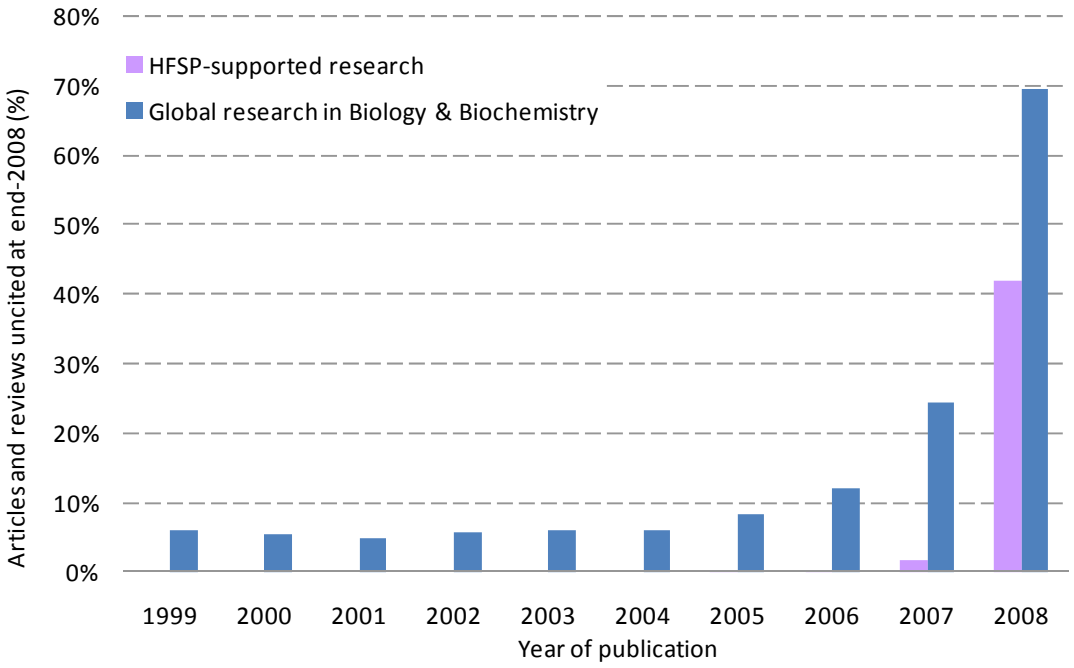


The chart above is based on one, unique journal category for each paper – this category is selected as the one in which the paper performs best, that is, has the highest impact within the category.

4.4 Percentage of publications which are uncited

Rates of citation accumulation are field-dependent. For the global research base as a whole, 10 years after publication there is a general plateau in citation profiles beyond which few additional citations would be expected. In biomedical sciences the plateau may be reached rapidly, often within two-three years after publication. A significant proportion of publications are never cited, however, in any 10-year sample, the majority of uncited articles would be less than three years old.

The following chart indicates the relative citation time-trend for HFSP-supported research and global research publications in Biology & Biochemistry. To get a valid comparison, noting the differences in citation behaviour between fields, this analysis has been restricted to articles published in the Thomson Reuters Essential Science Indicators macro field of Biology & Biochemistry.



Overall, HFSP-supported research is much more likely to have been cited than comparable papers published worldwide in this broad subject area. In addition, HFSP-supported research is cited more quickly than the global Biology & Biochemistry research and this gain is sustained even in older papers.

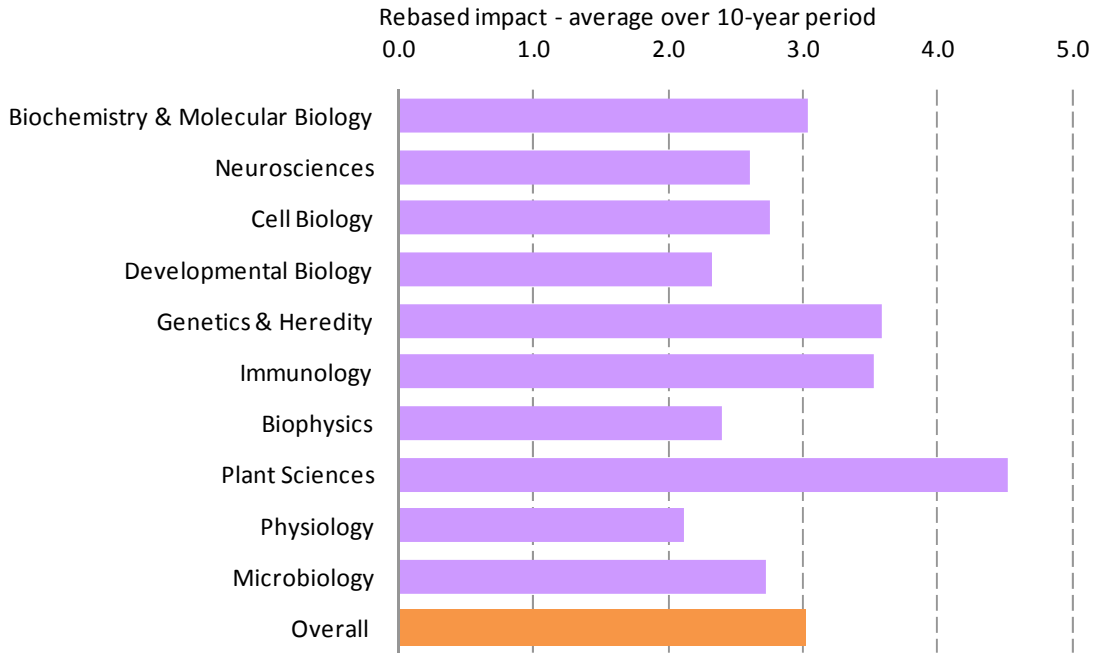
4.5 Impact of HFSP-supported research

The relative impact of HFSP-supported research publications is exceptional.

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 world average. This then allows appropriate comparison between years and between fields.

For example, the rebased impact (RBI) for HFSP-supported research in Biochemistry & Molecular Biology is the raw impact (citations/paper) of HFSP publications in Biochemistry & Molecular Biology for any specific year of publication divided by the raw impact (citations/paper) for all world publications in Biochemistry & Molecular Biology and in the same year of publication.

The overall rebased impact (RBI) of all of the HFSP-supported research publications is 3.03 (where world average is 1.0). For comparison, the global average rebased impact relative to world baseline for the 10-year period 1999-2008 in Biology & Biochemistry (Thomson Reuters Essential Science Indicators macro research field) was 1.38.

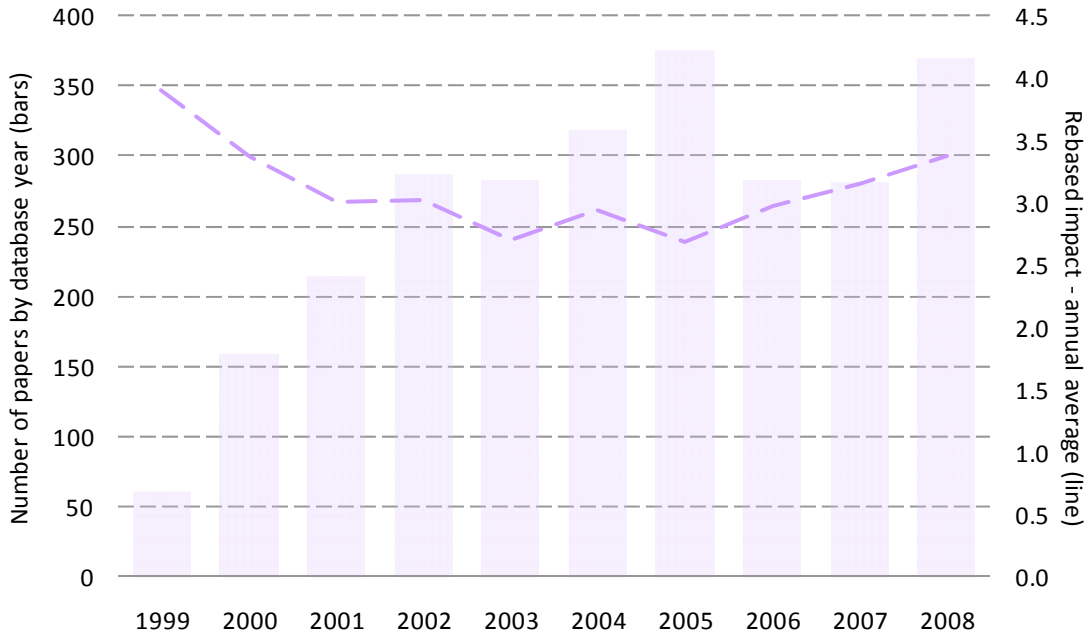


The chart shows that the average rebased impact of HFSP-supported research is well above world average in all fields.

The impact of the 81 articles published in Plant Sciences journals is four and a half times the world average.

4.6 Has the impact of HFSP-supported research changed during the 10-year period from 1999 to 2008?

Overall, the impact of HFSP-supported research in all research fields has been exceptionally good throughout the period covered by the data and the current indications are that impact has improved from a relatively low period in the early 2000s, though it must be stressed that even during this period rebased impact was more than 2.5 times the world average.



The database year is comparable to the publication year and refers to the date at which an article was abstracted by Thomson Reuters; this is the year against which it is appropriate to rebase the citations.

5 Impact Profiles® of HFSP-supported publications

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 for a typical, recent 10-year period. 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. However, it should not be assumed that this is the case for all such papers.

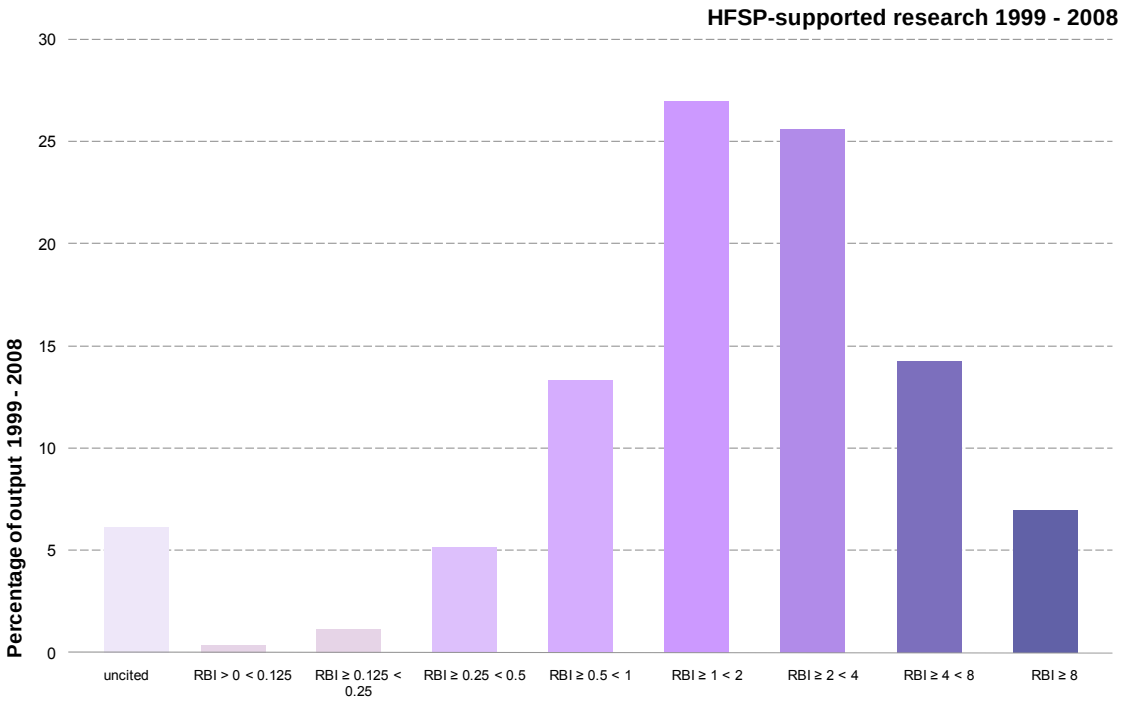
There is variation in non-citation between countries and between fields. For example, relatively more engineering papers tend to remain uncited than papers in other sciences, indicative of a disciplinary factor. There is also an obvious increase in the likelihood of citation over time but most papers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

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 eight 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.

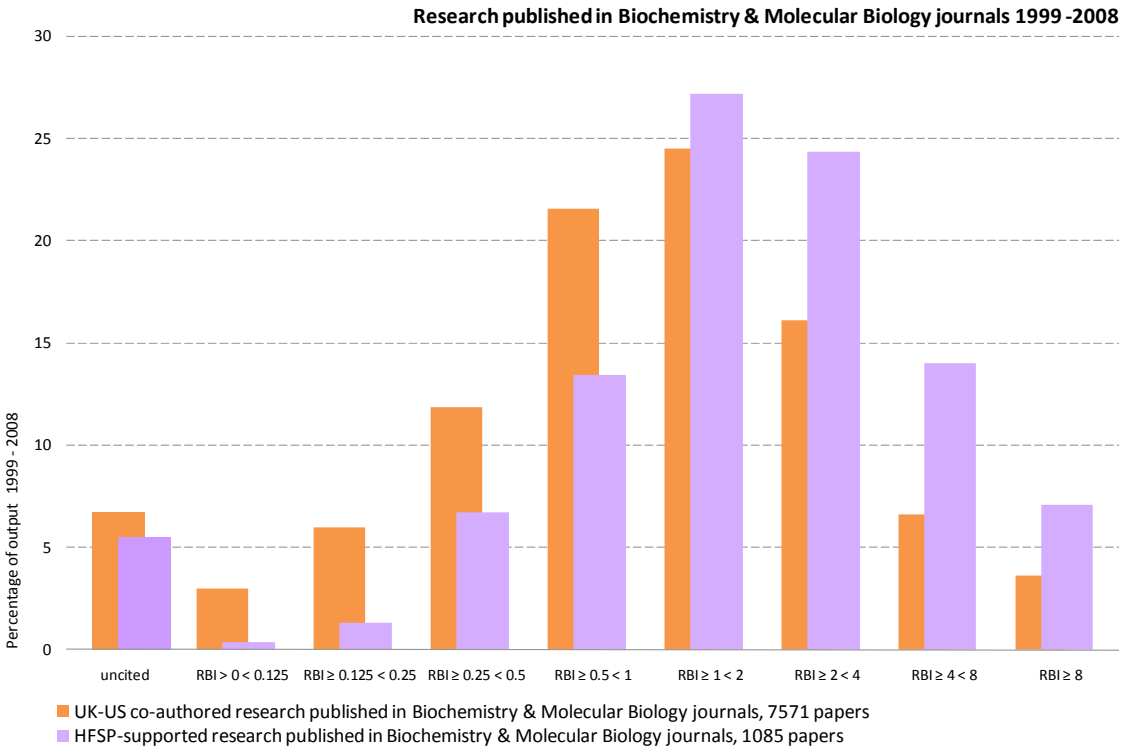
After reviewing the outcomes of a number of analyses, we have chosen a more relaxed definition for our descriptive and analytical work. We deem papers that are cited more often than four times the relevant world average to be relatively highly-cited for national comparisons. This covers the two most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

5.1 Impact Profile® for all HFSP-supported publications



- The Impact Profile® for the HFSP-supported research is clearly shifted to the right, higher impact categories over this period.
- HFSP-supported research is internationally important – more than one-fifth (21.3%) of papers are highly-cited (cited more than four times the relevant world average).
- Almost three quarters (73.9%) of HFSP-supported research has an impact more than the world average (RBI ≥ 1).
- Relatively few of the HFSP-supported papers are uncited.

5.2 Impact Profiles® for research published in Biochemistry and Molecular Biology journals, HFSP-supported compared to UK-US co-authored research



The benchmark dataset of UK-US co-authored papers has been chosen to represent research that is known to be very well regarded by researchers worldwide (**Patterns of international collaboration for the UK and leading partners** (2007): Adams, et al.). It has been restricted to research published in Biochemistry and Molecular Biology journals to enable a robust comparison – the total UK-US co-authored dataset would otherwise include research areas not important to the HFSP mission such as high energy physics.

- The Impact Profile® for the HFSP-supported research in Biochemistry and Molecular Biology is similarly shifted to the right, higher impact categories, as for the overall Impact Profile in Section 5.1.
- HFSP-supported research in Biochemistry and Molecular Biology is internationally well cited – more than one-fifth (21.1%) of papers are highly-cited (cited more than four times the relevant world average), proportionally twice as many as for UK-US co-authored research (10.2%).
- Almost three quarters (72.6%) of HFSP-supported research in Biochemistry and Molecular Biology has an impact more than the world average (RBI ≥ 1) compared to just over half of UK-US co-authored research (50.9%).

Annex A: Thomson Reuters definitions for the journal categories included in this report

A.1 Biochemistry & Molecular Biology

The Biochemistry & Molecular Biology category includes journals on general biochemistry and molecular biology topics such as carbohydrates, lipids, proteins, nucleic acids, genes, drugs, toxic substances, and other chemical or molecular constituents of cells, microbes, and higher plants and animals, including humans. Journals that focus on biochemistry in cells, tissues or organs and those whose primary focus is the organism of study, such as plants, microbes, and so forth are excluded. Also excluded are journals that focus on methods in biochemistry or molecular biology.

A.2 Neurosciences

Neurosciences covers journals on all areas of basic research on the brain, neural physiology, and function in health and disease. The areas of focus include neurotransmitters, neuropeptides, neurochemistry, neural development, and neural behaviour. Coverage also includes journals in neuro-endocrine and neuro-immune systems, somatosensory system, motor system and sensory motor integration, autonomic system as well as diseases of the nervous system.

A.3 Cell Biology

Cell Biology includes journals on all aspects of the structure and function of eukaryotic cells. The principle characteristic of journals in this category is an emphasis on the integration at the cellular level of biochemical, molecular, genetic, physiological, and pathological information. This category considers material on specific tissues, differentiated as well as embryonic.

A.4 Developmental Biology

Developmental Biology includes journals focussed on the specific mechanisms of cell, tissue, and organism development, as well as gametogenesis, fertilization, biochemistry and molecular genetic control of development, cell biology of gametes and zygotes, and embryology.

A.5 Genetics & Heredity

The Genetics & Heredity category includes journals that deal with the structure, functions, and properties of genes, and the characteristics of inheritance. This category also considers heritable traits, population genetics, frequency and distribution of polymorphism, as well as inherited diseases and disorders of the replicative process. The category is distinguishable from Biochemistry & Molecular Biology by its specific emphasis on the gene as a single functional unit, and on the gene's effect on the organism as a whole.

A.6 Immunology

The Immunology category includes journals that are dedicated to all aspects of immune response and regulation, at the cellular-molecular level as well as the clinical level. Other topics include studies of the interaction between pathogens and host immunity, as well as clinical immunology, emerging immunotherapies, and the immunologic contribution to disease course.

A.7 Biophysics

Biophysics covers journals that focus on the transfer and effects of physical forces and energy-light, sound, electricity, magnetism, heat, cold, pressure, mechanical forces, and radiation-within and on cells, tissues, and whole organisms.

A.8 Plant Sciences

Plant Sciences covers journals concerning many aspects of the study of plants including systematic, biochemical, agricultural, and pharmaceutical topics. This category includes materials on higher and lower plants, terrestrial and aquatic plants, plant cells, entire plants, and plant assemblages.

A.9 Physiology

Physiology covers journals concerned with the normal and pathologic functioning of living cells, tissues, and organisms. Topics include comparative physiology, molecular biochemistry of cell function, applied physiology, and pharmacological intervention in pathophysiological processes.

A.10 Microbiology

Journals in the Microbiology category focus on the interpretative application of microscope magnification to the study of materials that cannot be seen properly by the unaided eye. The instruments used in microscopy may be either optical in nature, or use radiation other than light for making enlarged images of minute objects (such as an electron microscope).