



---

evidence

---

## The Sainsbury Laboratory Bibliometric baselines

January 2008

### Contact details

*Evidence* Ltd 103 Clarendon Road, Leeds LS2 9DF  
t/ 0113 384 5680  
f/ 0113 384 5874  
e/ [enquiries@evidence.co.uk](mailto:enquiries@evidence.co.uk)

*Evidence* Ltd is registered in England, Company no 4036650, VAT registration 758 4671 85  
<http://www.evidence.co.uk>

Confidential: This document and attachments have been prepared solely for the customer indicated. They may not be disclosed to any third party without permission.



## 1 Executive Summary

- .e This is a report of bibliometric analyses of publications produced by The Sainsbury Laboratory (TSL) over the 10-year period between 1997 and 2006.
- .e For TSL, a higher proportion of publications are substantive journal articles than is the case for the UK as a whole. The most frequently used journals by TSL are leading journals in their respective research fields.
- .e The publication output from TSL has remained steady at around 40 papers and 35 substantive journal articles per year between 1997 and 2006.
- .e TSL publications are more likely to be cited than comparable papers in the UK. There is an exceptional proportion of output in the most-cited categories, well above UK and world benchmarks.
- .e TSL performs exceptionally well in the research field of Plant Sciences and also in fields associated with basic biochemistry. The publication profile in these fields suggests the use of well-regarded journals and the publications have high impact, both in the specific journals and in the wider research field.



## 2 Introduction

### 2.1 The Sainsbury Laboratory

The Sainsbury Laboratory (TSL) was set up in 1987 as a joint venture between the Gatsby Charitable Foundation (Gatsby), the John Innes Foundation (JIF), the University of East Anglia (UEA) and the Agricultural and Food Research Council (now part of BBSRC). Research at TSL is focussed on

plant interactions with microbes and viruses, using state-of-the-art genetics and molecular biology techniques.

This report provides background bibliometric analyses and UK benchmarks to set the publication performance of The Sainsbury Laboratory into context.

### 2.2 Outline of analyses and report

This report compares recent publication data (10-year period 1997 to 2006) for TSL with similar UK research publications. 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.

For the analyses, bibliometric data was sourced from Thomson Scientific, with which *Evidence* has a strategic alliance. Thomson's Philadelphia-based subsidiary, Thomson ISI®, maintains the most complete international data on research journal publications and their citations. Citations are subsequent references made to an article by later publications. Highly-cited work is recognised as having a greater impact and *Evidence* has shown that high citation rates are correlated with other measures of research excellence. World average impact data are sourced from the Thomson Scientific National Science Indicators 2006.

*Evidence* has performed a complete address reconciliation for the Thomson Scientific 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. The research publications for the analyses reported here were identified by *Evidence* as those with The Sainsbury Laboratory (and address variants) as at least one of the author addresses on publications from the most recent 10-year period (January 1997 to December 2006 inclusive). *Evidence* has crosschecked the data and modified

searches to ensure that the correct data have been extracted. Such searches are, inevitably, subject to some error because of mistakes made by authors or by publishers during the publication process or by Thomson Scientific during data inputting.

A time window of five to ten years is normally considered most appropriate for initial bibliometric analyses. Research by *Evidence* has shown that citation counts in the first two years after publication can be a general guide to long-term impact for collections of papers (e.g. from research programmes) but this cannot be applied with absolute confidence to single papers.<sup>1</sup>

From the publication data, bibliometric analyses have been carried out for all the publications produced by TSL and for publications 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 Scientific journal categories because these are well-established and are informed by extensive Thomson Scientific work with the research community over the last twenty years. Papers may be allocated to one or more categories according to which journal the paper is published in. For example, papers published in the journal *Cancer* are assigned both to Oncology and to Oncogenesis & Cancer categories. The Multidisciplinary journal category may include articles from prestigious journals such as *Nature* and *Science* although,

recently, most papers from these 'multidisciplinary' journals are assigned to specific categories based on the journal categories of the references cited by the article.

The journal categories fields used in this report were selected from the TSL dataset as the top 10 most-frequently used of the 106 'Current Contents' journal categories to which publications are assigned by Thomson Scientific. This selection coincided with fields with an output of more than 5 papers over the 10-year period. The categories, ranked by most-used to less-frequently used, were:

- Plant Sciences
- Biochemistry & Biophysics
- Molecular Biology & Genetics
- Microbiology
- Cell & Developmental Biology
- Biology
- Multidisciplinary
- Animal Sciences
- Biotechnology & Applied Microbiology
- Experimental Biology

---

### 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. This could be seen in some of the data analysed for this report. For this reason, the report indicates where analyses have used not a single-year average but a 5-year moving average. 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 interquartile mean, calculated from a usefully-truncated range

of values lying between the 25th centile and the 75th centile.

Citation data provided by Thomson Scientific are assigned on an annual census date referred to as the Article Time Period. For the majority of publications the Article Time Period is the same as the year of publication but for a few publications (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

**Papers:** Thomson Scientific abstracted publications include editorials, meeting abstracts, book reviews as well as full journal articles.

**Citations:** the citation count is the number of times that a given publication has been cited since it was published.

**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 publications in that field and year of publication divided by the impact for all world publications across that research field in the same year of publication

### 3 TSL publications – bibliometric indicators

This section describes the basic characteristics of the TSL publication dataset and compares it to those of the UK as a whole.

- The TSL dataset contains a total of 406 publications from 96 journals for the 10-year period between 1997 and 2006.
- A much higher proportion of TSL publications are substantive journal articles than for the UK as a whole (Section 3.1, 3.2).
- TSL publications are more likely to be cited than comparable papers in the UK (Section 3.3).
- Most of TSL research is published in journals associated with plant sciences (Section 3.4).
- The most frequently used journals by TSL are those which are leading journals in their respective research fields. In addition, TSL also publishes an outstanding proportion (more than 12.5%) of its papers in elite, multidisciplinary journals such as Science and Nature (Section 3.5).
- TSL research publications have a high impact compared to UK and world baselines. The average impact has been consistent across the period (Section 3.6, 3.7).

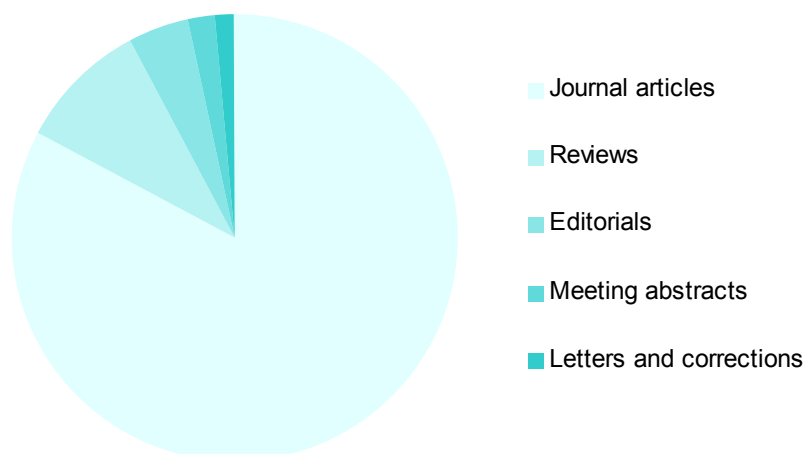
The indicators described in this section suggest that TSL publications are of higher quality than that of the general UK research base. The categorised impact profile of these publications is analysed further in the next section.

---

#### 3.1 Proportion of journal articles to other types of publication

The TSL dataset for 1997-2006 analysed for this report contained 406 journal articles and other published items collated from 96 journal titles.

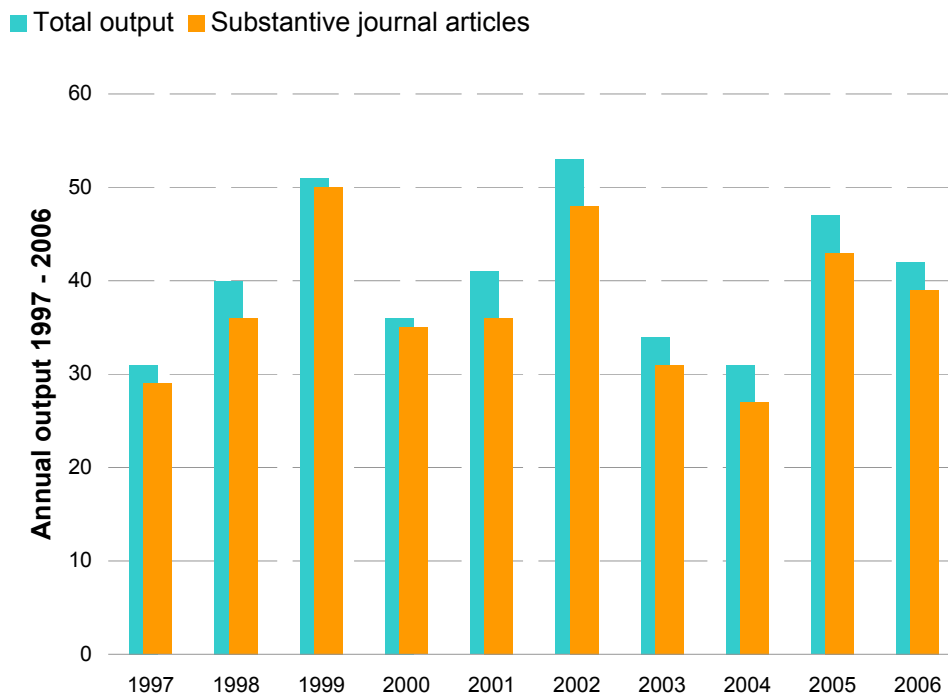
Substantive journal articles account for 83% of the TSL dataset, which is a greater proportion of outputs than is typical for the UK as a whole over the same 10-year time period (UK research base about 65%).



This analysis does not cover conference proceedings, books, chapters in books or grey literature such as reports. It therefore captures only a specific part of the total output generated by TSL over the period, but this part is usually recognised as describing the most direct contributions to the research base.

### 3.2 Trends in publication output between 1997 and 2006

The following figure indicates the annual total output and number of substantive journal articles published by TSL over the 10-year period between 1997 and 2006.



On average, TSL has published 30–40 papers per year in journals on Thomson Scientific databases.

As noted earlier, most TSL journal outputs are substantive articles.

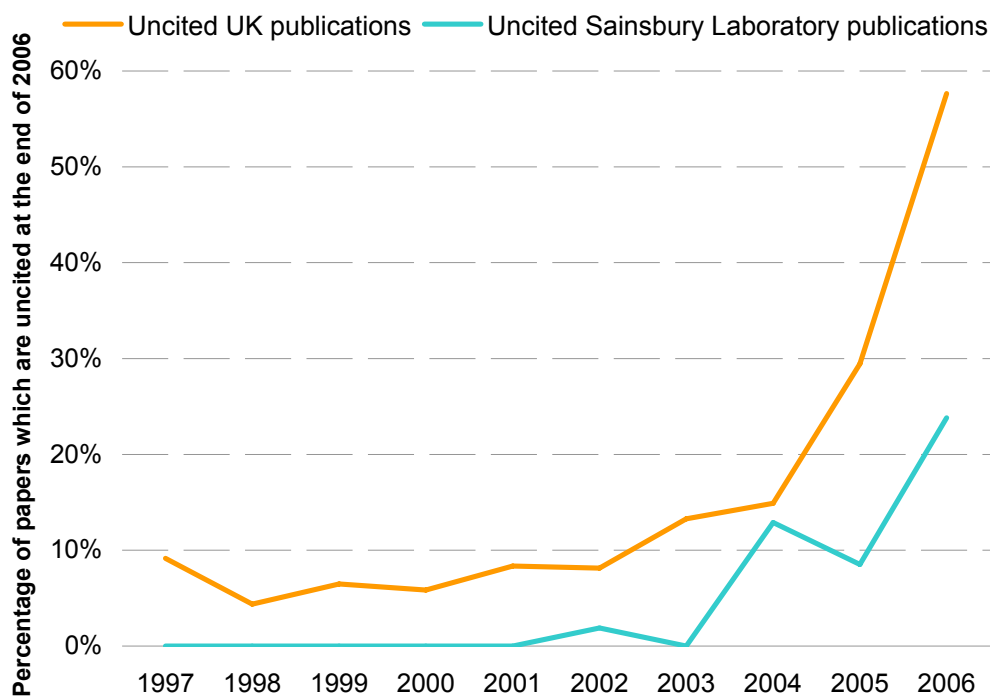
There was a marked drop in number of papers published in 2003 and 2004, which followed an exceptional peak in 2002. The last two years' data are more typical of the broad pattern for the period.

### 3.3 Percentage of publications which are uncited

The percentage of papers which remain uncited at the end of 2006 was lower for TSL publications than for the rest of the UK in Plant Sciences for any year of publication.

Rates of citation accumulation are field-dependent. For the UK research base as a whole, ten 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 2-3 years after publication. A significant proportion of publications are never cited, however, and typically around one-third of UK output is uncited in any ten-year sample (the majority of uncited articles would be less than three years old).

The following figure indicates the relative citation time-trend for TSL and UK publications. To get a valid comparison, noting the differences in citation behaviour between fields, this analysis has been restricted to articles published in Plant Sciences. A comparison between all TSL articles and all UK articles would include a large medical and physical science background of little relevance to TSL.



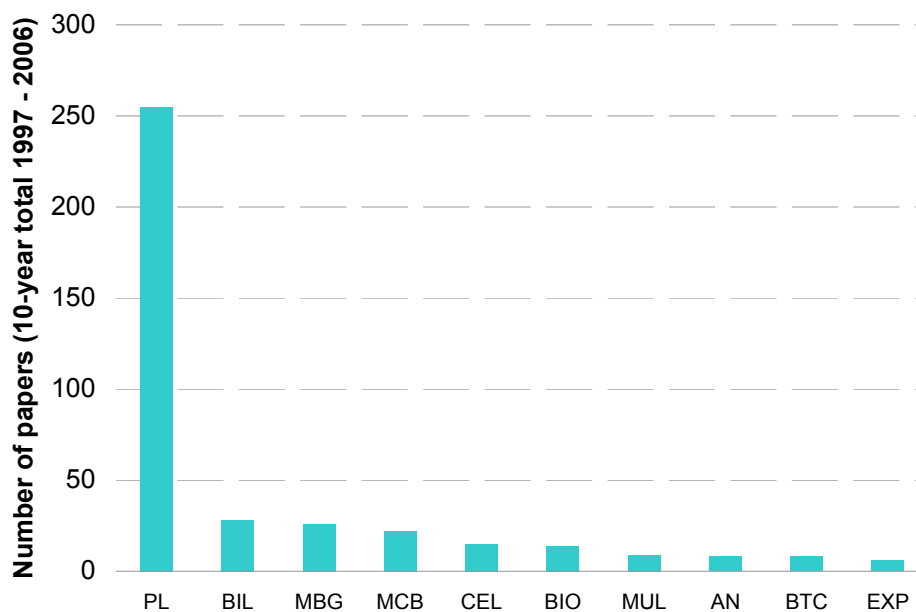
Overall, TSL publications are much more likely to have been cited than comparable papers in the UK. Most TSL publications are cited within three years whereas up to 10% of comparable publications elsewhere in the UK research base are never cited.

### 3.4 Most frequently used research fields

Most research publications from TSL appear in journals assigned to the Plant Sciences category.

About 63% of TSL journal articles are in Plant Science journals. This is a strongly focussed portfolio compared to many other research laboratories.

Other research fields frequently used by TSL in its publication output include basic biochemistry, molecular biology & genetics and microbiology (around 19% in total).



#### Key to Research Field codes

PL	Plant Sciences	BIO	Biology
BIL	Biochemistry & Biophysics	MUL	Multidisciplinary science
MBG	Molecular Biology & Genetics	AN	Animal Sciences
MCB	Microbiology	BTC	Biotechnology & Applied Microbiology
CEL	Cell & Developmental Biology	EXP	Experimental Biology

### 3.5 Most frequently used journals

The twenty journals used most frequently by authors at TSL are listed in the Table below.

Publications in these more commonly-used journals total 300 papers or 73.9% of the total output. The majority of these journals are ranked in the 'top' 10% (by Journal Impact Factor) of journals in their specific research fields. This indicates that the journals used by TSL contain papers that are well-regarded amongst their peers.

In the Plant Sciences category, TSL has published extensively in the most of the top 10 journals, with the exception of two review journals: *Annual Reviews in Plant Biology* and *Annual Reviews in Phytopathology*. We noted earlier that TSL publishes relatively few review articles.

TSL has published more than 50 papers in three elite, multidisciplinary journals: *Nature*, *Science* and *Proceedings of the National Academy of Sciences USA*.

#### Journals in which TSL has published most frequently

(ranked by total papers 1997-2006)

Journal Title	Number of papers	Impact Factor 2006	Research Field and rank by impact factor	
Plant Journal	51	6.565		PL (6/147)
Plant Cell	46	9.868	PL (3/147)	BMB (17/262) CLB (18/156)
Molecular Plant-Microbe Interactions	31	3.936	PL (10/147)	BMB (70/262) BTC (21/140)
Proceedings of the National Academy of Sciences USA	22	9.643		MUL (3/50)
Nature	19	26.681		MUL (2/50)
Current Opinion in Plant Biology	17	10.182		PL (2/147)
Plant Physiology	17	6.125		PL (7/147)
EMBO Journal	13	10.086	BMB (16/252)	CLB (16/156)
Theoretical and Applied Genetics	13	2.715	PL (19/147) HER (61/131)	AGR (2/49) HOR (1/21)
Science	11	30.028		MUL (1/50)
Cell	8	29.194	BMB (2/262)	CLB (2/156)
Genetics	7	4.242		HER (32/131)
Journal of Biological Chemistry	7	5.808		BMB (39/262)
Plant Molecular Biology	7	3.577		PL (12/147)
Trends in Plant Science	7	8.000		PL (5/147)
Current Biology	6	10.988		BMB (14/262)
Plant and Cell Physiology	6	3.324		PL (15/147)
Development	4	7.764		DEV (5/34)
New Phytologist	4	4.245		PL (8/147)
Nucleic Acids Research	4	6.317		BMB (36/262)
Trends in Biochemical Sciences	4	13.863		BMB (7/262)

Note that the fields used in this part of the analysis differ from and are slightly finer granularity than those used in the previous section.

**Key to Research Field codes for journal impact data**

AGR	Agronomy	HOR	Horticulture
BTC	Biotechnology & Applied Microbiology	BMB	Biochemistry & Molecular Biology
CLB	Cell Biology	MUL	multidisciplinary journals
DEV	Developmental Biology	PL	Plant Sciences
HER	Genetics & Heredity		

### 3.6 Impact of TSL research

The relative impact of research publications from TSL is outstanding.

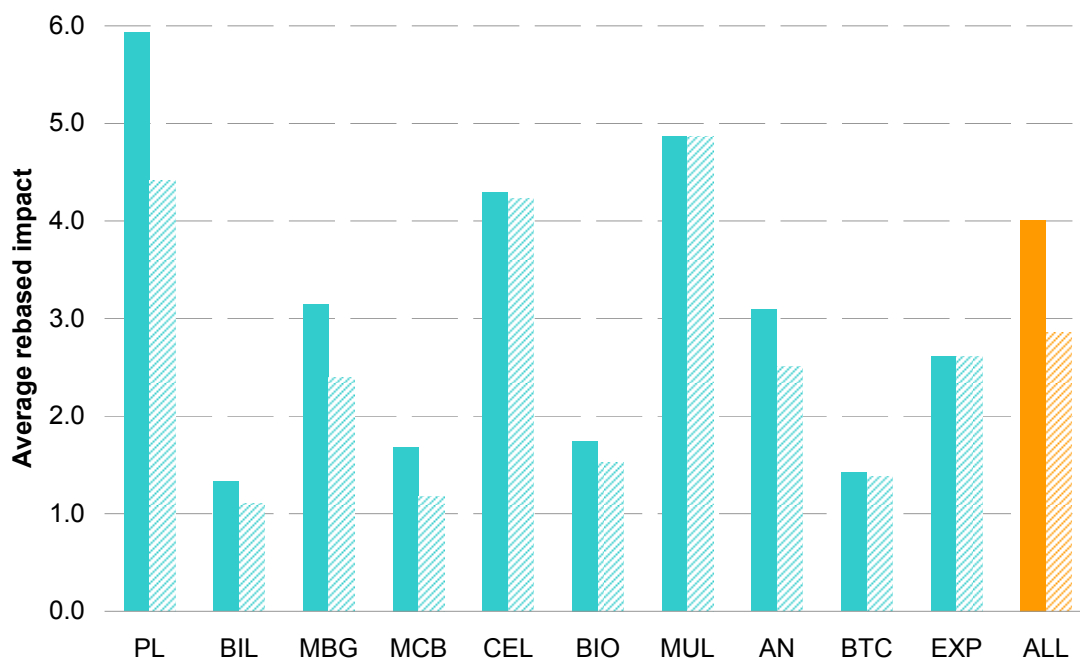
Impact of research, an index linked to citation accumulation, is field-dependent. All data presented in this report are therefore impact normalised, or rebased, to the world average. This then allows appropriate comparison between years and between fields.

For example, TSL’s rebased impact (RBI) for Plant Sciences is the raw impact (citations/paper) of TSL publications in Plant Sciences for any specific year of publication divided by the raw impact (citations/paper) for all world publications in Plant Sciences and in the same year of publication.

The overall rebased impact (RBI) of TSL research publications is 4.01 (where world average is 1.0). For comparison, the UK’s average gross impact relative to world baseline across all fields is about 1.3.

The interquartile mean for TSL’s RBI is 2.86. This indicates that although some of the high average could be due to the positive skew of a few exceptionally-cited publications, the average is supported by a substantial body of well-cited publications.

Rebased Impact 1997-2006; average (solid bars); interquartile mean (crosshatched bars)



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

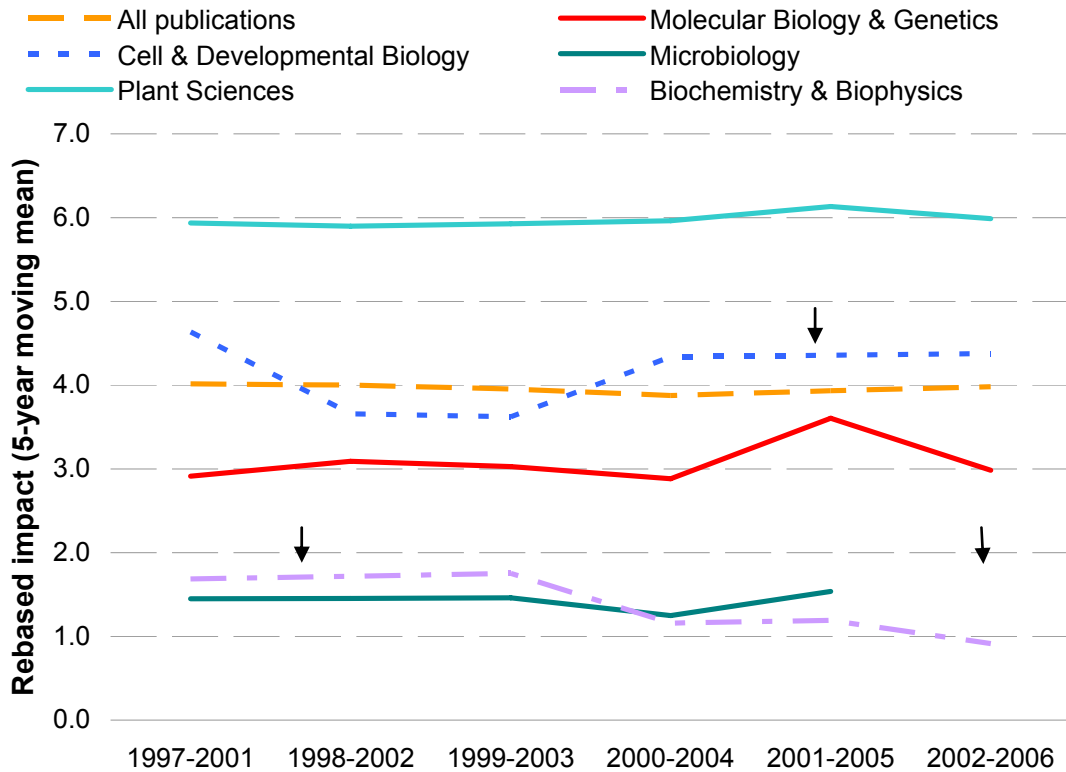
In some research fields it is exceptional, being on average more than four times world baseline in Plant Sciences, Multidisciplinary science, Cell & Developmental Biology. For comparison, less than 10% of the UK’s overall output is cited more often than four times world average.

Research in some other fields, although less exceptional, nonetheless maintains a good average, as it does in Biochemistry & Biophysics, Biotechnology & Applied Microbiology and Biology.

### 3.7 Trends in research impact between 1997 and 2006

The impact of TSL research in all research fields has been consistently high throughout the 10-year period covered by this report.

The figure below shows the 5-year moving-average RBI for all publications and for the top five research fields (data are too sparse for other research fields). The world average benchmark is 1.0.



Arrows mark points where the 5-year moving average has been estimated due to a missing annual data point.

It is evident that, across fields, the level of impact has been broadly maintained throughout the period. It remains consistently high in the fields in which most TSL research is published. It shows slightly more volatility in some of the less intensive fields.

## 4 Impact Profiles™ of TSL publications

Impact Profiles™ enable an examination and analysis of the balance of published outputs relative to world average and relative to a reference profile.<sup>2</sup> 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, ( $\geq 4$  x world,  $\geq 8$  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, engineering papers tend to remain uncited more often 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.

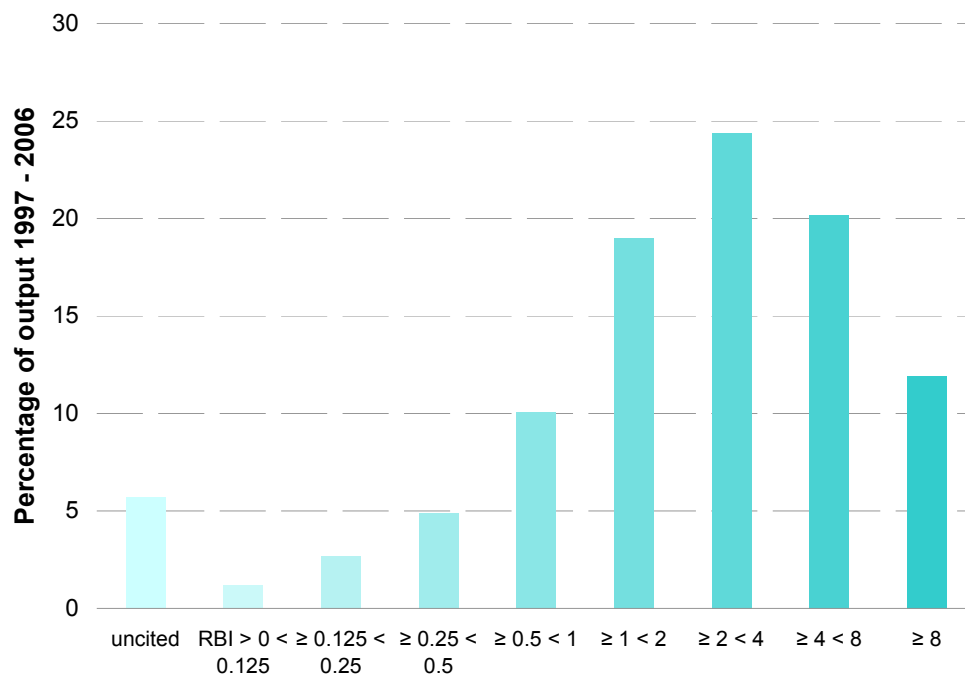
We work on the assumption that relative non-citation rates within a field are one of the indicators of the extent to which a body of work is regarded by others in the same field to be of greater or lesser significance to their subsequent work.

### What is the threshold for 'highly cited'?

Thomson Scientific 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.

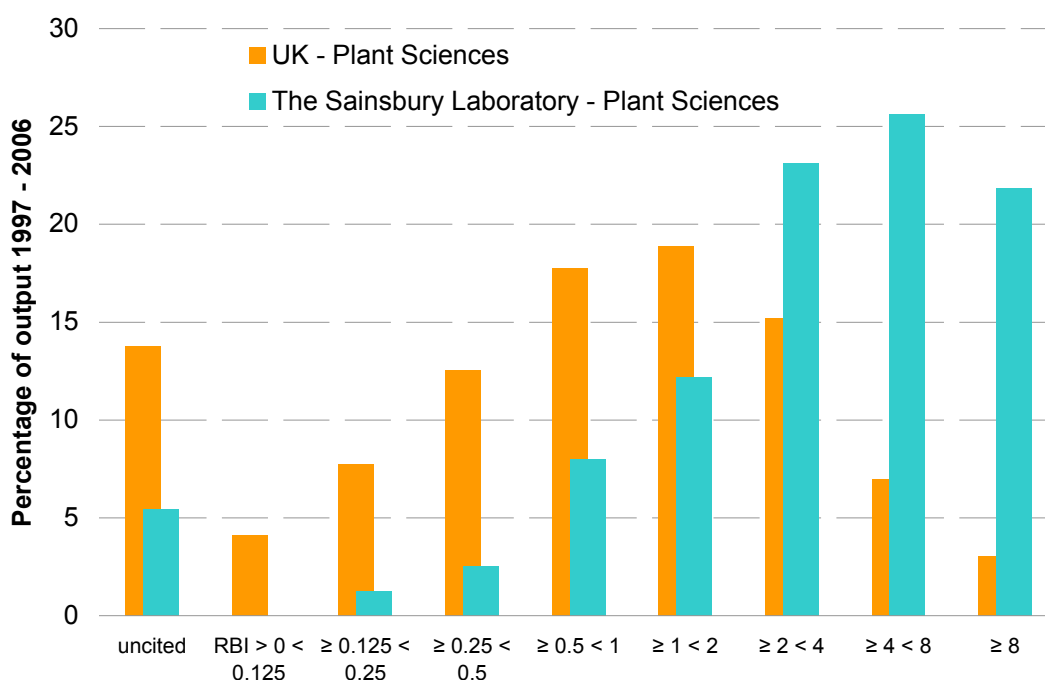
After reviewing the outcomes of a number of analyses, we have chosen a more relaxed definition for our descriptive and analytical work. We deem papers that are cited more often than 4 times the relevant world average to be relatively highly-cited for national comparisons. This covers the two most highly-cited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

## Impact Profile™ for all TSL publications



- Research published by TSL is outstanding - almost one-third (32.1%) of papers are highly-cited (cited more than 4 times the relevant world average).
- The modal group for TSL's cited papers is well above world average (1.0) at RBI 2–4 whereas it is below world average for the UK research base as a whole.
- TSL has relatively few uncited papers in this ten-year sample. The UK typically has some 25-30% of papers uncited in any similar sample.

## Impact Profiles™ for publications in Plant Sciences: TSL benchmarked against the UK



- The publication impact of TSL research in Plant Sciences is outstanding compared to the rest of the UK. Almost half (47.5%) of TSL papers are highly-cited (BBI > 4).
- TSL has an exceptional percentage of extremely highly-cited papers (RBI ≥ 8) where its advantage relative to the UK research base is even greater than in other categories above world average.
- The modal group for TSL Plant Science papers is well above world average (1.0) at RBI 4–8. The proportion of cited papers above the world average (1.0) is 82.8% (UK = 44.1%).
- TSL has many more papers cited above world average and many fewer papers cited below world average than the rest of the UK.
- In all impact categories above world average (to the right of the figure), publications from TSL perform better than the rest of the UK, indicating that Plant Sciences research from TSL is much more frequently cited than similar publications from the rest of the UK.
- In all impact categories below world average (to the left of the figure) TSL has a much lower percentage of papers than the rest of the UK. There is a much lower percentage of uncited TSL papers than for other UK research in this area (5.5% compared with 13.8%).

<sup>1</sup> Adams J. (2005) Early citation counts correlate with accumulated impact. *Scientometrics*, **65** (3), 567-581.

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