Bibliometric analysis of ongoing projects: Innovative Medicines Initiative Joint Undertaking (IMI)

IMI EXECUTIVE OFFICE

Second report: March 2013

Prepared by Thomson Reuters on behalf of IMI JU Executive Office under a public procurement procedure

1 EXECUTIVE SUMMARY

This report presents a bibliometric analysis of IMI research associated with funding Calls 1, 2 and 3, using citations as an index of research quality and co-authorship as an index of collaboration. The analyses use two sets of research publications, publications from IMI projects (IMI project research) and publications from IMI-supported researchers (IMI researchers).

The overall volume of IMI project research has increased rapidly since 2009. This is partly to be expected as the number of funded projects rises and those projects funded earliest in the program begin to publish. To date, IMI projects have produced 320 publications – over one-third of which have been published in the last six months – appearing in more than 150 journals including Nature, JAMA and, more recently, Science.

The volume of IMI research has also increased at the level of individual projects. EUROPAIN and NEWMEDS have been the most prolific projects funded in Call 1 while output from BTCure has increased rapidly in the last six months and is now in line with EUROPAIN and NEWMEDS. Of the 30 projects funded in Calls 1, 2 and 3, all but three (ABIRISK, EUPATI and PreDiCT-TB, which were Call 3 projects and were only initiated in 2012) have published once to date.

IMI project research is wide-ranging – the research portfolio from IMI projects covers a wide range of research fields and this diversity has increased in the last six months. Whilst IMI project research has again been published most frequently in Pharmacology & Pharmacy and Neurosciences journals, research in Mathematics & Computational Biology and Rheumatology journals has become more common.

IMI project research is well-regarded – the quality of IMI project research (as indexed by citation impact) has not only been maintained but has increased while output has grown. The average citation impact of IMI research is well above world and European averages and over twice the world average for specific research fields. Furthermore, over one-tenth of papers from IMI projects are 'highly-cited', that is, they belong to the world's top ten percent of papers in that journal category and year of publication, when ranked by number of citations received.

Researchers funded by IMI are also well-regarded by their peers. The total research published by IMI researchers funded in Calls 1, 2 and 3 (as opposed to that directly associated with IMI funding) is well cited with an average citation impact over twice world average and more than one-fifth of papers being 'highly-cited'. This indicates that IMI is funding researchers that perform at a high overall level.

These researchers are also highly collaborative. About two-fifths of all publications by IMI researchers were cross-sector, for example, between academic institutions and small medium enterprises (SMEs). About one-quarter of all publications were cross-project; this has increased slightly from the first analysis of Call 1 researchers.

A more detailed summary of key findings from this report cross-referenced to associated analyses is presented overleaf.

Summary of key findings – IMI project research

Currently there are 40 active IMI projects, of which 14 (almost half) were launched since 1 January 2011. It may take several years for a project to progress from inception to the point where it has generated sufficient data for a publication. It may take further years until it has produced its most valuable results. The IMI projects that are analysed here are young, and early bibliometric indicators may not fully reflect their eventual impact.

- A total of 320 publications from IMI-supported projects were identified for inclusion in this report on IMI research activity up to the end of January 2013. Over 95% of these documents were substantive articles and reviews (Section 4.1, Section 4.2).
- IMI project research continues to be published in highly-regarded journals (Science, Nature, JAMA and PNAS) and the core set of journals used by IMI projects reinforces the diversity of IMI-supported research with titles focused on bioinformatics, rheumatology, oncology and genetics (Section 4.4, Section 4.7).
- IMI project research is most frequently published in Pharmacology & Pharmacy and Neurosciences journals, however, IMI output in Mathematics & Computational Biology journals has expanded and this is now the third most frequent category (Section 4.5).
- The average citation impact for IMI project research is 1.34 for the 2-year period, 2010-2011, (where world average is 1.0) and more than one-tenth (11.6%) of papers are highly-cited. For comparison, the EU's average citation impact relative to world baseline for the same 2-year period in similar research fields was 1.14 (Section 4.6).
- Over one-third (35.0%) of the IMI project research portfolio was published in the last six months (Section 4.7).
- Output has increased for all three IMI funding Calls, though Call 1 continues to account for the highest share of IMI research. EUROPAIN and NEWMEDS have been the most prolific projects funded in Call 1 while BTCure has accelerated output in the last six months with output now in line with EUROPAIN and NEWMEDS (Section 5).

Summary of key findings – IMI researchers

The productivity, research performance and collaboration of researchers funded by IMI through Calls 1, 2 and 3 were assessed by analysing their total publication output (not limited to publications acknowledging IMI funding). Some 3 477 researchers were included in the analysis and 17 406 of their publications were identified for the period January 2007-January 2013 (Section 7.1).

- Publication output, as previously, is higher for IMI-supported researchers based in academic institutions and other research environments compared to industry and SMEs (Section 7.3).
- Since the initial report, researchers associated with patient organisations have published well-regarded research and now show the strongest research performance. Three (23.1%) have published at least one 'hot' paper, 4 (30.8%) have an h-index of at least 10 and all 13 have published exclusively in top-quartile journals (Section 7.4).
- Collaboration analysis was performed on the basis of co-authorship between IMI-supported researchers as well as between co-authors. Around three-quarters (73.1%) of IMI researchers collaborated (co-authored) with at least one other IMI researcher during the period of analysis (Section 7.5).
- Again, co-authorship is more common between researchers in the same sector than among researchers in different sectors. Cross-sector co-authorship accounts for around two-fifths (39.9%) of all co-authorship activities during the analysis period (Section 7.6)
- The same is true of co-authorship activities by project. The majority of collaborative relationships are among researchers associated with the same project with only around one-quarter (25.5%) being cross-project. The share of cross-project activity between researchers has, however, increased slightly relative to the initial report (Section 7.6).

2 INTRODUCTION

2.1 OVERVIEW

The Innovative Medicines Initiative Joint Undertaking (IMI) has commissioned Thomson Reuters to undertake periodic evaluation of its research portfolio using bibliometric and intellectual property indicators.

The commissioned evaluation comprises a series of bi-annual reports focussing on research publications and patents produced by IMI funded researchers. This report is the second evaluation in the series.

2.2 INNOVATIVE MEDICINES INITIATIVE JOINT UNDERTAKING (IMI)

The Innovative Medicines Initiative Joint Undertaking (IMI) is a public private partnership between the European Union and the European Federations of Pharmaceutical Industries and Associations (EFPIA). The purpose of the IMI is to increase the efficiency and effectiveness of the drug development process, thereby increasing production of safer and more effective medicines. IMI pools resources from the public and private sectors and is funded jointly through Framework Programme Seven, EFPIA and EFPIA member companies. IMI supports pre-competitive pharmaceutical research and development to deliver new approaches, methodologies, and technologies.

With a €2 billion euro budget, IMI supports collaborative research projects and builds networks of industrial and academic experts in Europe that will boost innovation in healthcare. By acting as a neutral third party to support the creation of innovative partnerships, IMI aims to build a more collaborative ecosystem for pharmaceutical research and development (R&D).

IMI supports research projects in the areas of safety and efficacy, knowledge management and education and training. Projects are selected through open Calls for proposals. Project participants are recruited through these open and competitive Calls based on independent peer review and concluded by a Grant Agreement and Project Agreement.

The research consortia participating in IMI projects consist of:

- large biopharmaceutical companies that are members of EFPIA

and a variety of other partners, such as:

- small- and medium-sized enterprises,
- patients' organisations,
- universities and other research organisations,
- hospitals,
- regulatory agencies,
- any other industrial partners.

To date, IMI have announced eight Calls for proposals to be funded under the initiative. The first funding call was announced in 2008 and the latest, 8th, funding call was launched on 17th December 2012.

This report will cover the research outputs (publications and patent data) from the first three calls which have resulted in 30 projects.

2.3 THOMSON REUTERS

Thomson Reuters is the world's leading source of intelligent information for business and professionals. We combine industry expertise with innovative technology to deliver critical information to leading decision makers in the financial, legal, tax and accounting, healthcare, science and media markets, powered by the world's most trusted news organisation. Visit our WEBPAGE for more information.

2.4 THOMSON REUTERS RESEARCH ANALYTICS

Thomson Reuters Research Analytics is a suite of products, services and tools that provide comprehensive research analysis, evaluation and management. For over half a century we have pioneered the world of citation indexing and analysis, helping to connect scientific and scholarly thought around the world. Today, academic and research institutions, governments, not-for-profits, funding agencies, and all others with a stake in research need reliable, objective methods for managing and measuring performance. Visit our WEBPAGE for more information.

2.5 THOMSON REUTERS CUSTOM ANALYTICS & ENGINEERED SOLUTIONS

Thomson Reuters Custom Analytics & Engineered Solutions provide reporting and consultancy services within Research Analytics using customised analyses to bring together several indicators of research performance in such a way as to enable customers to rapidly make sense and interpret of a wide-range of data points to facilitate research strategy decision-making.

Our consultants have up to 20 years' experience in research performance analysis and interpretation. We have extensive experience with databases on research inputs, activity and outputs and have developed innovative analytical approaches for benchmarking, interpreting and visualisation of international, national and institutional research impact.

2.6 SCOPE OF THIS REPORT

One of IMI's principal objectives is to support collaborative research projects and build networks of industrial and academic experts in Europe. This will deliver socio-economic benefits to European citizens, increase Europe's competitiveness globally and establish Europe as the most attractive place for pharmaceutical R&D.

The analyses and indicators presented in this report have been specified to provide an analysis of IMI research output (publications and patent data) for research management purposes:

- To provide bibliometric indicators to identify excellence in IMI-supported research and to benchmark this research, where possible, overall and at individual project level
- To provide bibliometric indicators at individual researcher level
- To show that collaboration; at all levels, researcher, institutional and country, is being encouraged through the projects funded by IMI
- To provide an analysis of patent activity by IMI-funded researchers as an indication of increased engagement with industry and successful knowledge transfer

Outline of report

- Section 3 describes the data sources and methodology used in this report along with definitions of the indicators and guidelines to interpretation
- Sections 4 and 5 present citation analyses of IMI project publications overall with Section 4.7 providing summary analyses of IMI project publications identified since the initial report
- Section 6, which presents an analysis of IMI patent data, has not been updated in this second report. Patent data and analyses for IMI research will be updated in Report 3 (October 2013).[1]
- Section 7 presents bibliometric indicators for IMI-supported researchers and analyses of collaboration between these individuals

---

[1] For information on recent patent activity resulting from IMI-supported projects, details of five new patent applications from IMI-supported projects have been provided by IMI personnel and are listed in this Section.

3 DATA SOURCES, INDICATORS AND INTERPRETATION

3.1 BIBLIOMETRIC DATA AND CITATION ANALYSIS

3.1.1 BACKGROUND

Research evaluation is increasingly making wider use of bibliometric data and analyses. Bibliometrics is the analysis of data derived from publications and their citations. Publication of research outcomes is an integral part of the research process and is a universal activity. Consequently, bibliometric data have a currency across subjects, time and location that is found in few other sources of research-relevant data. The use of bibliometric analysis, allied to informed review by experts, increases the objectivity of and confidence in evaluation.

Research publications accumulate citation counts when they are referred to by more recent publications. Citations to prior work are a normal part of publication, and reflect the value placed on a work by later researchers. Some papers get cited frequently and many remain uncited. Highly cited work is recognised as having a greater impact and Thomson Reuters (*Evidence)* has shown that high citation rates are correlated with other qualitative evaluations of research performance, such as peer review.[2] This relationship holds across most science and technology areas and, to a limited extent, in social sciences and even in some humanities subjects.

Indicators derived from publication and citation data should always be used with caution. Some fields publish at faster rates than others and citation rates also vary. Citation counts must be carefully normalised to account for such variations by field. Because citation counts naturally grow over time it is essential to account for growth by year. Normalisation is usually done by reference to the relevant global average for the field and for the year of publication.

Bibliometric indicators have been found to be more informative for core natural sciences, especially for basic science, than they are for applied and professional areas and for social sciences. In professional areas the range of publication modes used by leading researchers is likely to be diverse as they target a diverse, non-academic audience. In social sciences there is also a diversity of publication modes and citation rates are typically much lower than in natural sciences.

Bibliometrics work best with large data samples. As the data are disaggregated, so the relationship weakens. Average indicator values (e.g. of citation impact) for small numbers of publications can be skewed by single outlier values. At a finer scale, when analysing the specific outcome for individual departments, the statistical relationship is rarely a sufficient guide by itself. For this reason, bibliometrics are best used in support of, but not as a substitute for, expert decision processes. Well-founded analyses can enable conclusions to be reached more rapidly and with greater certainty, and are therefore an aid to management and to increased confidence among stakeholders, but they cannot substitute for review by well-informed and experienced peers.

3.1.2 PUBLICATION AND CITATION DATA SOURCES

For this project, the Thomson Reuters data platform *ScienceWire*® has been used to identify publications associated with IMI funding and individual researchers. This platform has been developed to support program evaluation and research analytics using up-to-date multi-source data on research publications, funded research projects, patents and other research-related activities. It includes publications data from MEDLINE as well as the Thomson Reuters *Web of Science*® as well as data on other entities in publicly available and proprietary databases.

Citation data have been sourced from Thomson Reuters databases underlying the *Web of Knowledge*℠, which gives access to conference proceedings, patents, websites, and chemical structures, compounds and reactions in addition to journals. It has a unified structure that integrates

---

[2] *Evidence* Ltd. (2002) Maintaining Research Excellence and Volume: A report by Evidence Ltd to the Higher Education Funding Councils for England, Scotland and Wales and to Universities UK. (*Adams J, et al.)* 48pp .

all data and search terms together and therefore provides a level of comparability not found in other databases. It is widely acknowledged to be the world's leading source of citation and bibliometric data. The *Web of Science* is part of the *Web of Knowledge*, and focuses on research published in journals and conferences in science, medicine, arts, humanities and social sciences. The authoritative, multidisciplinary content covers over 12 000 of the highest impact journals worldwide, including Open Access journals and over 150 000 conference proceedings. 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'. Thomson Reuters (*Evidence)* 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.

Granularity of analysis is an important issue. Unduly fine analysis at the level of research groups provides little comparability or connectedness, while coarse analysis may miss spikes of excellence in key areas.

Journals are mapped to one or more subject categories, and every article within that journal is subsequently assigned to that category. Thomson Reuters (*Evidence*) uses these categories as the basis for bibliometric analysis because they are well-established and informed by extensive work with the research community since inception. Papers from prestigious, 'multidisciplinary' and general 'biomedical' journals such as Nature, Science, BMJ, The Lancet, New England Journal of Medicine and the Proceedings of the National Academy of Sciences (PNAS) are assigned to specific categories based on the journal categories of the citing and cited references in each article. Further information about the journals included in the citation databases and how they are selected is available here: http://scientific.thomsonreuters.com/mjl/.

The bibliometric evaluation of research covered in this report has been based principally on citation analysis of research published between January 2010 and January 2013 with citation counts as at end-January for all 'current' indicators and citation counts as at end-2011 for all indicators calculated with reference to world citation baselines (e.g. normalised citation impact).

Annex 4 provides the standard methodology and data definitions used in bibliometric and citation analyses. A summary of bibliometric and citation data definitions is given in Section 3.1.3.

3.1.3 BIBLIOMETRIC AND CITATION DATA DEFINITIONS AND INDICATORS

Citations: The citation count is the number of times that a citation has been recorded for a given publication since it was published. Not all citations are necessarily recorded since not all publications are indexed. However, the material indexed by Thomson Reuters is estimated to attract about 95% of global citations.

Citation impact: 'Citations per paper' is an index of academic or research impact (as compared with economic or social impact). It is calculated by dividing the sum of citations by the total number of papers in any given dataset (so, for a single paper, raw impact is the same as its citation count). Impact can be calculated for papers within a specific research field such as Clinical Neurology, or for a specific institution or group of institutions, or a specific country. Citation count declines in the most recent years of any time-period as papers have had less time to accumulate citations (papers published in 2007 will typically have more citations than papers published in 2010).

Citation velocity/hot papers: Citation velocity is the rate at which a paper accumulates citations. Most papers reach their citation peak some time after publication. A small number of papers, however, accumulate citations rapidly (high citation velocity) and may represent breakthroughs in the field(s) to which they relate.

Field-normalised citation impact (NCI_F): Citation rates vary between research fields and with time, consequently, analyses must take both field and year into account. In addition, the type of publication will influence the citation count. For this reason, only citation counts of papers (as defined above) are used in calculations of citation impact. The standard normalisation factor is the world average citations per paper for the year and journal category in which the paper was published. This normalisation is also referred to as 'rebasing' the citation count.

H-index: The h-index was developed by JE Hirsch as in indicator of both productivity and impact.[3] The value of the index h is equal to the number of papers (N) in the list that have N or more citations, while the remaining papers have fewer than N citations. Therefore, a researcher who has published 30 papers, of which 17 have received 17 or more citations while the remaining 13 have received fewer than 17 citations, has an h-index of 17. Irrespective of research impact, older researchers in more prolific fields tend to have a higher h-index.

Thomson Reuters **Hot Papers** database tracks and identifies papers with high citation velocities relative to their field and age. To identify hot papers, papers published in the last two years are selected and frequency distributions compiled for citations received in the most recent two-month period. To correct for variation in citation rates between different research fields, separate distributions are made for each field. The 22 *Essential Science Indicators*® fields used in this classification are documented here: http://archive.sciencewatch.com/about/met/fielddef/. Thresholds are set to find the top fraction of papers in each field – typically 0.1% of papers meet this threshold and are classified as **hot papers**.

Interdisciplinarity/diffusion score: This is indicated by the number and disparateness of the fields from which publications citing an IMI publication originate, summarised in a diffusion score developed by Carley and Porter.[4] The diffusion score is a measure of the applicability of new knowledge across subject areas and represents a measure of the robustness of the findings in the published article. The diffusion score incorporates features of traditional measures of diversity in assessing the balance and distribution of citations arising from different subject categories that in substance very different from one another. For example, while an article A receiving 5 citations from Physics, Applied and 5 citations from Chemistry, Physical and an article B receiving 5 citations from Physics, Applied and 5 citations from Physiology would have the same diversity, the diffusion score would be greater for article B since the two fields from which the citations originate are very different from one another.

Journal-normalised citation impact (NCI_J): Another bibliometric indicator which can be very useful in small datasets is the journal-normalised citation impact, NCI_J. This indicator is calculated from the citation impact relative to the specific journal in which the publication appears.

For the publication in Annex 4 which has been cited 71 times to the end-December 2010, the expected citation rate for a publication in Acta Biomaterialia published in 2005 would be 18.6 and the NCI_J would be 3.82. Therefore, this publication has been cited more than expected for the journal.

For a set of publications, we calculate the quality index as the percentage of publications which are cited more than expected for the relevant journals.

This indicator should be considered alongside that of field-normalised citation impact as they are complementary. For example, a given set of publications may have a high quality index and relatively low average field-normalised citation impact. This would imply that these publications were well cited in relation to other papers in that journal and that year but when considered in relation to other publications in the same research field did not perform as well. The interpretation would be that the publications are in relatively low impact journals.

Journal Impact Factor (JIF): In the same way that citation impact can be used as an index of research quality, the average number of citations per paper can be used to indicate the impact and/or importance of a journal. The Impact Factor for a journal (JIF) is calculated using data for a three-year period. For example, the 2011 Impact Factor for a given journal is calculated is calculated by Thomson Reuters as the average number of times which articles from the journal published in the past two years (2009 and 2010) were cited in 2011. Thus, a JIF of 2.0 means that, on average, the articles published in 2009 or 2011 have been cited twice. Citing articles may be from the same journal; however, most citing articles are from other journals.

---

[3] Hirsch, J.E. (2005) *Proceedings of the National Academy of Sciences of the United States of America* **102** (46): 16569-16572

[4] Carley S, Porter A (2012). A forward diversity index. Scientometrics, 90:407-427.

For the journal *Vaccine*, the 2011 journal Impact Factor would be calculated as follows:

Cites in 2011 to items published in 2010 =	3 729	Number of items published in 2010 =	1 105
Cites in 2011 to items published in 2009 =	4 702	Number of items published in 2009 =	1 134
Total	**8 431**		**2 239**

$$\frac{Number\ of\ citations}{Number\ of\ \text{items}} = \frac{8\,431}{2\,239} = \mathbf{3.766}$$

The calculation of the journal Impact Factor is fully described on the Thomson Reuters website at: http://thomsonreuters.com/products_services/science/free/essays/impact_factor/.

When looking at journal Impact Factor data it is important to remember that, as citation rates vary between research fields and publication type, these will affect the JIF. That is a JIF of 3.766 ranks the journal *Vaccine* 23rd out of 109 journals in the Research & Experimental Medicine journal category and therefore in the top quartile. However, the journal *Cell Calcium* with the same JIF of 3.766 is ranked in the second quartile (71st out of 178 journals) in the journal category Cell Biology.

Journal top quartile: This indicator is defined as the quartile in which the journal appears when ranked by Journal Impact Factor among all journals in that category.

Mean normalised citation impact (mNCI): The mean NCI indicator for any specific dataset is calculated as the mean of the field-normalised citation impact (NCI_F) of all papers within that dataset.

Papers/publications: Thomson Reuters abstracts publications including editorials, meeting abstracts and book reviews as well as research journal articles. The terms 'paper' and 'publication' are often used interchangeably to refer to printed and electronic outputs of many types.

For clarity, in this report:

- **Publication** is used inclusively to refer to all IMI publications whether linked to Thomson Reuters citation data or not.
- ***Web of Science* publication** is used exclusively to refer to those IMI publications which have been linked to Thomson Reuters citation data.
- **Paper** is used exclusively to refer only to substantive *Web of Science* publications (journal articles, reviews and some proceedings papers) that have been linked to Thomson Reuters citation data. This definition excludes editorials, meeting abstracts or other types of publication. Papers are the subset of publications for which citation data are most informative and which are used in calculations of citation impact.

Percentage of highly-cited papers: For the purpose of this report, highly-cited papers have been defined as those articles and reviews which belong to the world's top decile of papers in that journal category and year of publication, when ranked by number of citations received. A percentage that is above 10 indicates above-average performance.

Research field: Standard bibliometric methodology uses journal category as a proxy for research field. Journals are assigned to one or more categories, and every article within that journal is subsequently assigned to that category. Publications from prestigious, 'multidisciplinary' and general medical journals such as Nature, Science, The Lancet, BMJ, The New England Journal of Medicine and the Proceedings of the National Academy of Sciences (PNAS) are assigned to specific categories based on the journal categories of the references cited in the article. The selection procedures for the journals included in the citation databases are documented here http://scientific.thomsonreuters.com/mjl/. For this evaluation, the standard classification of *Web of Science* journal categories has been used.

3.1.4 INTERPRETATION OF BIBLIOMETRIC INDICATORS AND CITATION ANALYSES

The following points should be borne in mind when considering the results of these analyses.

- IMI JU only started to fund projects in May 2009. Of the 40 active projects 14 (almost half) were launched since 1 January 2011. It may take several years for a project to progress from inception to the point where it has generated sufficient data for a publication. It may take further years until it has produced its most valuable results. The IMI JU projects that will be analysed are therefore relatively young, and early bibliometric indicators may not fully reflect their eventual impact.

- Although additional papers for the authors have been identified by our analysts this is still a relatively small dataset. Bibliometrics work best with large data samples. As the data are disaggregated, so the relationship weakens. Average indicator values (e.g. of citation impact) for small numbers of publications can be skewed by single outlier values. At a finer scale, when analysing the specific outcome for individual departments, the statistical relationship is rarely a sufficient guide by itself. For this reason, bibliometrics are best used in support of, but not as a substitute for, expert decision processes. Well-founded analyses can enable conclusions to be reached more rapidly and with greater certainty, and are therefore an aid to management and to increased confidence among stakeholders, but they cannot substitute for review by well-informed and experienced peers.
- As noted above most of the publications associated with IMI JU-funded projects are relatively recent. Publications accumulate citations over time and it may take years until a given publication is cited. While citation counts in early years have been shown to reflect long-term citation performance,[5] indicators based on citation counts may be relatively more volatile in the years immediately following publication.
- 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 biomedical sciences than physical sciences, and natural sciences generally cite at a higher rate than social sciences.

INDICATOR THRESHOLDS

Papers: The minimum number of papers suitable as a sample for quantitative research evaluation is a subject of widespread discussion. Larger samples are always more reliable, but a very high minimum may defeat the scope and specificity of analysis. Experience has indicated that a threshold between 20 and 50 papers can generally be deemed appropriate. For work that is likely to be published with little contextual information, the upper boundary (≥ 50) is a desirable starting point. For work that will be used primarily by an expert, in-house group then the lower boundary (≥ 20) may be approached. Because comparisons for in-house evaluation often involve smaller, more specific research groups (compared to broad institutional comparisons) a high volume threshold is self-defeating. Smaller samples may be used but outcomes must be interpreted with caution and expert review should draw on multiple information sources before reaching any conclusions.

Field normalised citation impact: such values for individual papers vary widely and it is more useful to consider the average for a set of papers. This average can be at several granularities: field (either journal category or field), annual and overall (total output under consideration). When considering such average data points, care must be taken to understand that these data are highly skewed and the average can be driven by a single, highly-cited paper (this would be highlighted in accompanying text though not apparent from Tables & Figures). The world average is 1.0, so any value higher than this indicates a paper, or set of papers, which are cited more than average for similar research worldwide. For research management purposes, experience suggests that values between 1.0 and 2.0 should be considered to be indicative of research which is influential at a national level whilst that cited more than twice the world average has international recognition.

Research field: A problem frequently encountered in the analysis of data about the research process is that of ‘mapping’. For example, a funding body allocates money for chemistry but this goes to researchers in biology and engineering as well as to chemistry departments. Clinicians publish in mathematics and education journals. Publications in environmental journals come from a diversity of

[5] Adams, J. et al. (2002) Maintaining Research Excellence and Volume: A report by Evidence Ltd to the Higher Education Funding Councils for England, Scotland and Wales and to Universities UK, http://www.hefce.ac.uk/pubs/rdreports/2002/rd08_02/rd08_02.pdf

disciplines. This creates a problem when we try to define, for example, 'Parasitology research'. Is this the work funded under Parasitology programmes, the work of researchers in Parasitology units or the work published in Parasitology journals? For the first two options we need to track individual grants and researchers to their outputs, which is feasible but not within the scope of this study nor for every comparator institution. Therefore, to create a simple and transparent dataset of equal validity across time and geography, we rely on the set of journals associated with Parasitology as a proxy for the body of research reflecting the field.

3.1.5 DATASET DEFINITIONS USED IN THE BIBLIOMETRIC INDICATORS AND CITATION ANALYSES

IMI researcher publications/papers: This dataset comprises publications by IMI-supported researchers as described in Section 7 and outlined in Figure 7.1.1.

IMI project publications/papers: This dataset comprises publications from IMI-supported projects as described in Section 4.1 and outlined in Figure 4.1.1.

Similar European research: this benchmark dataset has been created using the EU-27 grouping of countries: Thomson Reuters *National Science Indicators* 2011 database and only research falling into the same journal categories as in the IMI project dataset.

3.2 PATENT DATA AND ANALYSIS

Thomson Reuters patent data and analyses for IMI research will be updated in Report 3 (October 2013).

For information on recent patent activity resulting from IMI-supported projects, details of five new patent applications from IMI-supported projects have been provided by IMI personnel and are listed in Section 6.

4 CITATION ANALYSIS – IMI-SUPPORTED PUBLICATIONS OVERALL

This Section of the report presents analyses of the output and citation impact of IMI projects considered overall and compared to the IMI-researcher dataset collated for all researchers supported by IMI (Section 7.1) and similar European research (see footnote on page 24).

Publications for analyses include all IMI-supported publications identified in *Web of Science* to date – that is, publications new to this report (Section 4.7) as well as publications identified in the initial report. The census point for inclusion of publications in the initial report was mid-August 2012. The updated second report therefore reflects changes in IMI activity since that point.

Citation counts have been updated from the mid-August census point used in the initial report.

4.1 PUBLICATIONS FROM IMI-SUPPORTED PROJECTS

Publications from IMI-supported projects were identified using bibliographic data supplied by IMI, or through specific keyword searches using funding acknowledgment data in Thomson Reuters *Web of Science.*

The aggregated list of publications was reviewed by Thomson Reuters (*Evidence*) and supplied to IMI for further verification prior to inclusion in the analyses. Four publications have not been assigned to specific projects despite review by IMI personnel.

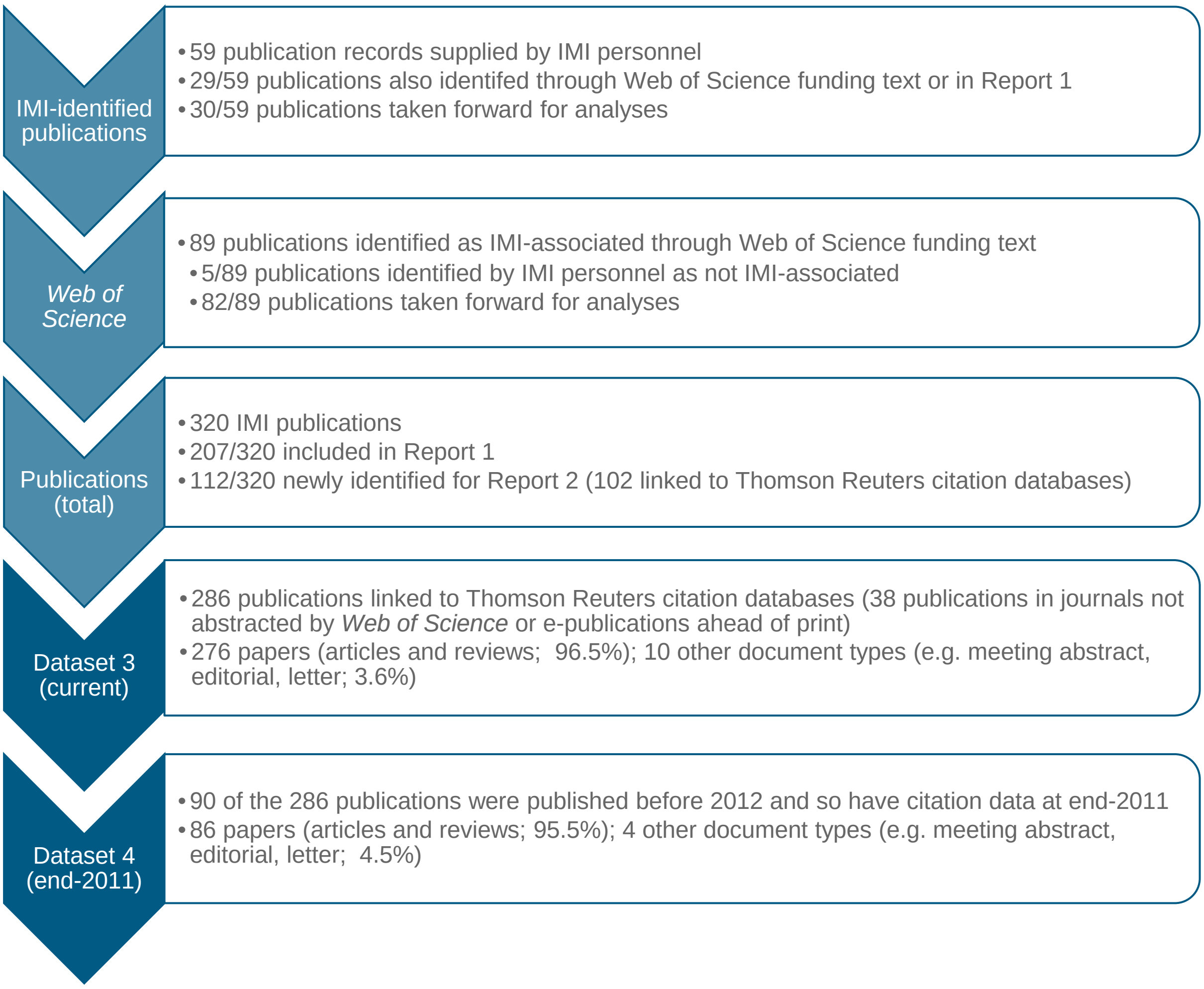
The process of identifying publications from IMI-supported projects which have Thomson Reuters citation data is outlined in Figure 4.1.1.

Two datasets are used in the citation analyses:

- Dataset 3: This is an expansion of Dataset 1 in the initial report and comprises 286 *Web of Science* publications. Of these 286, 102 are new to this report.
- Dataset 4: This includes one additional paper with normalised citation impact data at end-2011 but is otherwise unchanged relative to Dataset 2 in the initial report (see Section 4.7).

To benchmark IMI project research performance, these cumulative datasets are also compared to publications collated for all researchers supported by IMI in Calls 1, 2 and 3 (Section 7.1) and similar European research (see footnote on page 24).

FIGURE 4.1.1 IDENTIFYING PUBLICATIONS FROM IMI-SUPPORTED PROJECTS WITH THOMSON REUTERS CITATION DATA



Dataset 3 is an expansion of Dataset 1 used in the initial report. Dataset 4 includes one additional paper as described earlier. The other new papers were abstracted in *Web of Science* in 2012 and, as the world citation baselines for 2012 are not yet available, the normalised citation impact of these papers cannot yet be calculated.

World citation baselines for 2012 will be available in March and normalised citation impact indicators for IMI-supported papers will therefore be updated in the third report (October 2013).

4.1.1 CITATION DATA FOR PUBLICATIONS FROM IMI-SUPPORTED PROJECTS

A total of 320 publications resulting from IMI-supported projects were identified and 286 of these publications linked to records in *Web of Science*. Citation counts have been sourced from the citation databases which underlie Thomson Reuters *Web of Knowledge* and have been extracted at two distinct census points – current (end-January 2013) and end-2011.

The 'current' census point (Dataset 3) allows assessment of the performance of IMI research from as up-to-date a viewpoint as possible through calculation of 'raw' citation impact (see Section 3.1.3). This, however, does not allow benchmarking of IMI research performance against the world and European average.

Dataset 4 is used to evaluate the citation impact of IMI-supported research relative to world average (normalised citation impact) and has same end-2011 census point used in the calculation of global citation baselines (see Section 3.1.3).

Normalised bibliometric indicators for Dataset 4 were calculated using standard methodology and the Thomson Reuters *National Science Indicators* (NSI) database for 2011.

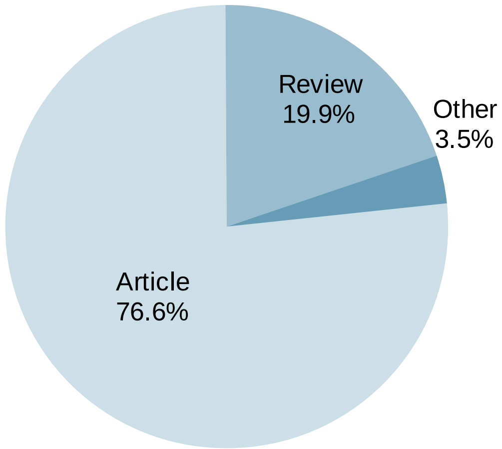
4.2 SHARE OF PAPERS RELATIVE TO OTHER PUBLICATION TYPES

FIGURE 4.2.1 CATEGORISATION OF IMI PROJECT RESEARCH BY DOCUMENT TYPE

Figure 4.2.1 shows the share of articles and reviews (papers) relative to other document types, for all *Web of Science* publications from IMI-associated projects. Papers are the subset of publications for which citation data are most informative and which are used in calculations of normalised citation impact.

IMI project research comprises 286 unique *Web of Science* publications linked to citation data (as outlined in Figure 4.1.1). Over 95% of these documents (96.5%) were substantive articles and reviews with only ten documents not falling into this grouping. These documents (classified as 'Other') include five editorials, two meeting abstracts, two letters and one news item.

The distribution of document types is similar to that observed in the initial report to IMI.



4.3 TRENDS IN PUBLICATION OUTPUT

Figure 4.3.1 compares the annual numbers of *Web of Science* publications between this second report and the initial report to IMI.

Publication output has increased substantially in successive years from 2010 to 2012. Output in 2012 has more than doubled from 2011 with just over one-third (35.0%) of these publications new to this report (Section 4.7).

FIGURE 4.3.1 NUMBER OF *WEB OF SCIENCE* PUBLICATIONS BY YEAR AND REPORT

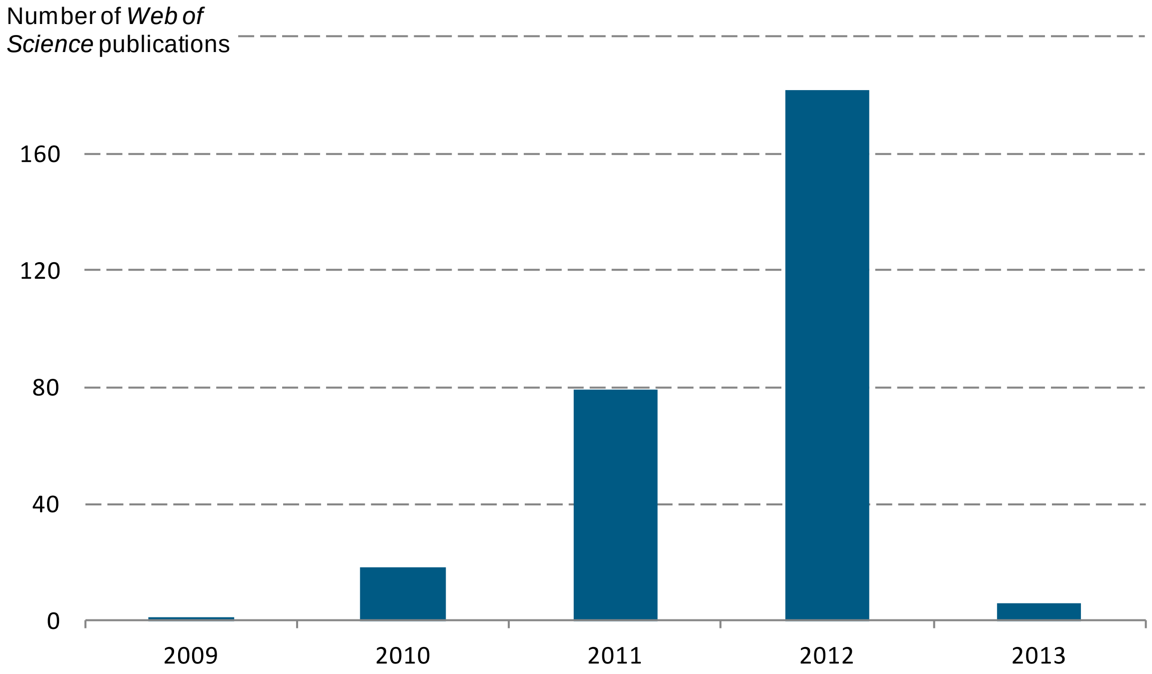
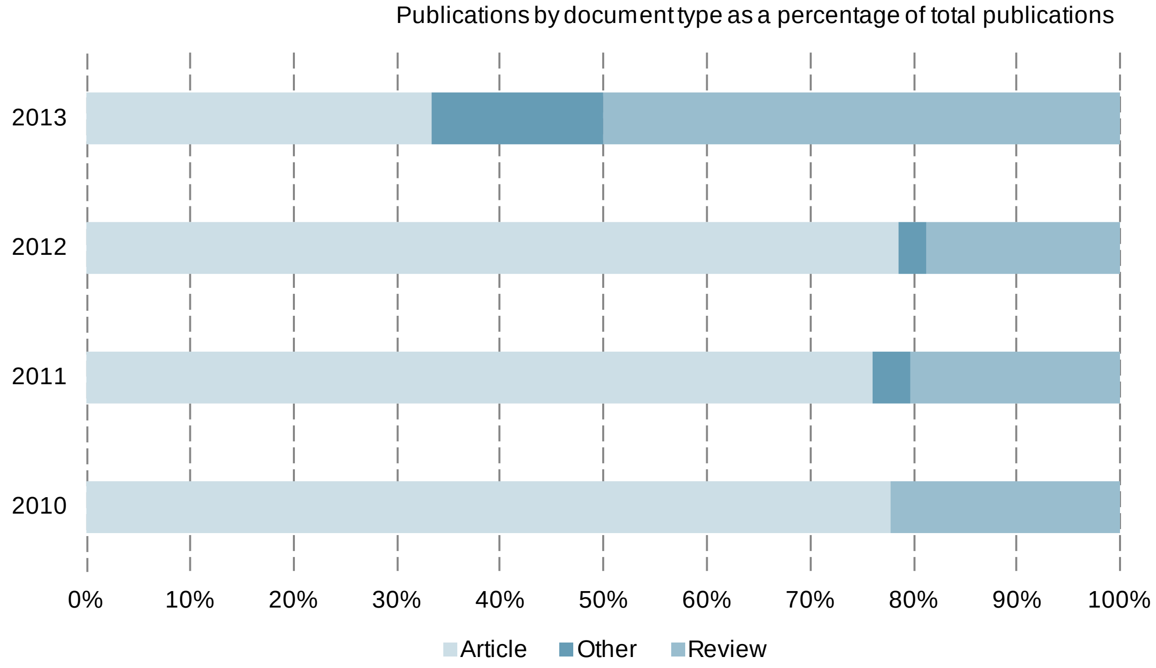


Figure 4.3.2 shows the proportion of papers (articles and reviews) relative to other document types for IMI project research from 2010 to date.[6]

IMI projects continue to generate a high proportion of papers relative to other document types with reviews typically accounting for around one-fifth of output.

[6] 2009 publications comprise a single meeting abstract – this has been omitted from Figure 5.3.2 for clarity.

FIGURE 4.3.2 CATEGORISATION OF *WEB OF SCIENCE* PUBLICATIONS BY YEAR AND DOCUMENT TYPE



4.4 IN WHICH JOURNALS DO IMI PROJECT PUBLICATIONS APPEAR MOST FREQUENTLY?

The 11 journals appearing most frequently[7] in the IMI project publications dataset, 2009-2013 are listed in Table 4.4.1. A total of 49 journal titles are used more than once.

Together, the 11 most frequently used journals cover 74 *Web of Science* publications, just over one-quarter (26.0%) of the total number of items in the dataset.

This core set of journals continues to highlight the diversity of IMI-supported projects – the top 11 journals include titles focused on bioinformatics, rheumatology, oncology and genetics. Bioinformatics- and genetics-focused journals were not in the most frequently used journals in the initial report.

All journals in Table 4.4.1 are in the top quartile when ranked by Journal Impact Factor among all journals in that category.

IMI project publications have been published in a total of 171 journals, of which 121 are ranked in the top quartile (by Journal Impact Factor) of journals in their specific research fields. A total of 227 publications (79.4% of IMI project publications) have been published in these well regarded journals

[7] Table 4.4.1 uses the same frequency threshold of at least three publications as in the initial report.

TABLE 4.4.1 JOURNALS IN WHICH IMI PROJECT PUBLICATIONS HAVE BEEN PUBLISHED MOST FREQUENTLY (2009-2013), RANKED BY NUMBER OF *WEB OF SCIENCE* PUBLICATIONS

Journal	Number of *Web of Science* publications	Number of papers	Journal Impact Factor (2011)	Journal categories
PLOS One	13	13	4.092	Biology
Molecular Informatics	10	10	2.390	Medicinal Chemistry; Mathematical & Computational Biology
Pain	8	8	5.777	Anaesthesiology; Clinical Neurology; Neurosciences
Annals of the Rheumatic Diseases	7	6	8.727	Rheumatology
Arthritis and Rheumatism	6	5	7.866	Rheumatology
European Journal of Cancer	6	6	5.536	Oncology
Diabetes	5	5	8.286	Endocrinology & Metabolism
Diabetologia	5	5	6.814	Endocrinology & Metabolism
Journal of Clinical Investigation	5	5	13.069	Research & Experimental Medicine
Pharmacogenomics	5	5	3.974	Pharmacology & Pharmacy
Drug Discovery Today	4	4	6.828	Pharmacology & Pharmacy

Table 4.4.2 lists the 20 journals with highest Journal Impact Factor (JIF) in the IMI-supported publications dataset.

Overall, there are 34 publications in journals with an impact factor of 10 or above, an increase of 11 publications over the initial report. Of these, some 13 publications appear in journals with an impact factor of 20 or above, 5 more than in the initial report. These new publications include elite multidisciplinary journals as noted in Section 4.7.

Together the top 20 journals by Journal Impact Factor account for more than one-tenth (11.9%) of all IMI-supported publications indicating that IMI research continues to be published in highly-regarded journals.

TABLE 4.4.2 JOURNALS IN WHICH IMI PROJECT PUBLICATIONS HAVE BEEN PUBLISHED MOST FREQUENTLY (2009-2013), TOP TWENTY RANKED BY JOURNAL IMPACT FACTOR

Journal	Number of *Web of Science* publications	Number of papers	Journal Impact Factor (2011)	Journal categories
Nature	2	2	36.280	Multidisciplinary Sciences
Nature Genetics	3	1	35.532	Genetics & Heredity
Science	1	1	31.201	Multidisciplinary Sciences
JAMA-Journal of the American Medical Association	1	1	30.026	General & Internal Medicine
Nature Reviews Drug Discovery	2	0	29.008	Biotechnology & Applied Microbiology; Pharmacology & Pharmacy
Nature Immunology	1	1	26.008	Immunology
Lancet Neurology	2	2	23.462	Clinical Neurology

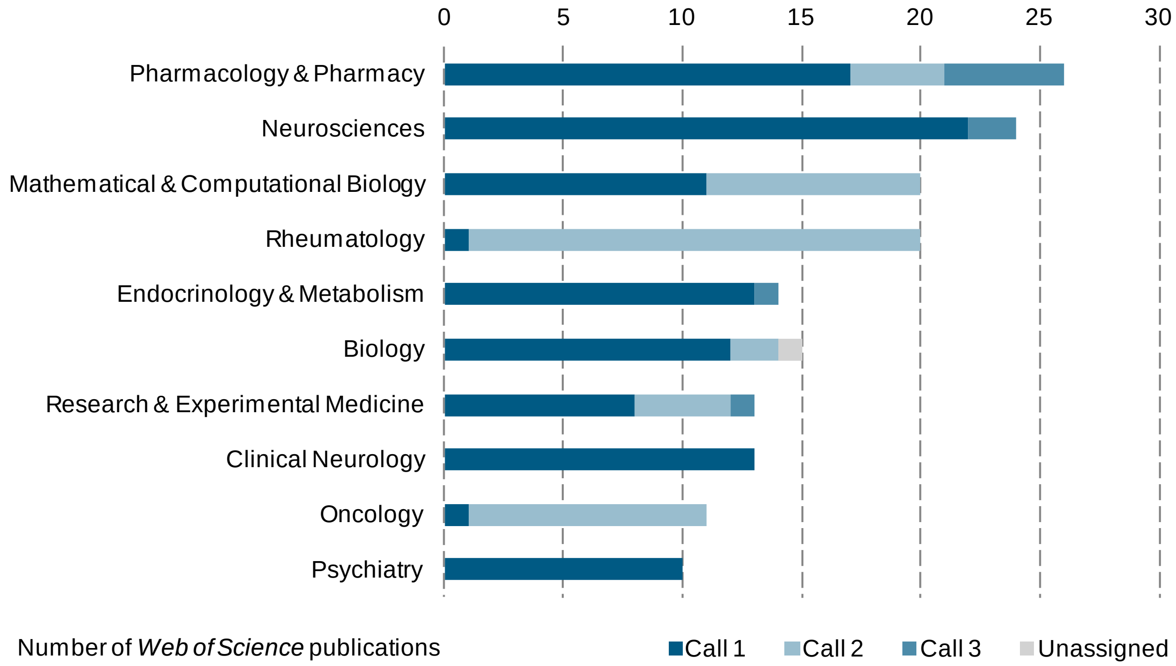
Journal	Number of *Web of Science* publications	Number of papers	Journal Impact Factor (2011)	Journal categories
Nature Biotechnology	1	0	23.268	Biotechnology & Applied Microbiology
PLOS Medicine	1	1	16.269	General & Internal Medicine
Nature Neuroscience	1	1	15.531	Neurosciences
British Medical Journal	2	2	14.093	General & Internal Medicine
Journal of Experimental Medicine	1	1	13.853	Research & Experimental Medicine
Molecular Psychiatry	3	3	13.668	Biochemistry & Molecular Biology; Neurosciences; Pyschology
Journal of Clinical Investigation	5	5	13.069	Research & Experimental Medicine
American Journal of Psychiatry	1	1	12.539	Psychiatry
Nature Reviews Neurology	1	1	12.461	Clinical Neurology
Nature Reviews Clinical Oncology	1	1	11.963	Oncology
Journal of Allergy and Clinical Immunology	3	3	11.003	Allergy; Immunology
Trends in Pharmacological Sciences	1	1	10.927	Pharmacology & Pharmacy
ACS Nano	1	1	10.774	Chemistry, Multidisciplinary; Materials Science, Multidisciplinary; Nanoscience & Nanotechnology; Physical Chemistry

4.5 WHICH RESEARCH FIELDS ACCOUNT FOR THE HIGHEST VOLUME OF IMI PROJECT PUBLICATIONS?

Figure 4.5.1 shows the top ten *Web of Science* journal categories[8] associated with IMI project research.

[8] Journals can be associated with more than one *Web of Science* category. This analysis is based on best-performing category (i.e. that in which it ranks highest in terms of overall citations relative to journal category and year).

FIGURE 4.5.1 TOP TEN *WEB OF SCIENCE* JOURNAL CATEGORIES IN WHICH IMI PROJECT RESEARCH IS PUBLISHED



- Almost one-tenth (9.5%) of IMI project research is assigned to the journal category of Pharmacology & Pharmacy which has overtaken Neurosciences (8.4%) as the most frequently used journal category in this dataset.
- Mathematics & Computational Biology was not amongst the most frequently used journal categories in the initial report to IMI. Output in this category has expanded from five to 13 publications, the majority of which are associated with eTOX or Open PHACTS.
- Output in Rheumatology journals has more than doubled since the initial report; all but one of the publications in this category are associated with the BTCure project (Call 2) with one publication from PROTECT (Call 1). This reinforces that BTCure has been quick to publish in scientific journals as observed in the initial report.
- Endocrinology & Metabolism now includes substantial output from SUMMIT as well as IMIDIA (both Call 1); 4 and 6 publications, respectively.
- Anaesthesiology is no longer among the top ten most frequently used journal categories as output in this journal category has not increased.
- The most frequently used journal categories in Figure 4.5.1 continue to reflect breadth and depth in the IMI research portfolio which contains cross-cutting and more specialised research.

The analysis presented in Figure 4.5.1 includes all publication types and spans the full time period of IMI-supported publications (2009-2013). As publication numbers are small and citation impact is calculated only on papers, the identification of the top ten journal categories, when considering citation impact, differs from that above (Section 4.6).

Standard descriptions of the scope of these journal categories are given in Annex 1.

4.6 IS IMI PROJECT RESEARCH WELL-CITED?

Citation impact of research, an indicator linked to the accumulation of citations, is subject specific. Typically, papers published in areas such as biomedical research receive more citations than papers published in subjects such as engineering even if the papers are published in the same year. All citation impact data presented in this report are therefore normalised, or rebased, to the relevant world average to allow comparison between years and fields.

Tables 4.6.1 and 4.6.2 present a summary of the citation analyses of research from IMI-supported projects compared with the IMI researcher dataset. Table 4.6.1 presents a viewpoint of IMI-supported papers at the end of 2011 using indicators where citation impact has been normalised against world average values. Table 4.6.2 presents a more recent (but also more descriptive) viewpoint using indicators based on current (end-January) citation counts (see Section 4.1.1)

The average citation impact for IMI project papers is 1.55 (where world average is 1.0) for the 2-year period, 2010-2011. This has increased from 1.34 reported in the initial report due to the inclusion of a single, highly-cited paper. This newly-identified paper is associated with the Call 1 project Pharma-Cog and is also designated as a ‘hot paper’ (Annex 2).

For comparison, the EU’s average citation impact[9,10] relative to world baseline for the same 2-year period in similar research fields was 1.14 and for the IMI researchers dataset was 2.36 (Table 4.6.1). The average citation impact of European benchmark research has not changed from that in the initial report as this dataset is unchanged (Section 4.1.1). The average citation impact of research from IMI-supported researchers, however, has changed as this benchmark dataset now includes researchers from Calls 2 and 3 in addition to those working on projects funded in Call 1.

Similarly, using current citation counts and raw citation impact, IMI project research performs less well compared to the IMI-supported researchers dataset (3.96 compared to 6.76; Table 4.6.2).

Overall, these data suggest that the IMI-supported researchers typically publish well-cited papers but that those papers associated with IMI projects are not cited as frequently. This may be due to the more applied nature of these papers supported by IMI funding compared to the typical more academic publication output of the researchers.

However, other indicators such as the average percentile and % *Web of Science* publications in top quartile journals suggest that IMI-supported papers remain more likely to be published in a well-regarded journal than typical research published by IMI-supported researchers (this corroborates the journal analysis in Section 4.4).

More than one-tenth of publications from IMI projects or IMI-supported researchers are published in open access journals (11.3% and 11.1%, respectively – Table 4.6.2). These values are above the global average reported by a 2011 study reviewing accessibility in the journal literature[11] though it should be borne in mind that most IMI research has been published since 2009 and the share of publications that are openly accessible is likely to have grown.

The IMI projects dataset is small in comparison with the IMI researchers dataset and these analyses should not be taken as evidence of poor performance rather that IMI project funding is being awarded to researchers that perform at a high overall level.

---

[9] EU-27 grouping of countries: Thomson Reuters *National Science Indicators* 2011 database; similar research has been defined as including the same journal categories as in the IMI project dataset.

[10] For this analysis, only papers are considered since only these publication types have normalised citation impact data (see Section 3.1.3). Two of the top journal categories in Figure 4.5.1 (Rheumatology and Oncology) drop out of the analyses in Figure 4.6.1 because most publications in these categories were published in 2012 and therefore do not yet have normalised citation impact data.

[11] Laakso *et al*. (2011) The Development of Open Access Journal Publishing from 2003 to 2009, PLOS ONE, 6(6), e20961.

TABLE 4.6.1 SUMMARY CITATION ANALYSIS FOR IMI RESEARCH – CITATIONS TO END-2011

		Citation impact			
2010-2011	**Number of papers**	**Normalised at field level**	**Normalised at journal level**	**Average percentile**	**% Highly-cited papers**[12]
IMI projects	86	1.55	0.98	65.69	11.6%
IMI researchers	5 706	2.23	1.29	52.08	21.2%

TABLE 4.6.2 SUMMARY CITATION ANALYSIS FOR IMI RESEARCH – CITATIONS TO CURRENT

	IMI publications		***Web of Science* publications**			
2010 to current	**Total**	**% Open access journals**[13]	**Total**	**Citations**	**Raw citation impact**	**% Top quartile journals**[14]
IMI projects	320	11.3%	286	1 133	3.96	79.4%
IMI researchers	n/a	11.1%	11 444	77 315	6.76	70.2%

Disaggregation by journal category shows strengths in the IMI project publications dataset.

Figure 4.6.1 shows that the citation impact of IMI project research in the majority of the journal categories is, on average, well above the citation impact of similar European research.

IMI project research in Biology, Clinical Neurology and Psychiatry has higher citation impact, on average, than similar research by IMI-supported researchers.

[12] 'Highly-cited' refers those articles and reviews belonging to the world's top decile of papers for journal category and year of publication. A percentage that is above 10 indicates above-average performance.

[13] For this report, we have considered a journal as open access if listed in the Directory of Open Access journals (http://www.doaj.org/)

[14] This indicator is based upon the quartile in which the journal appears when ranked by Journal Impact Factor among all journals in that category. Journal ranking data have been sourced from Thomson Reuters *Journal Citation Reports* database.

FIGURE 4.6.1 CITATION IMPACT OF IMI-SUPPORTED PAPERS, BY RESEARCH FIELD (JOURNAL CATEGORY) BENCHMARKED AGAINST PAPERS BY IMI-SUPPORTED RESEARCHERS AND SIMILAR PAPERS FROM THE EUROPEAN RESEARCH BASE

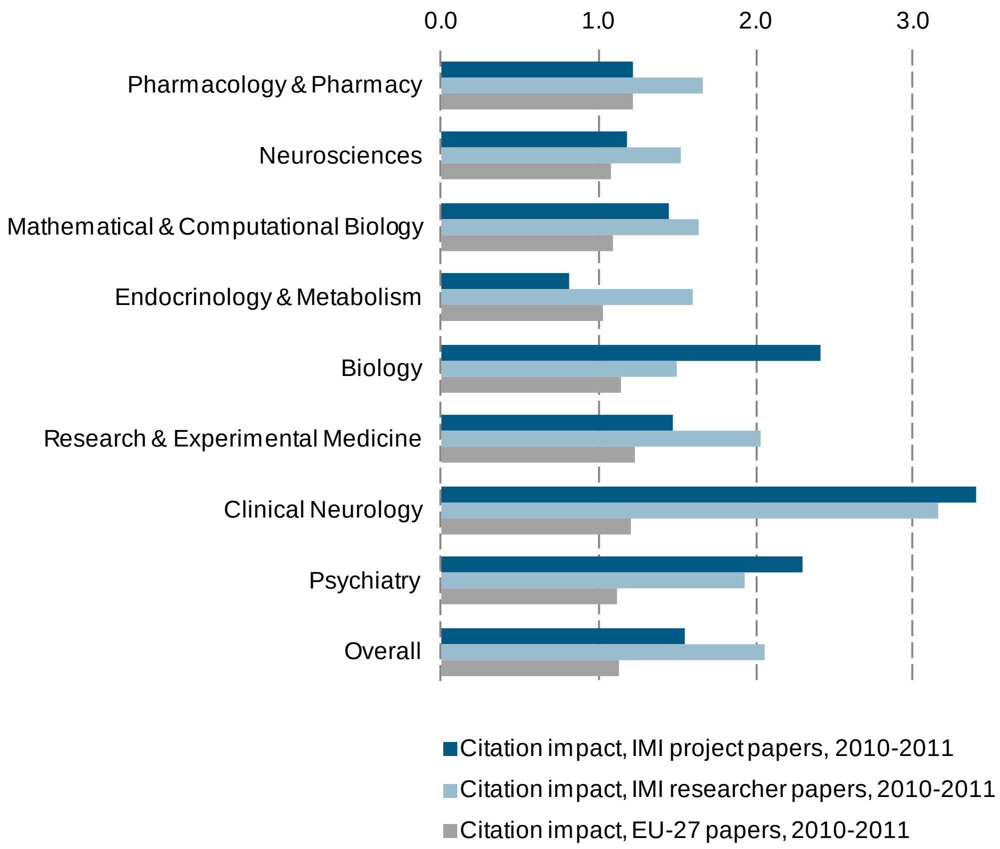


TABLE 4.6.3 SUMMARY OF PUBLICATION OUTPUT AND 2-YEAR AVERAGE CITATION IMPACT FOR IMI-SUPPORTED RESEARCH BY TOP *WEB OF SCIENCE* JOURNAL CATEGORIES, 2010-2011 BENCHMARKED AGAINST IMI RESEARCHERS DATASET AND SIMILAR PUBLICATIONS FROM THE EUROPEAN RESEARCH BASE

	IMI projects		IMI researchers		EU-27	
***Web of Science* journal category**	**Number of papers**	**Citation impact**	**Number of papers**[15]	**Citation impact**	**Number of papers**[14]	**Citation impact**
Pharmacology & Pharmacy	13	1.22	554	1.66	21 999	1.22
Neurosciences	21	1.18	552	1.53	26 862	1.07
Mathematical & Computational Biology	5	1.44	94	1.64	4 091	1.09
Endocrinology & Metabolism	7	0.81	330	1.59	13 135	1.03
Biology	5	2.41	350	1.50	17 551	1.15

[15] Papers can be assigned to more than one journal category and so may be counted towards the number of papers in more than one category.

Web of Science journal category	IMI projects		IMI researchers		EU-27	
	Number of papers	Citation impact	Number of papers[15]	Citation impact	Number of papers[14]	Citation impact
Research & Experimental Medicine	3	1.48	151	2.03	9 400	1.23
Clinical Neurology	12	3.40	341	3.16	18 856	1.20
Psychiatry	5	2.30	339	1.93	12 567	1.11
Overall	**86**	**1.55**	**5 709**	**2.23**	**568 436**	**1.14**

It is important to note that IMI projects have many fewer papers in each of these categories than either benchmark and that low paper numbers can mean that citation impact values will be more susceptible to skew by especially well-cited papers or large numbers of uncited papers.

These analyses therefore give a useful early indication of IMI project research performance relative to comparators but it should be borne in mind that this performance may change as IMI paper numbers increase.

Standard definitions of the scope of the journal categories in Figure 4.6.1 and Table 4.6.1 are given in Annex 1

4.7 BIBLIOMETRIC ANALYSIS – NEW IMI-SUPPORTED PUBLICATIONS

This Section of the report presents summary analyses of IMI publications identified since mid-August 2012. These summary analyses should be borne in mind when considering IMI project research overall.

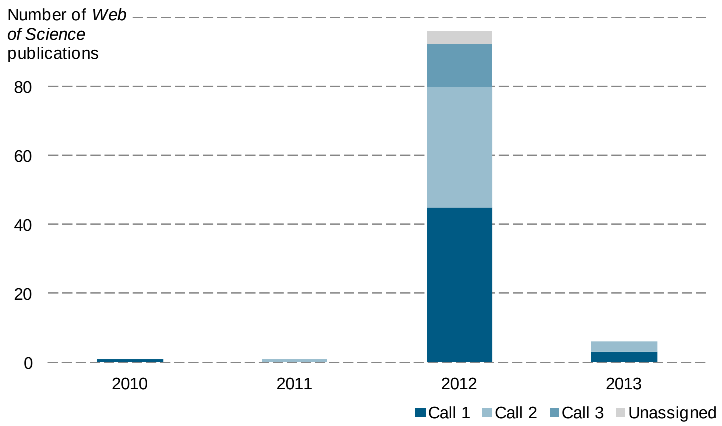
4.7.1 SUMMARY OF NEW IMI-SUPPORTED PUBLICATIONS – OUTPUT

A total of 112 new IMI-supported publications were identified for this report (as outlined in Section 4.1, 95 of which (84.8%) were abstracted in *Web of Science*.

- Most of this research was published in 2012 (Figure 4.7.1A).
- Two pre-2012 papers not found previously are new to the dataset.
    - One paper was published in 2010 and abstracted in *Web of Science* in the same year; the remaining paper was published in late 2011 and was not abstracted in *Web of Science* until 2012. Consequently, there is only one additional paper with normalised citation impact data (as outlined in Section 4.1.1).
    - The additional paper from 2010 (associated with the Call 1 project Pharma-Cog) is highly-cited and is designated as a ‘hot paper’ (Annex 2).
- The categorisation of these new publications by document type is very similar to that of the overall dataset (Section 4.2) with over 95% of new publications classified as articles and reviews (Figure 4.7.1B).

FIGURE 4.7.1 SUMMARY OF NEW IMI-SUPPORTED PUBLICATION OUTPUT

(A) Number of new *Web of Science* publications by year and call



(B) Categorisation of new *Web of Science* publications by document type

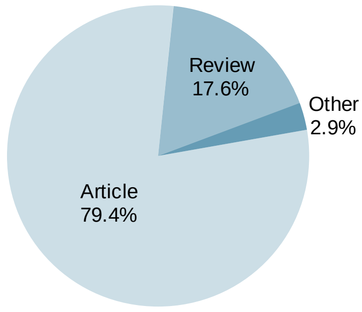


Figure 4.7.1A shows that the majority of *Web of Science* publications were associated with Calls 1 and 2 projects with some 12 publications from Call 3 projects.[16] Output from Call 3 has increased 4-fold since the initial report in October 2012. Though publication numbers are small, this represents substantial growth over a short time period and shows strong activity even in Call 3 which was launched in late 2010.

The additional publications have extended the range of journals in which IMI project research is published. Table 4.7.1 shows all new journals used more than once while Table 4.7.2 presents the top ten new journals with highest Journal Impact Factor.

TABLE 4.7.1 NEW JOURNALS IN WHICH NEW IMI PROJECT PUBLICATIONS HAVE BEEN PUBLISHED MOST FREQUENTLY (2009-2013)

Journal	**Number of *Web of Science* publications**	**Number of papers**	**Journal Impact Factor (2011)**	**Journal categories**
Journal of Neural Transmission	2	2	2.73	Clinical Neurology; Neurosciences
Proteomics	2	2	4.505	Biochemical Research Methods; Biochemistry & Molecular Biology

TABLE 4.7.2 TOP TEN NEW JOURNALS IN WHICH NEW IMI PROJECT PUBLICATIONS HAVE BEEN PUBLISHED, RANKED BY JOURNAL IMPACT FACTOR

Journal	**Number of *Web of Science* publications**	**Number of papers**	**Journal Impact Factor (2011)**	**Journal categories**
Science	1	1	31.201	Multidisciplinary Sciences
Nature Immunology	1	1	26.008	Immunology
PLOS Medicine	1	1	16.269	General & Internal Medicine

[16] Four 2012 publications have not been assigned to specific IMI projects despite review by IMI personnel.

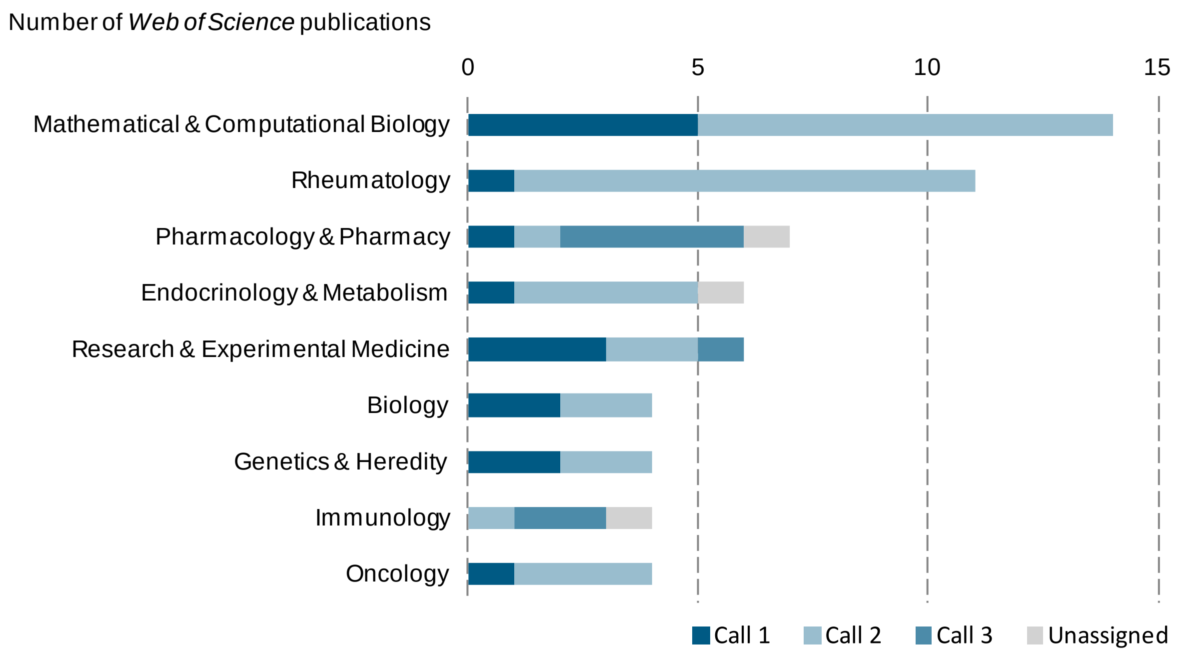
Journal	Number of *Web of Science* publications	Number of papers	Journal Impact Factor (2011)	Journal categories
Nature Reviews Clinical Oncology	1	1	11.963	Oncology
Nature Reviews Neurology	1	1	12.461	Clinical Neurology
Trends in Pharmacological Sciences	1	1	10.927	Pharmacology & Pharmacy
PLOS Genetics	1	1	8.694	Genetics & Heredity
Antioxidants & Redox Signaling	1	1	8.456	Biochemistry & Molecular Biology; Endocrinology & Metabolism
Nature Reviews Rheumatology	1	1	8.388	Rheumatology
Small	1	1	8.349	Chemistry, Multidisciplinary; Chemistry, Physical; Nanoscience & Nanotechnology; Materials Science, Multidisciplinary; Physics, Applied; Physics, Condensed Matter
Cancer Research	1	1	7.856	Oncology

These journal titles encompass a broad range of research disciplines and confirm the cross-disciplinary nature of IMI project research. Furthermore, in the last 6 months, IMI projects have produced publications in elite journals such as Science (EU-AIMS) and Nature Publishing Group titles (BTCure, Quic-Concept).

Standard bibliometric methodology uses journal category as a proxy for research field. Journals are assigned to one or more categories and every publication within that journal is subsequently assigned to that category.

Figure 4.7.2 shows the top nine *Web of Science* journal categories associated with new IMI project publications. Nine rather than ten categories are included as output below this threshold is too low for meaningful analysis.

FIGURE 4.7.2 TOP *WEB OF SCIENCE* JOURNAL CATEGORIES IN WHICH NEW IMI PROJECT PUBLICATIONS ARE PUBLISHED



Two of these categories, Mathematics & Computational Biology and Genetics & Heredity are new to this report. This is indicative of increased diversity in the IMI project research portfolio and confirms the expansion of subject coverage noted in Section 4.5.

Output in Mathematics & Computational Biology currently accounts for more than one-tenth (12.8%) of all new publications. It is associated mainly with IMI projects eTOX and Open PHACTS while output in Genetics & Heredity is associated with several different IMI projects.

4.7.2 SUMMARY OF NEW IMI-SUPPORTED PUBLICATIONS – CITATIONS

A summary of new IMI-supported publications is shown in Table 4.7.3. Although these publications are relatively recent, over one-fifth (22.3%) have already been cited.

TABLE 4.7.3 SUMMARY INDICATORS FOR NEW IMI PROJECT PUBLICATIONS

Call	Project	Number of IMI publications	Number of *Web of Science* publications	Number of cited *Web of Science* publications	Total citations (current)
n/a	(unassigned)	4	4	2	2
1	eTOX	6	5	2	2
1	EUROPAIN	8	8	2	2
1	IMIDIA	3	3	0	0
1	MARCAR	1	1	0	0
1	NEWMEDS	9	7	0	0
1	Pharma-Cog	5	3	1	123
1	PRO-active	1	1	0	0
1	PROTECT	6	6	1	2

Call	Project	Number of IMI publications	Number of *Web of Science* publications	Number of cited *Web of Science* publications	Total citations (current)
1	SAFE-T	1	1	0	0
1	SUMMIT	9	8	3	3
1	U-BIOPRED	6	6	1	2
2	BTCure	22	18	2	8
2	DDMoRe	3	2	0	0
2	Onco Track	5	3	2	2
2	Open PHACTS	12	12	5	5
2	Predect	1	1	0	0
2	Quic-Concept	2	2	1	1
2	RAPP-ID	1	1	0	0
3	BioVacSafe	5	3	1	1
3	DIRECT	1	1	0	0
3	EU-AIMS	6	6	3	16
3	MIP-DILI	2	2	1	1
	Overall	**112**	**102**	**27**	**170**

Of the 30 IMI projects in Calls 1, 2 and 3, eight had no new *Web of Science* publications (EMTRAIN, EU2P, Pharmatrain, SafeSciMet, EHR4CR, ABIRISK, EUPATI, and PreDiCT-TB) and are not listed in Table 4.7.3.

These projects had no *Web of Science* publications in the initial report.

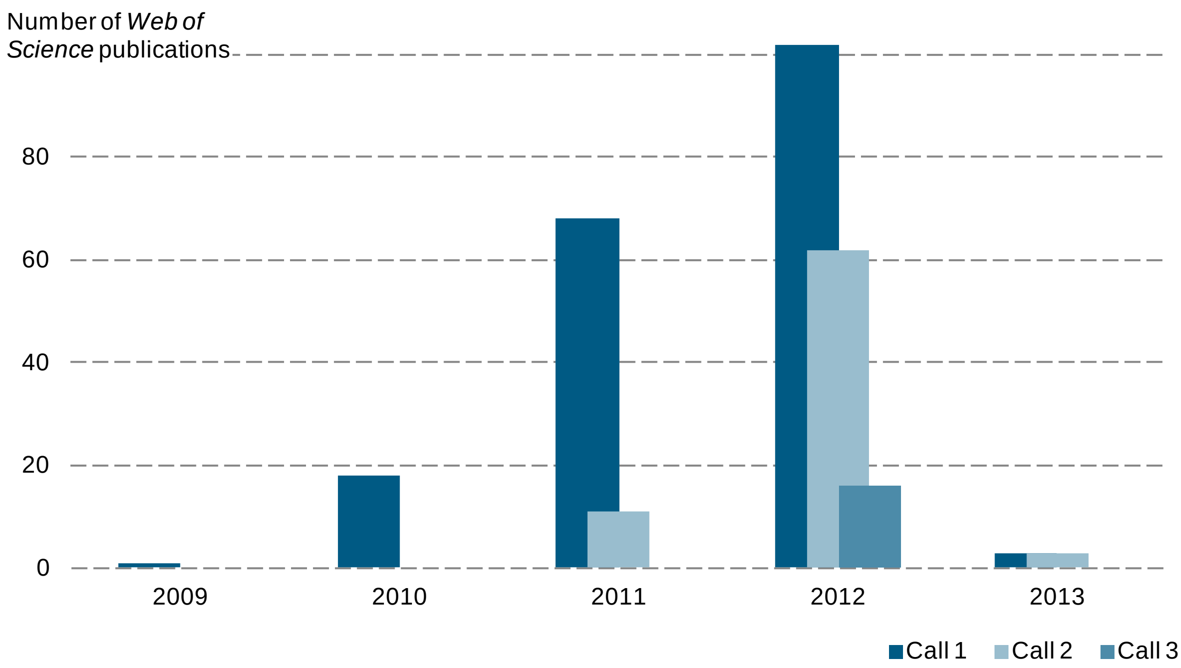
5 CITATION ANALYSIS – AT IMI PROJECT LEVEL

This Section of the report presents project level analyses of the publication output and citation impact of IMI research.

5.1 TRENDS IN PUBLICATION OUTPUT BY IMI FUNDING CALL

The data in Figure 5.1.1 show that the majority of IMI-supported publications and papers are associated with Call 1 but that output has increased from all three calls since mid-August 2012.

FIGURE 5.1.1 NUMBER OF *WEB OF SCIENCE* PUBLICATIONS BY YEAR AND FUNDING CALL



Growth in output is particularly pronounced for Call 2 where output for 2012 has more than doubled (27 publications – initial report; 62 publications – second report). Output associated with Call 3 has increased 4-fold compared to the initial report (4 publications – initial report; 16 publications – second report) though, as publication numbers are small, this trend should be interpreted with caution until more data are available.

5.2 SUMMARY BIBLIOMETRIC ANALYSES FOR IMI PROJECTS – CALL 1

Figure 5.2.1 presents a 'bubble-chart' visualisation of IMI-supported research for those projects with at least 4 papers over the time period (2010-2011). The number of papers, 2-year average citation impact and share of highly-cited papers are compared. The area of the 'bubble' is proportional to the share of highly-cited papers. The solid horizontal line indicates the average citation impact for all IMI project papers.

FIGURE 5.2.1 PAPER NUMBERS, 2-YEAR AVERAGE CITATION IMPACT AND SHARE OF HIGHLY-CITED RESEARCH FOR SELECTED IMI PROJECTS – CALL 1

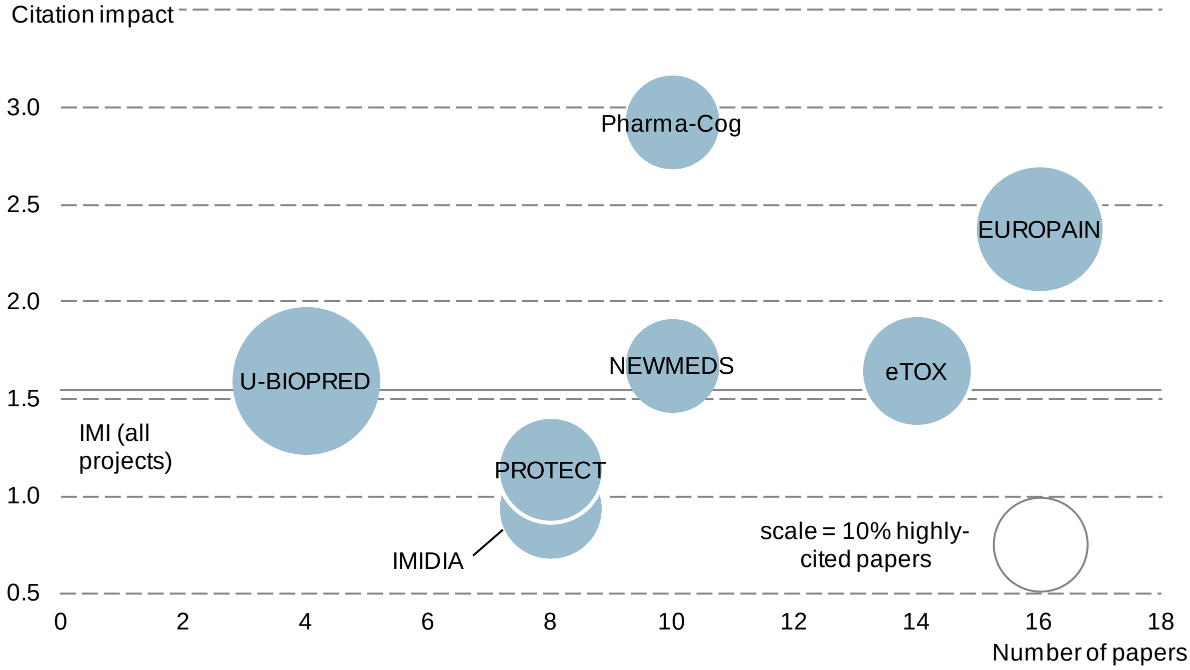


Figure 5.2.1 has been updated from the initial report to include publications from the Call 1 project Pharma-Cog. There were no highly-cited papers associated with this project in the initial report but subsequent identification of an additional paper means that the number of papers meets the threshold above.

Tables 5.2.1 and 5.2.2 compare bibliometric indicators for all projects in Call 1. Table 5.2.1 presents indicators where citation impact has been normalised against world average values and is an expansion of the data used in Figure 5.2.1. Table 5.2.2 presents a more recent (but also more descriptive) viewpoint using indicators based on current (end-January) citation counts (see Section 4.1.1).

Four Call 1 projects (EMTRAIN, EU2P, Pharmatrain, and SafeSciMET) have no *Web of Science* publications at the current time (grey text). Each of these projects has one publication associated with them but the journals in which the publications appear are not currently abstracted in *Web of Science*.

TABLE 5.2.1 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 1 – CITATIONS TO END-2011

		Citation impact			
Project	**Number of papers**	**Normalised at field level**	**Normalised at journal level**	**Average percentile**	**% Highly-cited papers[17]**
EMTRAIN	0	0.00	0.00	0.00	0.0%
eTOX	14	1.64	1.10	54.70	14.3%
Eu2P	0	0.00	0.00	0.00	0.0%
EUROPAIN	16	2.37	1.62	60.58	18.8%
IMIDIA	8	0.94	0.51	69.75	12.5%
MARCAR	1	2.41	2.78	23.15	0.0%
NEWMEDS	10	1.67	1.17	53.44	10.0%
Pharma-Cog	10	2.92	1.16	51.43	10.0%
Pharmatrain	0	0.00	0.00	0.00	0.0%
PRO-active	2	0.00	0.00	100.00	0.0%
PROTECT	8	1.13	1.10	78.59	12.5%
SafeSciMET	0	0.00	0.00	0.00	0.0%
SAFE-T	1	0.00	0.00	100.00	0.0%
SUMMIT	4	0.00	0.00	100.00	0.0%
U-BIOPRED	4	1.59	0.87	58.10	25.0%
Overall (IMI projects)	**86**	**1.55**	**0.98**	**65.70**	**11.6%**

TABLE 5.2.2 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 1 – CITATIONS TO CURRENT

	IMI publications		*Web of Science* publications			
Project	**Total**	**% Open access journals**	**Total**	**Citations**	**Raw citation impact**	**% Top quartile journals**
EMTRAIN	1	0.0%	0	0	0.00	0.0%
eTOX	30	26.7%	28	108	3.86	71.4%
Eu2P	1	0.0%	0	0	0.00	0.0%
EUROPAIN	42	7.1%	39	282	7.23	66.7%
IMIDIA	18	5.6%	16	91	5.69	100.0%
MARCAR	4	50.0%	4	9	2.25	100.0%
NEWMEDS	37	5.4%	33	159	4.82	90.9%
Pharma-Cog	19	0.0%	14	169	12.07	71.4%
Pharmatrain	1	0.0%	0	0	0.00	0.0%
PRO-active	6	83.3%	6	2	0.33	33.3%
PROTECT	21	9.5%	20	32	1.60	70.0%
SafeSciMET	1	0.0%	0	0	0.00	0.0%

[17] 'Highly-cited' refers those articles and reviews belonging to the world's top decile of papers for journal category and year of publication. A percentage that is above 10 indicates above-average performance.

	IMI publications		*Web of Science* publications			
Project	**Total**	**% Open access journals**	**Total**	**Citations**	**Raw citation impact**	**% Top quartile journals**
SAFE-T	4	25.0%	3	7	2.33	100.0%
SUMMIT	20	35.0%	16	36	2.25	87.5%
U-BIOPRED	13	0.0%	13	73	5.62	76.9%
Overall (IMI projects)	**320**	**11.3%**	**286**	**1 133**	**3.96**	**79.4%**

Bibliographic references for all highly-cited papers from IMI projects and the five papers with the highest citation velocity or interdisciplinarity (see Section 3.1.3) are listed in Annex 2. Summary Tables of bibliometric indicators Call 3 projects are listed Annex 3.

5.3 SUMMARY BIBLIOMETRIC ANALYSES FOR IMI PROJECTS – CALL 2

The numbers of papers from Call 2 projects at end-2011 are too few to allow a 'bubble-chart' visualisation of IMI-supported research at project level. This visualisation will be included in the next report (October 2013).

The number of publications from Call 2 projects has increased since the initial report and Table 5.3.2 presents summary bibliometric indicators for Call 2 projects using indicators based on current (end-January) citation counts (see Section 4.1.1). There were no additional Call 2 publications at end-2011 and Table 5.3.1 is unchanged from Table A3.2 in the initial report.

TABLE 5.3.1 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 2 – CITATIONS TO END-2011

		Citation impact			
Project	**Number of papers**	**Normalised at field level**	**Normalised at journal level**	**Average percentile**	**% Highly-cited papers[18]**
BTCure	4	0.00	0.00	100.00	0.0%
DDMoRe	0	0.00	0.00	0.00	0.0%
EHR4CR	0	0.00	0.00	0.00	0.0%
Onco Track	3	0.51	0.22	74.73	0.0%
Open PHACTS	1	0.00	0.00	100.00	0.0%
Predect	0	0.00	0.00	0.00	0.0%
Quic-Concept	0	0.00	0.00	0.00	0.0%
RAPP-ID	0	0.00	0.00	0.00	0.0%
Overall (IMI projects)	**86**	**1.55**	**0.98**	**65.70**	**11.6%**

[18] 'Highly-cited' refers those articles and reviews belonging to the world's top decile of papers for journal category and year of publication. A percentage that is above 10 indicates above-average performance.

TABLE 5.3.2 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 2 – CITATIONS TO CURRENT

	IMI publications		*Web of Science* publications			
Project	**Total**	**% Open access journals**	**Total**	**Citations**	**Raw citation impact**	**% Top quartile journals**
BTCure	36	5.6%	32	51	1.59	84.4%
DDMoRe	4	25.0%	3	2	0.67	66.7%
EHR4CR	1	0.0%	0	0	0.00	0.0%
Onco Track	14	14.3%	12	38	3.17	66.7%
Open PHACTS	19	0.0%	18	33	1.83	94.4%
Predect	1	0.0%	1	0	0.00	100.0%
Quic-Concept	8	0.0%	8	20	2.50	87.5%
RAPP-ID	2	0.0%	2	4	2.00	100.0%
Overall (IMI projects)	**320**	**11.3%**	**286**	**1 133**	**3.96**	**79.4%**

There are no *Web of Science* publications associated with EHR4CR. This project had one IMI-associated publication, but the journal in which the publication appears is not currently abstracted in *Web of Science*.

Among all Call 2 projects, only Onco Track had published cited research at end-2011 though the majority of projects have published cited research subsequently.

5.4 TRENDS IN PUBLICATION OUTPUT AND RAW CITATION IMPACT FOR IMI PROJECTS – CALL 1

Figure 5.4.1 and Figure 5.4.2 show the publication output and raw citation impact of *Web of Science* publications associated with projects in Call 1. For clarity, the projects are split into two groups in descending order of total publication volume.

FIGURE 5.4.1 TRENDS IN (A) OUTPUT AND (B) RAW CITATION IMPACT FOR RESEARCH FROM IMI-SUPPORTED PROJECTS IN CALL 1: ETOX, EUROPAIN, IMIDIA, NEWMEDS, PHARMACOG AND PROTECT

Figure 5.4.1A shows that EUROPAIN and NEWMEDS continue to account for the highest output among the group and have increased this output over the time period. Both these projects have already generated publications in *Web of Science* in 2013. As in the analyses of output trends at overall level (Section 4.3) there are jumps in activity between successive years as projects yield results for publication.

PROTECT, in particular, has increased output compared to the initial report.

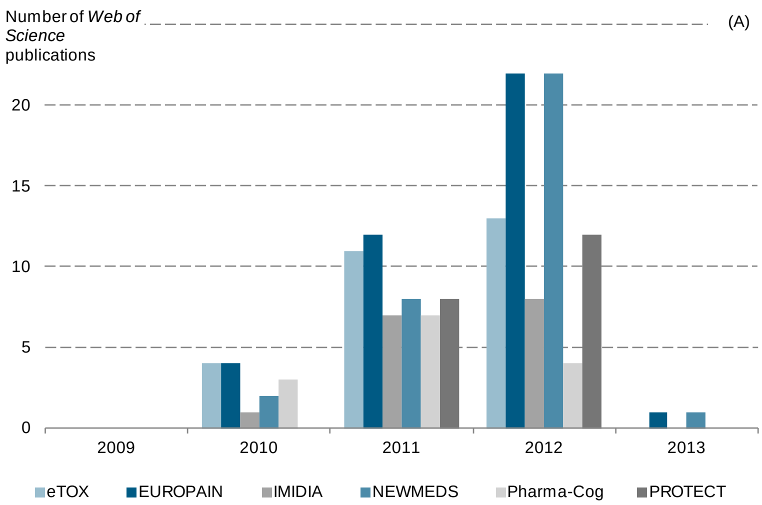


Figure 5.4.1B again shows a 'spike' in the raw citation impact of EUROPAIN publications in 2010 and this has increased from the initial report. This is again attributable to two highly-cited papers one of which is defined as a 'hot paper' (Annex 2).

Figure 5.4.1B shows an increase in the raw citation impact of Pharma-Cog publications in 2010 over the initial report. As noted in Section 4.7, this is due to the inclusion of a single highly-cited 2010 paper that has been newly identified as associated with this project and is designated as a 'hot paper' (Annex 2).

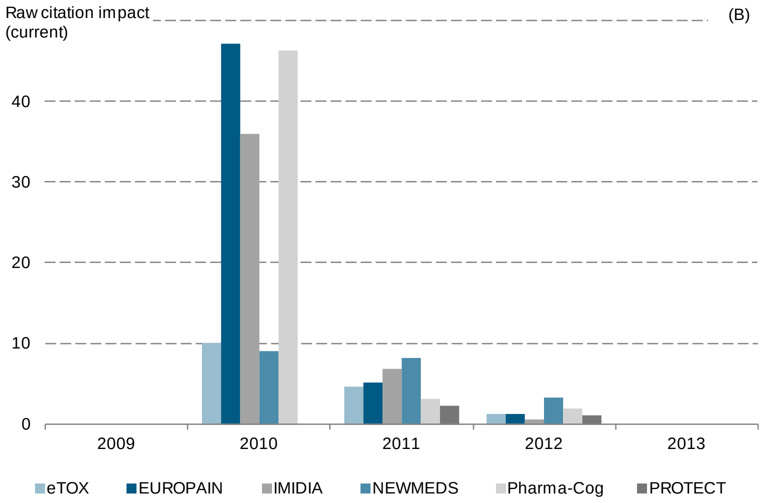


FIGURE 5.4.2 TRENDS IN (A) OUTPUT AND (B) RAW CITATION IMPACT FOR RESEARCH FROM IMI-SUPPORTED PROJECTS IN CALL 1: MARCAR, PROACTIVE, SAFE-T, SUMMIT AND U-BIOPRED.

Figure 5.4.2A shows that, of the projects in the second group, SUMMIT and U-BIOPRED continue to account for the highest share of publications. Output from both these projects has increased substantially since the initial report with SUMMIT overtaking U-BIOPRED in 2012.

Three projects in the second group (MARCAR, SUMMIT and U-BIOPRED) have already generated publications in *Web of Science* in 2013.

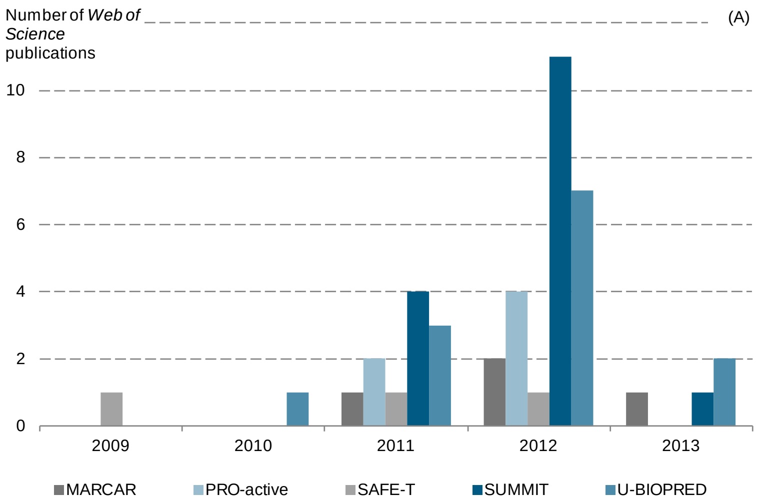
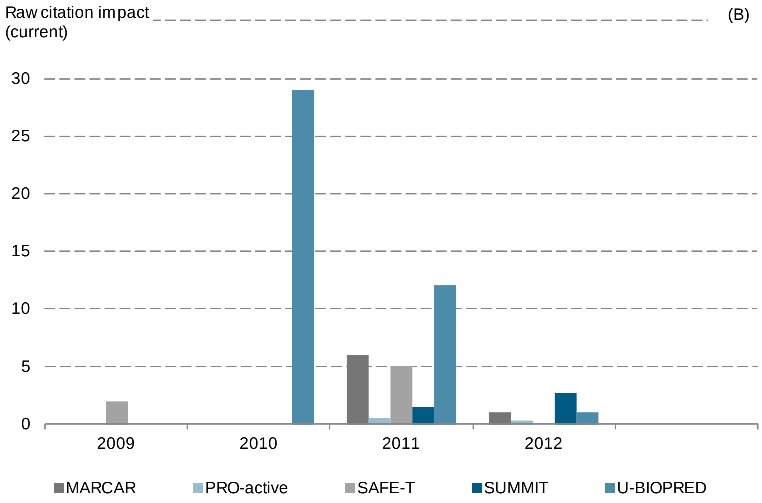


Figure 5.4.2B shows that U-BIOPRED continues to have higher raw citation impact than other projects in the group and that this now applies to research published in 2011 as well as 2010.



As there has been an increase in the number of Call 2 publications compared to the initial report, Section 5.5 presents similar trends analyses for Call 2 projects.

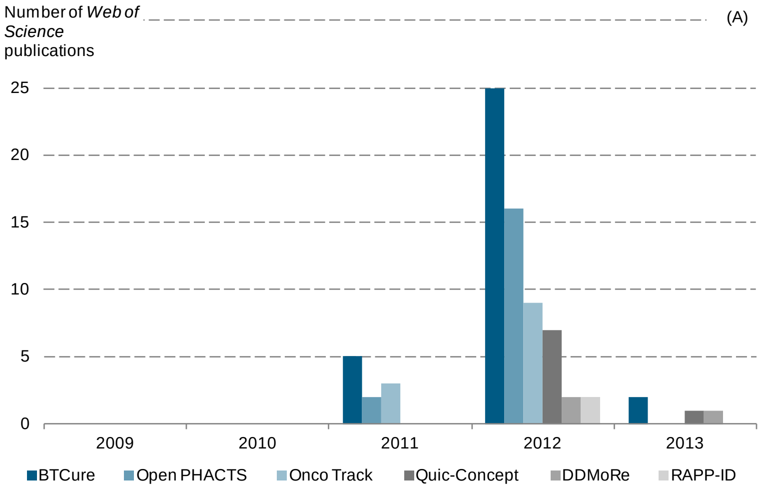
5.5 TRENDS IN PUBLICATION OUTPUT AND RAW CITATION IMPACT FOR IMI PROJECTS – CALL 2

Figure 5.5.1 shows the publication output and raw citation impact of *Web of Science* publications associated with projects in Call 2.

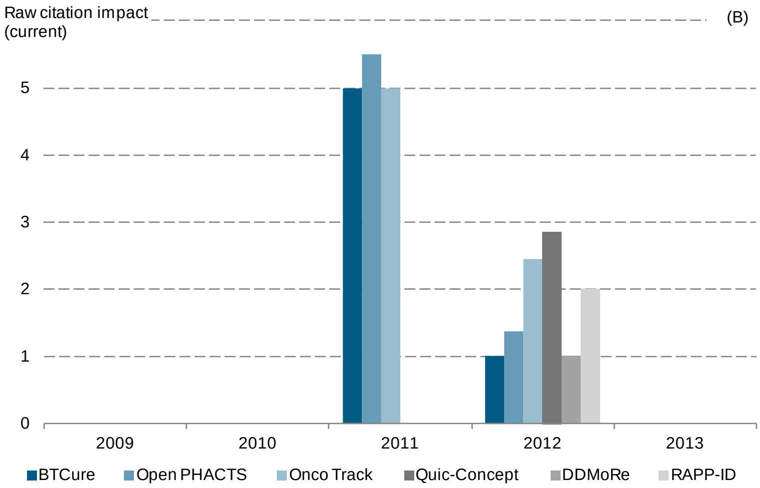
FIGURE 5.5.1 TRENDS IN (A) OUTPUT AND (B) RAW CITATION IMPACT FOR RESEARCH FROM IMI-SUPPORTED PROJECTS IN CALL 2: BTCURE, OPEN PHACTS, ONCO TRACK, QUIC-CONCEPT, DDMORE AND RAPP-ID

Figure 5.5.1A shows that BTCure is the most prolific of the Call 2 projects with output in line with the top Call 1 projects. Open PHACTS is the next most prolific Call 2 project with other projects some way behind this.

Three projects in Call 2 (BTCure, DDMoRe and Quic-Concept) have already generated publications in *Web of Science* in 2013.



Early indications are that Call 2 research is being well-cited, particularly that associated with Quic-Concept none of whose 2012 publications are uncited (from Figure 5.5.1B).



EHR4CR and Predect are not included in the above analyses. EHR4CR has not yet generated publications while the single publication for Predect was published in late 2012 and is as yet uncited.

6 PATENT ANALYSIS – INNOVATION AND TECHNOLOGY TRANSFER

Four patent applications and one patent award have been made since the first IMI report in October 2012 and are listed below.

TABLE 6.1 NEW PATENT APPLICATIONS AND AWARDS FROM IMI-SUPPORTED PROJECTS

IMI project	Patent title	Patent body	Patent number
IMIDIA	Human pancreatic beta cell lines for diagnostic of diabetes	US Patent and Trademark Office	20110318389
RAPP-ID	A device and protocol related to breath-born aerosol sampling	data not yet available	
SUMMIT	A new ultrasound-based method for non-invasive assessment of atherosclerotic plaque	data not yet available	
	A rat model of diabetic complications	data not yet available	
	Desmosine assay as biomarker of extracellular matrix degradation and vascular disease	data not yet available	

Patent titles have been provided by IMI personnel. Thomson Reuters patent data and analyses for IMI research will be updated in Report 3 (October 2013).

7 BIBLIOMETRIC INDICATORS FOR IMI RESEARCHERS: PRODUCTIVITY, RESEARCH PERFORMANCE AND COLLABORATION

This Section of the report presents analyses of the publication output and citation impact of IMI researcher publications as well as collaborative activities between IMI researchers

7.1 PUBLICATIONS BY IMI-SUPPORTED RESEARCHERS

Publications by IMI-supported researchers were identified using researcher names, projects, and affiliations supplied by IMI. For this initial report, data and analyses are limited to those 3 477 researchers associated with projects funded by the first three IMI funding calls (Call 1,2 and 3) from 2008 to 2010.

Names of researchers associated with funded projects were provided by IMI personnel along with organisational affiliation. Combining these two data elements with the assumption that researchers from the same project are likely to co-author with one another, candidate publications authored by these individuals were identified using an automated process in *Web of Science* for the period January 2007 to January 2013. These matches were further reviewed and edited by IMI personnel.

It is important to note that this dataset includes all identified output from IMI-supported researchers as described above, and is not restricted to that output specifically resulting from IMI funding. With the assumption that the quality of the researcher does not change depending on the source of their funding, these analyses illustrate the quality of researchers who are supported by IMI funds.

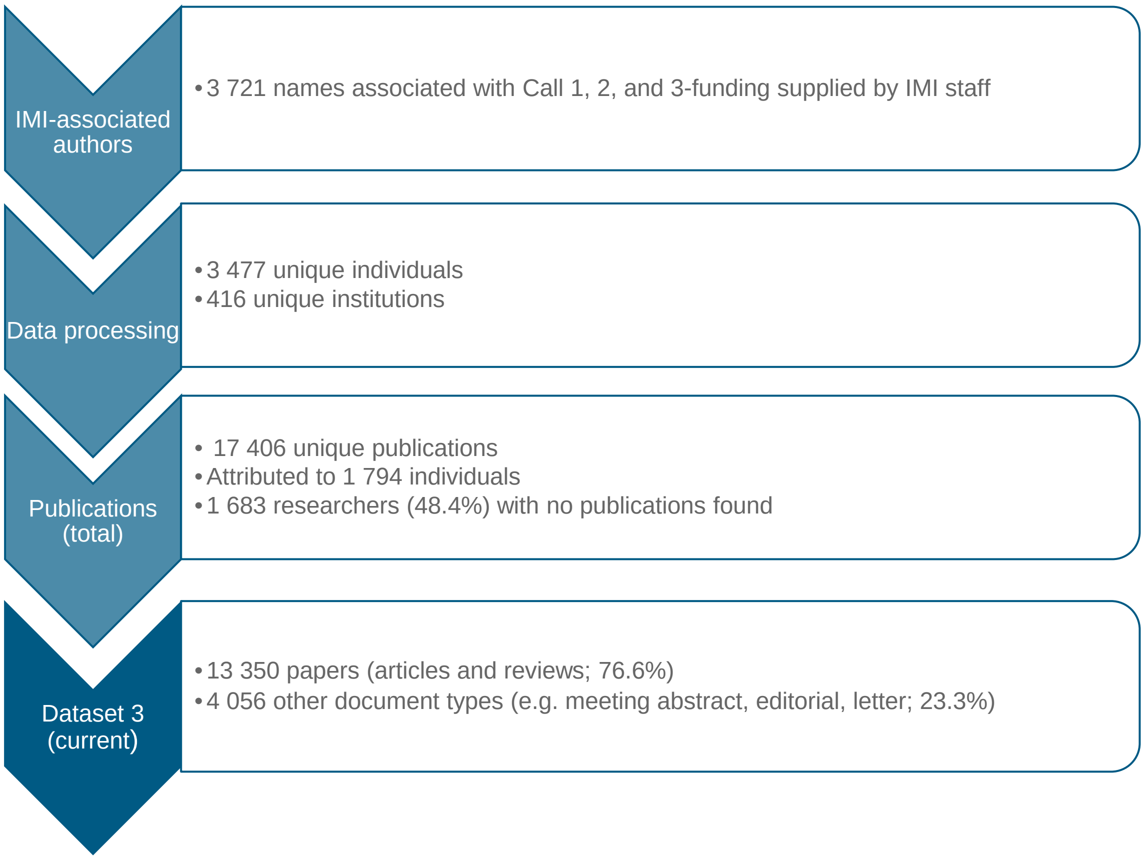
These data will also provide a basis for benchmarking how well research from IMI-supported projects compares with research by researchers that IMI funds in comparisons in Sections 4 and 5.

7.2 CITATION DATA FOR PUBLICATIONS BY IMI-SUPPORTED RESEARCHERS

A total of 17 406 publications by IMI-supported researchers were identified. The process of identifying publications by IMI-supported researchers with Thomson Reuters citation data is outlined in Figure 7.2.1.

Citation counts for these 17 406 publications have been sourced from the citation databases which underlie Thomson Reuters *Web of Science* and were extracted at end-January 2013.

FIGURE 7.2.1 IDENTIFYING PUBLICATIONS BY IMI-SUPPORTED RESEARCHERS WITH THOMSON REUTERS CITATION DATA



7.3 BIBLIOMETRIC INDICATORS FOR IMI-SUPPORTED RESEARCHERS: PRODUCTIVITY

Publication output, as previously, is higher for IMI-supported researchers based in academic institutions and other research environments (Table 7.3.1).

Overall, 51.6% of researchers have at least one publication.[19]

TABLE 7.3.1 PRODUCTIVITY: PUBLICATION OUTPUT, OVERALL AND BY SECTOR

	Number of researchers		% researchers with publications
Sector	**Total**	**With publications**	
Academic	1 345	797	59.3%
Corporate	1 133	545	48.1%
Patient Organisation	50	13	26.0%
Regulatory Agency	68	24	35.3%
Research (Other)	609	346	56.8%
Small Medium Enterprise	224	69	30.8%

[19] This proportion, however, is statistically significantly different by sector ($p<0.0001$, Chi-Sq =105.3532 test of the equality of proportions, 6 degrees of freedom).

Sector	Number of researchers		% researchers with publications
	Total	With publications	
No assignment	48	19	39.6%
Total researchers	3 477	**1 813**	**51.6%**

7.4 BIBLIOMETRIC INDICATORS FOR IMI-SUPPORTED RESEARCHERS: RESEARCH PERFORMANCE

The bibliometric indicators presented in Table 7.4.1 have been calculated for each individual IMI-supported researcher and aggregated by sector.

Since the initial report, researchers associated with patient organisations have published well-regarded research. Three (23.1%) have published at least one 'hot' paper (defined in Section 3.1.3), 4 (30.8%) have an h-index of at least 10 and all 13 have published exclusively in top-quartile journals.

Of the 797 publishing academic-based researchers, 142 researchers (17.8%) have published at least one 'hot paper', 134 (16.8%) have an h-index of at least 10 and just over half have published most frequently in top quartile journals.

Similarly, researchers based in other research environments have published research which has performed well. Sixty-four of these researchers (18.5%) have published a minimum of one 'hot paper', 42 researchers (12.1%) have h-index of at least 10 and around half of these researchers have published in top quartile journals more frequently than in less well-regarded journals.

By contrast, many IMI-supported researchers working in companies also have published relatively frequently in top quartile journals but these publications appear to be less well-cited as their 'hot papers' indicator and h-indices are generally lower.

TABLE 7.4.1 RESEARCH PERFORMANCE: BIBLIOMETRIC INDICATORS, OVERALL AND BY SECTOR

Sector	Researchers		With 'hot papers'		h-index ≥ 10		Publishes most often in top quartile journals	
	Total	Publishing	N	%	N	%	N	%
Academic	1 345	797	142	17.8%	134	16.8%	674	84.6%
Corporate	1 133	545	40	7.3%	11	2.0%	398	73.0%
Patient Organisation	50	13	3	23.1%	4	30.8%	13	100.0%
Regulatory Agency	68	24	2	8.3%	0	0.0%	15	62.5%
Research (other)	609	346	64	18.5%	42	12.1%	302	87.3%
Small Medium Enterprise	224	69	12	17.4%	5	7.2%	61	88.4%
No assignment	48	19	3	15.8%	0	0.0%	19	100.0%
Total researchers	3 477	1 813	**266**	**14.7%**	**196**	**10.8%**	**1 482**	**81.7%**

TABLE 7.4.2 RESEARCH PERFORMANCE: BIBLIOMETRIC INDICATORS, CONTINUOUS OUTCOMES

*For each metric (diffusion index and citation velocity) the mean per researcher and the maximum per researcher were calculated and those averaged withi sectors to obtain the summary metrics below. *Note: Data for researchers associated with patient organisations is not available as there was only 1 researcher with 1 publication which has not been cited*

Sector	Researchers		Mean Diffusion Index		Maximum Diffusion Index		Mean Citation Velocity		Maximum Citation Velocity	
	Total	Publishing	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Academic	48	797	0.538	0.134	0.674	0.144	0.011	0.019	0.051	0.105
Corporate	1 345	545	0.533	0.145	0.621	0.140	0.009	0.017	0.020	0.042
Patient Organisation*	1 133	13	0.495	0.108	0.602	0.068	0.030	0.047	0.075	0.085
Regulatory Agency	50	24	0.528	0.116	0.599	0.120	0.055	0.216	0.064	0.215
Research (other)	68	346	0.525	0.150	0.645	0.161	0.015	0.058	0.058	0.129
Small Medium Enterprise	609	69	0.572	0.148	0.634	0.134	0.017	0.029	0.060	0.146
No Assignment	224	19	0.520	0.161	0.640	0.132	0.014	0.016	0.026	0.031
Total researchers	**3 477**	**1 813**	**0.535**	**0.141**	**0.651**	**0.147**	**0.012**	**0.039**	**0.044**	**0.101**

7.5 COLLABORATION BETWEEN IMI-SUPPORTED RESEARCHERS AT INDIVIDUAL LEVEL

The projects funded by IMI are collaborative in nature. However, collaboration between researchers can manifest in many different ways – only one of which is in co-authorship in published materials. Using this definition of collaboration, social network analysis was used to assess the extent to which collaboration occurs, the nature of collaborations between researchers, and identify opportunities to foster collaboration.

Overall, 1 813 researchers (52.1% of 3 477 IMI researchers in total) published any documents that were indexed in *Web of Science*. About three quarters of these researchers (N=1 326, 73.1% of 1 813) collaborated (co-authored) with at least one other IMI researcher during the period January 2007-January 2013.

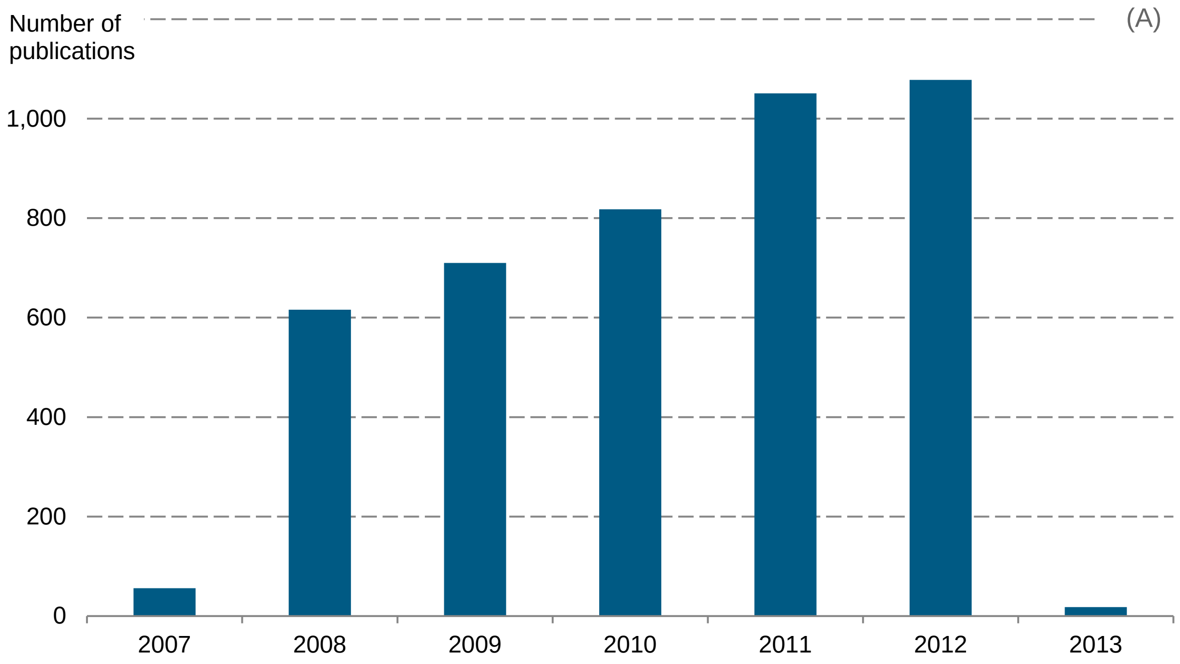
The frequency of collaborative activities are shown over the period of January 2007 to December 2012 by year in Figure 7.5.1 and Table 7.5.1 and further illustrated among researchers in Figure 7.5.2.

TABLE 7.5.1 BREAKDOWN OF COLLABORATIVE ACTIVITY BY YEAR

	2007	2008	2009	2010	2011	2012	2013
Publications	56	616	710	818	1 051	1 078	18
Within-Sector Collaborations	58	461	632	937	1 127	1 103	24
Cross-Sector Collaborations	13	152	384	541	663	555	1
% Cross-Sector	18.3%	24.8%	37.8%	36.6%	37.0%	33.5%	4.0%

FIGURE 7.5.1 DISTRIBUTION OF COLLABORATIVE ACTIVITIES BY YEAR

(A) Number of publications co-authored by two or more IMI researchers by year. (B) Collaborations defined as distinct researcher dyads within and across sectors appearing on one or more publications during the given publication year.



The number of individual researchers with collaborative activity has increased over time from 96 in 2007, to 493 in 2008, 618 in 2009, 666 in 2010, 793 in 2011, 807 in 2012, and 35 in 2013 to January.

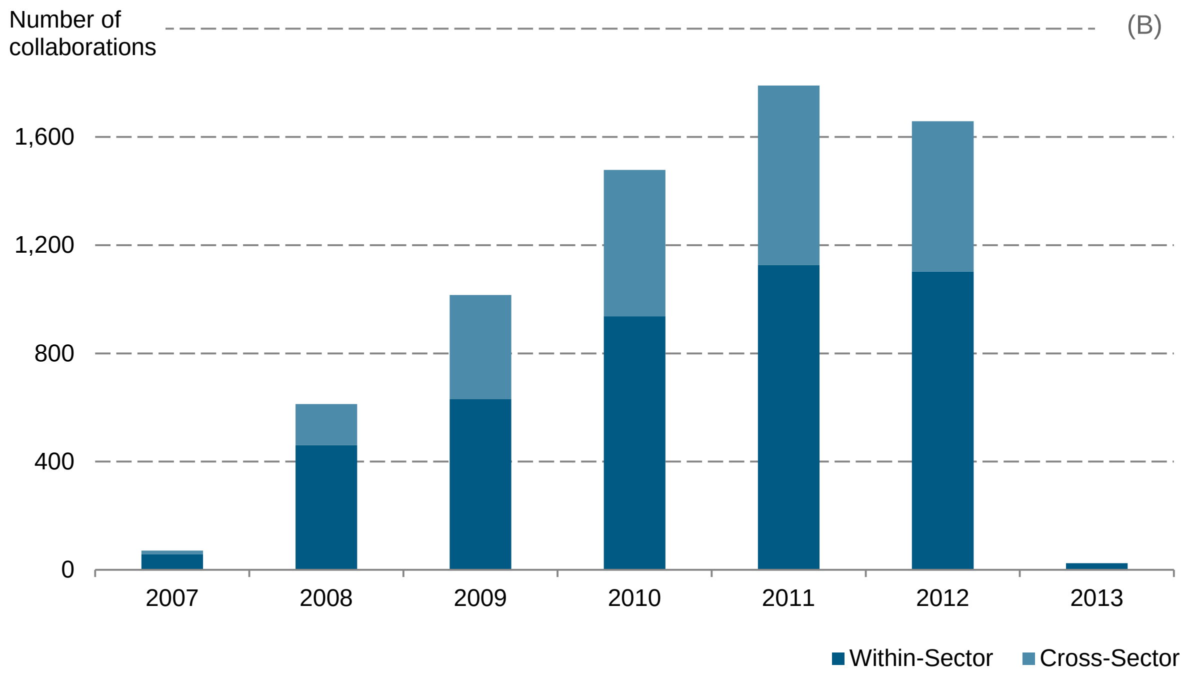
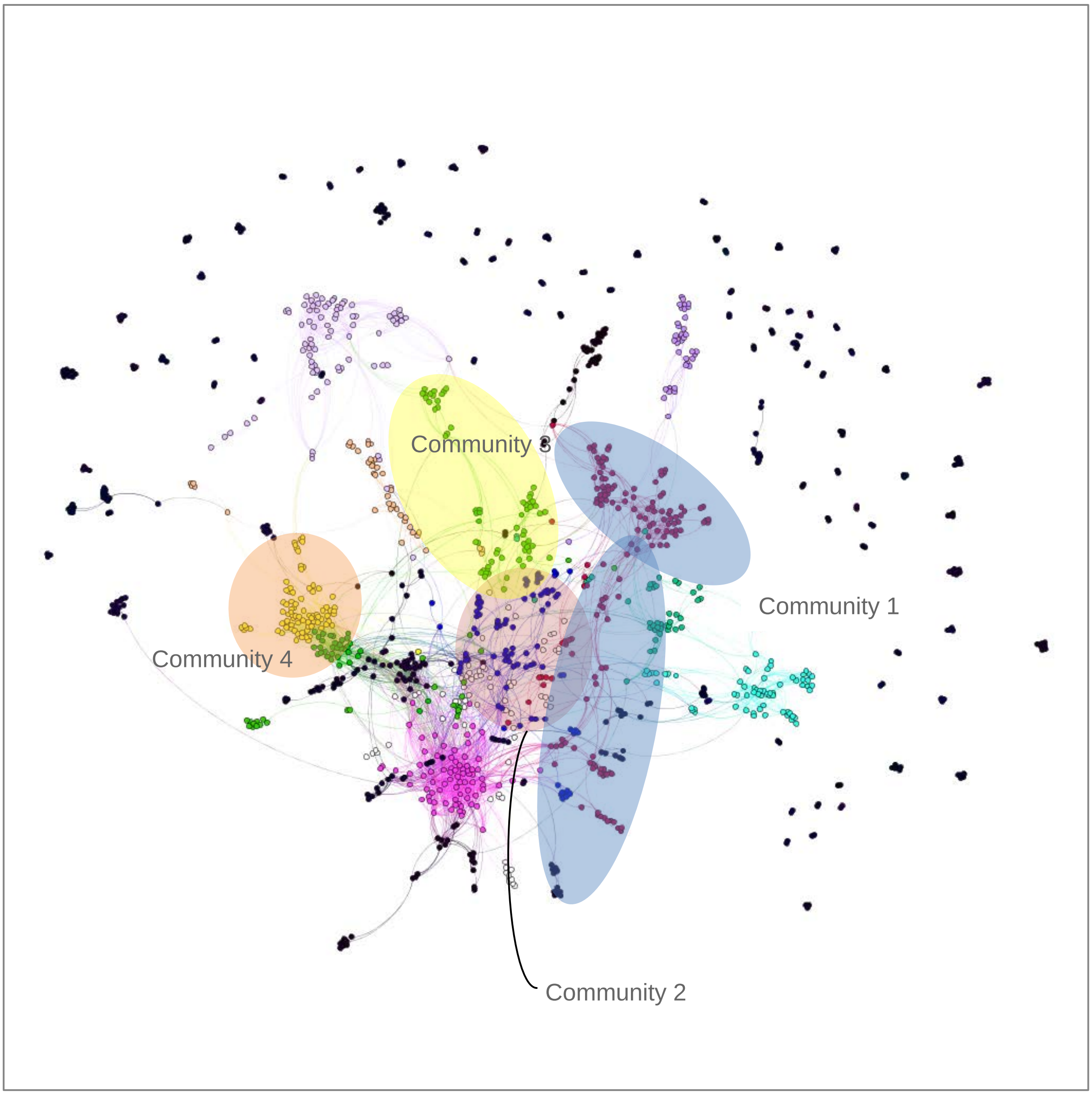


FIGURE 7.5.2 MAP OF 660 IMI PROJECT RESEARCHERS WHO HAVE CO-AUTHORED WITH AT LEAST ONE OTHER RESEARCHER WITHIN THE NETWORK BASED ON CO-AUTHORSHIP ACTIVITIES FROM JANUARY 2007 - AUGUST 2012.

Each individual is represented as a single node coloured with respect to the sector of their organisation. Ties between researchers are instances where co-authorship has occurred in a published work. The largest group of inter-connected researchers is composed of 24 communities of which the 4 largest are shown in shaded ovals. Graphics produced using Gephi, applying the Force Atlas 2 layout.[20] *Communities identified using a resolution of 1.*[21] *Isolated communities not connected to the main group of researchers are shown in black.*



[20] Jacomy, M. (2009). Force-Atlas Graph Layout Algorithm. *URL: http://gephi.org/2011/forceatlas2-the-new-version-of-our-home-brew-layout/*

[21] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

7.6 COLLABORATION BETWEEN IMI-SUPPORTED RESEARCHERS AT SECTOR LEVEL

TABLE 7.6.1 DISTRIBUTION OF SECTORS WITHIN SELECT COMMUNITIES BASED ON CO-AUTHORSHIP ACTIVITIES FROM JANUARY 2007 – AUGUST 2012.

Thirty-five isolated communities exist composed of between 2 and 10 researchers each. The largest group of inter-connected researchers (N=1 075 researchers) is composed of 24 distinct communities the largest four of which are described at below.

	Isolated Communities		Connected Communities		Connected Community							
					1		2		3		4	
Sector	N	%	N	%	N	%	N	%	N	%	N	%
No assignment	4	1.6%	9	0.8%	1	0.7%	--		--		--	
Academic	95	37.8%	548	51.0%	65	43.0%	48	53.9%	31	36.9%	52	63.4%
Corporate	88	35.1%	223	20.7%	39	25.8%	11	12.4%	24	28.6%	6	7.3%
Patient Organisation	--	--	8	0.7%	1	0.7%	1	1.1%	--	--	--	--
Regulatory Agency	2	0.8%	11	1.0%	1	0.7%	--	--	8	9.5%	--	--
Research (other)	53	21.1%	237	22.0%	39	25.8%	20	22.5%	21	25.0%	24	29.3%
Small Medium Enterprise	9	3.6%	39	3.6%	5	3.3%	9	10.1%	--	--	--	--
Total	**251**		**1075**		**151**		**89**		**84**		**82**	

The largest component, shown at the centre of Figure 7.5.2 and defined as groups of researchers where all individuals are connected with one another directly or indirectly via other IMI researchers, consisted of 1 051 researchers representing all six sectors (Table 7.6.1). Within this set of researchers, twenty-four communities were identified within which there are more frequent and closely inter-related co-authorship activities. The largest four of these twenty-four communities are shown enclosed by coloured ovals. A complete depiction of all twenty-four communities can be found in Figure 7.6.1D.

The largest community, shown in two blue ovals in Figure 7.5.2, is composed of individuals who are all closely positioned suggesting that there is high collaborative activity between researchers, as well as high collaborative activity between co-authors of a given researcher's co-authors. This group is largely composed of academic researchers (43.0%, Table 7.6.1), although it has a large contribution from each of the other five sectors, perhaps with a more even distribution among these sectors than among IMI funded researchers overall.

The co-authorship activities in group 2 (shown in red in Figure 7.5.2) are central to the larger group of all connected investigators in the main component. Composition of this group reflects a distribution of researchers more skewed towards academics than in community 1 (Table 7.6.1). However, the collaborative nature of the IMI projects does still promote cross-sector co-authorship.

While the majority of publishing researchers are connected to one another and are in the main connected component, 16.1% of collaborating researchers (N=251 of 1 326) collaborate within isolated communities composed of between 2 and 10 researchers. Eighty-three isolated groups exist (shown on the periphery of Figure 7.5.2), of which 59 (71.1%) are composed of researchers from only one sector.

This main component includes researchers from 201 distinct organisations, 34.8% (N=70) of which span across communities (Table 7.6.2). Within this set there are 127 academic organisations, 23 corporate organisations, 3 patient organisations, 5 regulatory agencies, 61 research/other entities, and 22 small medium enterprise organisations. The two entities which span the most communities are Astra Zeneca (corporate) and Karolinska Institutet (academic). This is a change from the initial report which identified Astra Zeneca and Imperial College London as spanning the most communities. Overall, these organisational affiliations include only 7.1% (N=75 of 1 051) of researchers in the main component.

Co-authorship is more common among researchers in the same sector than among researchers in different sectors (Figure 7.6.1B). This is expected given the principle of homophily which suggests that individuals are more likely to interact with individuals who are like them.[22] However, there are substantial co-authorship activities among researchers from different sectors (Figure 7.6.1B). Of a total 4 242 distinct co-authorship relationships, 1 693 are cross-sector and involve 717 total researchers from all 6 sectors. This accounts for 39.9% of all co-authorship activities during the analysis period.

The same is true of co-authorship activities by project. The majority of collaborative relationships are among researchers associated with the same project with only 1 080 of 4 242 of co-authorship relationships (25.5%) being cross-project.

7.6.1 STRENGTHS AND LIMITATIONS

These data rely on publication matching from researcher productivity analysis and are restricted to the period January 2007 – January 2013. Although this includes all document types some publications may have been missed in the effort to only match researchers to publications for which we are fairly certain they are the author.

Researchers with multiple organisational affiliations and/or multiple sectors were assigned to a single organisational affiliation, and sector based on the judgment of IMI personnel. No similar assignment was made for investigators affiliated with multiple projects – that is, multiple project affiliations were preserved for all investigators who were involved with multiple projects.

[22] McPherson et al. "Birds of a Feather: Homophily in Social Networks". Annu Rev Sociol, 2001, 27: 415-44.

FIGURE 7.6.1 THE LARGEST GROUP OF INTER-CONNECTED IMI RESEARCHERS IS COMPOSED OF 554 RESEARCHERS FROM ALL 15 PROJECTS AND ALL SIX SECTORS.

(A) Researchers with any collaborative activity are shown coloured by sector. (B) Researchers are coloured by community. Isolated communities outside of the main connected component are independent of one another and all shown in black. (C) Researchers are shown coloured by disease area. (D) Researchers in the main component are shown coloured by community.

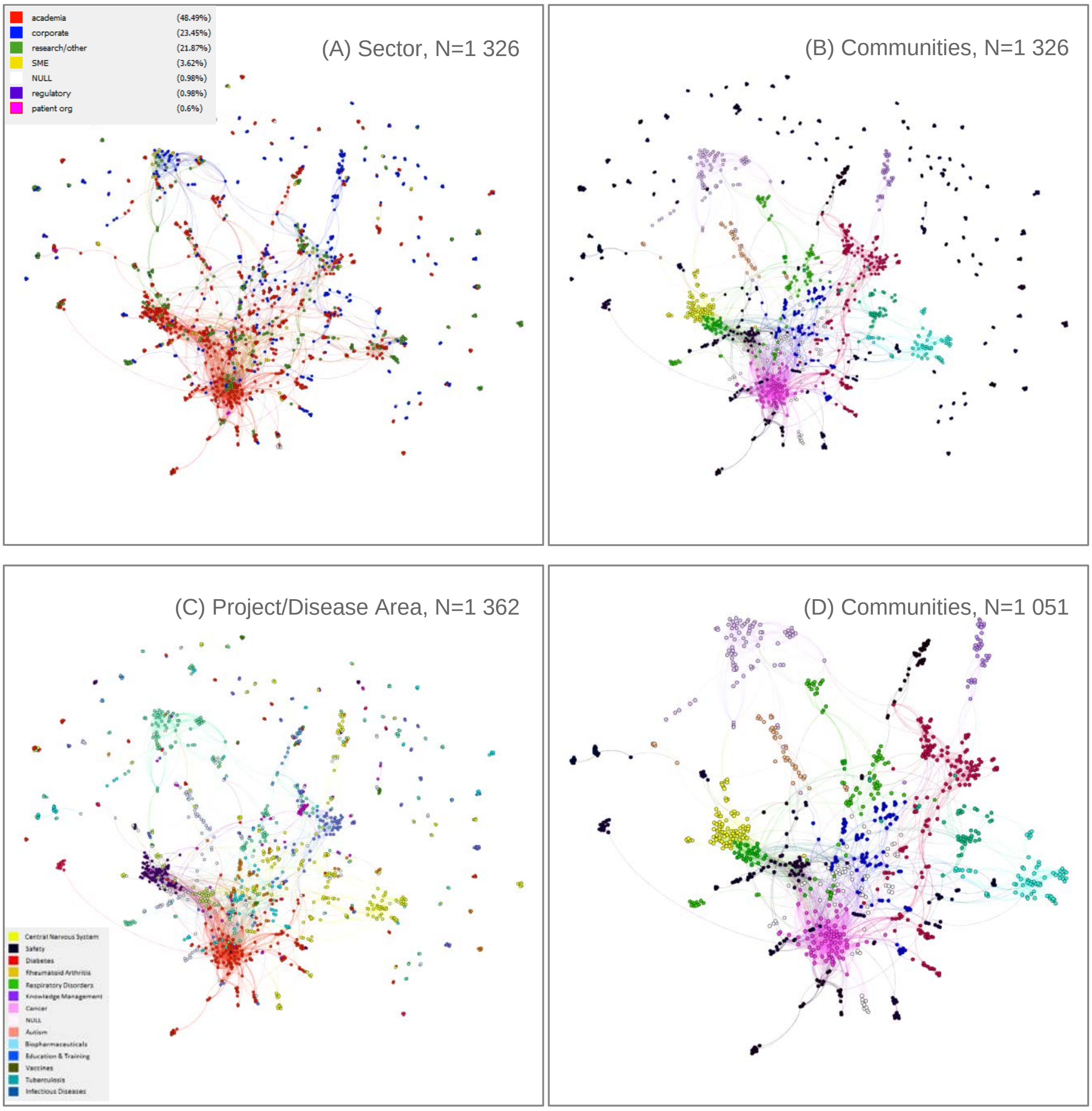


TABLE 7.6.2 ORGANISATIONS AND THE NUMBER OF ASSOCIATED COMMUNITIES AND RESEARCHERS WITHIN THE MAIN INTER-CONNECTED COMPONENT

In all, 275 organisations with collaborative co-authorship activity were identified, of which 104 (37.8%) span two or more communities and 47 (17.1%) span three or more communities. The top 25 organisations by number of communities which they span, is shown below.

Organisation	Sector	Number of communities	Number of researchers
Astra Zeneca	Corporate	20	46
Roche	Corporate	15	38
Novartis	Corporate	13	28
Karolinska Institutet	Academia	11	47
Eli Lilly	Corporate	10	23
GSK	Corporate	10	14
Pfizer	Corporate	10	33
INSERM	Research (other)	8	17
Lundbeck	Corporate	7	13
Servier	Corporate	7	11
King's College London	Academia	6	21
Sanofi-Aventis	Corporate	6	14
University of Dundee	Academia	6	12
Bayer	Corporate	5	7
Boehringer Ingelheim	Corporate	5	9
Imperial College London	Academia	5	13
Janssen	Corporate	5	26
Max-Planck-Gesellschaft	Research (other)	5	17
UCB	Corporate	5	6
University of Lille	Academia	5	9
University of Manchester	Academia	5	13
Amgen	Corporate	4	5
Centre National de la Recherche Scientifique	Research (other)	4	14
Universiteit Utrecht	Academia	4	10
University College London	Academia	4	10

ANNEX 1: DEFINITION AND SCOPE OF *WEB OF SCIENCE* JOURNAL CATEGORIES

The **Biology** category includes journals that have a broad or interdisciplinary approach to biology. In addition, it includes materials that cover a specific area of biology not covered in other categories such as theoretical biology, mathematical biology, thermal biology, cryobiology, and biological rhythm research.

Clinical Neurology covers journals on all areas of clinical research and medical practice in neurology. The focus is on traditional neurological illnesses and diseases such as dementia, stroke, epilepsy, headache, multiple sclerosis, and movement disorders that have clinical and socio-economic importance. This category also includes journals on medical specialties such as paediatric neurology, neurosurgery, neuroradiology, pain management, and neuropsychiatry that affect neurological diagnosis and treatment.

Endocrinology & Metabolism includes journals focused on endocrine glands; the regulation of cell, organ, and system function by the action of secreted hormones; the generation and chemical/biological properties of these substances; and the pathogenesis and treatment of disorders associated with either source or target organs. Specific areas covered include neuroendocrinology, reproductive endocrinology, pancreatic hormones and diabetes, regulation of bone formation and loss, and control of growth.

Genetics & Heredity 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.

Immunology covers journals 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.

Mathematical & Computational Biology includes journals concerning the use of mathematical, statistical and computational methods to address data analysis, modelling, and information management in biological problems, processes and systems. Among the areas covered are biostatistics, bioinformatics, biometrics, modelling of biological systems, and computational biology.

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.

Oncology covers journals on the mechanisms, causes, and treatments of cancer including environmental and genetic risk factors, and cellular and molecular carcinogenesis. Aspects of clinical oncology covered include surgical, radiological, chemical, and palliative care. This category is also concerned with journals on cancers of specific systems and organs.

Pharmacology & Pharmacy contains journals on the discovery and testing of bioactive substances, including animal research, clinical experience, delivery systems, and dispensing of drugs. This category also includes journals on the biochemistry, metabolism, and toxic or adverse effects of drugs.

Psychiatry covers journals that focus on the origins, diagnosis, and treatment of mental, emotional, or behavioural disorders. Areas covered in this category include adolescent and child psychiatry, forensic psychiatry, geriatric psychiatry, hypnosis, psychiatric nursing, psychiatric rehabilitation, psychosomatic research, and stress medicine.

Research & Experimental Medicine includes journals describing general medical research with a particular emphasis on extremely novel techniques and clinical interventions in a broad range of medical specialisations and applications, including vaccine development, tissue replacement, immunotherapies, and other experimental therapeutic strategies. Journals in this category reflect clinical interventions that are in early stages of development, using in vitro or animal models, and small-scale clinical trials.

Rheumatology covers journals on clinical, therapeutic, and laboratory research about arthritis and rheumatism, the chronic degenerative autoimmune inflammatory diseases that primarily affect joints and connective tissue.

ANNEX 2: BIBLIOGRAPHY OF HIGHLY-CITED PAPERS, 'HOT PAPERS' AND THOSE PAPERS WITH HIGHEST INTERDISCIPLINARITY

For the purpose of this report, highly-cited papers have been defined as those articles and reviews which belong to the world's top decile of papers in that journal category and year of publication, when ranked by number of citations received. A percentage that is above 10 indicates above-average performance.

Section A2.1 lists the ten papers in the IMI project publications dataset that have been identified as highly-cited. Papers are listed in ascending alphabetical order (project, first author) and those papers also identified as 'hot papers' (Section A2.2) are highlighted (bold text).

A2.1 HIGHLY-CITED PAPERS ASSOCIATED WITH IMI PROJECTS

- eTOX: Obiol-Pardo, C *et al*. (2011) A Multiscale Simulation System for the Prediction of Drug-Induced Cardiotoxicity, Journal Of Chemical Information And Modeling, 51: 483-492, doi: 10.1021/ci100423z
- eTOX: Taboureau, O *et al*. (2011) ChemProt: a disease chemical biology database, Nucleic Acids Research, 39: D367-D372, doi: 10.1093/nar/gkq906
- EUROPAIN: Aasvang, EK *et al*. (2010) Predictive Risk Factors for Persistent Postherniotomy Pain, Anesthesiology, 112: 957-969, doi: 10.1097/ALN.0b013e3181d31ff8
- **EUROPAIN: Finnerup, NB *et al*. (2010) The evidence for pharmacological treatment of neuropathic pain, Pain, 150: 573-581, doi: 10.1016/j.pain.2010.06.019**
- EUROPAIN: Phillips, TJC *et al*. (2010) Pharmacological Treatment of Painful HIV-Associated Sensory Neuropathy: A Systematic Review and Meta-Analysis of Randomised Controlled Trials, Plos One, 5: doi: 10.1371/journal.pone.0014433
- IMIDIA: Roggli, E *et al*. (2010) Involvement of MicroRNAs in the Cytotoxic Effects Exerted by Proinflammatory Cytokines on Pancreatic beta-Cells, Diabetes, 59: 978-986, doi: 10.2337/db09-0881
- NEWMEDS: Ingason, A *et al*. (2011) Maternally Derived Microduplications at 15q11-q13: Implication of Imprinted Genes in Psychotic Illness, American Journal Of Psychiatry, 168: 408-417, doi: 10.1176/appi.ajp.2010.09111660
- Pharma-Cog: Frisoni, GB *et al*. (2010) The clinical use of structural MRI in Alzheimer disease, Nature Reviews Neurology, 6: 67-77, doi: 10.1176/appi.ajp.2010.09111661
- PROTECT: Eussen, SRBM *et al*. (2010) Effects of the use of phytosterol/-stanol-enriched margarines on adherence to statin therapy, Pharmacoepidemiology And Drug Safety, 19: 1225-1232, doi: 10.1002/pds.2042
- U-BIOPRED: Auffray, C *et al*. (2010) An Integrative Systems Biology Approach to Understanding Pulmonary Diseases, Chest, 137: 1410-1416, doi: 10.1378/chest.09-1850

'Hot papers' have been defined as papers which are cited quickly to their research field (Section 3.1.3).

Section A2.2 lists the five papers from IMI projects that have been identified as 'hot papers'. Papers are listed in ascending alphabetical order (project, first author) and those papers also identified as highly cited (Section A2.2) are highlighted (bold text).

Papers from the eTOX, Open PHACTS and Pharma-Cog projects are new to this list while papers associated with EUROPAIN and NEWMEDS were identified as 'hot papers' in the previous report to IMI.

With the exception of the EUROPAIN and Pharma-Cog papers, none of these publications had accumulated citations at end-2011 when the percentile ranking used to define highly-cited papers is calculated.

A2.2 HOT PAPERS' ASSOCIATED WITH IMI PROJECTS

- eTOX: Arighi, CN *et al*. (2011) Overview of the BioCreative III Workshop, BMCBioinformatics, 12: , doi: 10.1186/1471-2105-12-S8-S1
- **EUROPAIN: Finnerup, NB *et al*. (2010) The evidence for pharmacological treatment of neuropathic pain, Pain, 150: 573-581, doi: 10.1016/j.pain.2010.06.019**
- NEWMEDS: Jacquemont, S *et al*. (2011) Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus, Nature, 478: 97-U111, doi: 10.1038/nature10406
- NEWMEDS: Kirov, G *et al*. (2012) De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia, Molecular Psychiatry, 17: 142-153, doi: 10.1038/mp.2011.154
- Open PHACTS: Williams, AJ *et al*. (2012) Towards a gold standard: regarding quality in public domain chemistry databases and approaches to improving the situation, Drug Discovery Today, 17: 685-701, doi: 10.1016/j.drudis.2012.02.013
- **Pharma-Cog: Frisoni, GB *et al*. (2010) The clinical use of structural MRI in Alzheimer disease, Nature Reviews Neurology, 6: 67-77, doi: 10.1176/appi.ajp.2010.09111661**

Papers with the highest interdisciplinarity have been defined as those with highest diffusion score as defined Carley and Porter (Section 3.1.3).[23]

Section A2.3 lists the ten papers from IMI projects scoring highest on interdisciplinarity. These papers were selected using similar criteria as in the previous report to IMI.

Papers are listed in ascending alphabetical order (project, first author). Three of these papers were also identified as highly-cited (Section A2.1) and are highlighted (bold text). None of the five papers were also identified as 'hot papers'.

A2.3 TOP FIVE PAPERS WITH HIGHEST DIFFUSION SCORE THAT ARE ASSOCIATED WITH IMI PROJECTS

- eTOX: Audouze, K *et al*. (2010) Deciphering Diseases and Biological Targets for Environmental Chemicals using Toxicogenomics Networks, PLOS Computational Biology, 6, doi: 10.1371/journal.pcbi.1000788

[23] Carley S, Porter A (2012). A forward diversity index. Scientometrics, 90:407-427.

- **EUROPAIN: Lasry-Levy, E *et al*. (2011) Neuropathic Pain and Psychological Morbidity in Patients with Treated Leprosy: A Cross-Sectional Prevalence Study in Mumbai, Plos Neglected Tropical Diseases, 5: , doi: 10.1371/journal.pntd.0000981**

- **EUROPAIN: Wildgaard, K *et al*. (2011) Consequences of persistent pain after lung cancer surgery: a nationwide questionnaire study, Acta Anaesthesiologica Scandinavica, 55: 60-68, doi: 10.1111/j.1399-6576.2010.02357.x**

- Open PHACTS: Sansone, SA *et al*. (2012) Toward interoperable bioscience data, Nature Genetics, 44: 121

- **U-BIOPRED: Auffray, C *et al*. (2010) An Integrative Systems Biology Approach to Understanding Pulmonary Diseases, Chest, 137: 1410-1416, doi: 10.1378/chest.09-1850**

ANNEX 3: SUMMARY BIBLIOMETRIC DATA FOR IMI PROJECTS – CALL 3

TABLE A3.1 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 3 – CITATIONS TO END-2011

	Citation impact				
Project	**Number of papers**	**Normalised at field level**	**Normalised at journal level**	**Average percentile**	**% Highly-cited papers[24]**
ABIRISK	0	0	0	0	0.0%
BioVacSafe	0	0	0	0	0.0%
DIRECT	0	0	0	0	0.0%
EU-AIMS	0	0	0	0	0.0%
EUPATI	0	0	0	0	0.0%
MIP-DILI	0	0	0	0	0.0%
PreDiCT-TB	0	0	0	0	0.0%

TABLE A3.2 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 3 – CITATIONS TO CURRENT

	IMI publications		*Web of Science* publications			
Project	**Total**	**% Open access journals**	**Total**	**Citations**	**Raw citation impact**	**% Top quartile journals**
ABIRISK	0	0.0%	0	0	0.00	0.0%
BioVacSafe	6	0.0%	4	1	0.25	50.0%
DIRECT	1	0.0%	1	0	0.00	0.0%
EU-AIMS	9	0.0%	8	30	3.75	87.5%
EUPATI	0	0.0%	0	0	0.00	0.0%
MIP-DILI	3	0.0%	3	1	0.33	100.0%
PreDiCT-TB	0	0.0%	0	0	0.00	0.0%
Overall (IMI projects)	**320**	**11.3%**	**286**	**1 133**	**3.96**	**79.4%**

All *Web of Science* publications associated with Call 3 projects were published in 2012 and do not have normalised citation impact data at end-2011.

[24] ‘Highly-cited’ refers those articles and reviews belonging to the world’s top decile of papers for journal category and year of publication. A percentage that is above 10 indicates above-average performance.

ANNEX 4: BIBLIOMETRICS AND CITATION ANALYSIS

Bibliometrics are about publications and their citations. The academic field emerged from 'information science' and now usually refers to the methods used to study and index texts and information.

Publications cite other publications. These citation links grow into networks, and their numbers are likely to be related to the significance or impact of the publication. The meaning of the publication is determined from keywords and content. Citation analysis and content analysis have therefore become a common part of bibliometric methodology. Historically, bibliometric methods were used to trace relationships amongst academic journal citations. Now, bibliometrics are 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', which is interpreted as significance or influence on 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 tool that could be applied to 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 IP & Science business of Thomson Reuters).[25]

We can count citations, but they are only 'indicators' of impact or quality – not metrics. Most impact indicators 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.

A4.1 DATA SOURCE

The data we use come from the Thomson Reuters databases underlying the *Web of Knowledge*℠, which gives access not only to journals but also to conference proceedings, books, patents, websites, and chemical structures, compounds and reactions. It has a unified structure that integrates all data and search terms together and therefore provides a level of comparability not found in other databases. It is widely acknowledged to be the world's leading source of citation and bibliometric data. The *Web of Science*℠ is one part of the *Web of Knowledge*, and focuses on research published in journals, conferences and books in science, medicine, arts, humanities and social sciences.

The *Web of Science* was created as an awareness and information retrieval tool but it has acquired an important secondary use as a tool for research evaluation, using citation analysis and bibliometrics. Data coverage is both current and retrospective in the sciences, social sciences, arts and humanities, in some cases back to 1900. Within the research community this data source is often still referred to by the acronym 'ISI'.

Unlike other databases, the *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 12,000 of the highest impact journals worldwide, including Open Access journals, and over 150,000 conference proceedings. The abstracted journals encompass the majority of significant, frequently cited 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 unduly from year to year, which increases the usability of such data for performance evaluation.

[25] Garfield, E (1955) Citation Indexes for Science – New dimension in documentation through association of ideas. *Science*: **122**, 108-111.

Evidence, now as part 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.

A4.2 DATABASE CATEGORIES

The source data can be grouped in various classification systems. Most of these are based on groups of journals that have a relatively high cross-citation linkage and naturally cluster together. Custom classifications use subject maps in third-party data such as the OECD categories set out in the Frascati manual.

Thomson Reuters typically uses the broader field categories in the *Essential Science Indicators* system and the finer journal categories in the *Web of Science*. There are 22 fields in *Essential Science Indicators* and 254 fields in *Web of Science*. In either case, the bibliometric analyses draw on the full range of data available in the underlying database, so analyses in our reports will differ slightly from anything created 'on the fly' from data in the web interface.

The lists of journal categories in these systems are attached at the end of this document.

Most analyses start with an overall view across the data, then move to a view across broad categories and only then focus in at a finer level in the areas of greatest interest to policy, programme or organisational purpose.

A4.3 ASSIGNING PAPERS TO ADDRESSES

A paper is assigned to each country and each organisation whose address appears at least once for any author on that paper. One paper counts once and only once for each assignment, however many address variants occur for the country or organisation. No weighting is applied.

For example, a paper has five authors, thus:

Author	Organisation	Country		
Gurney, KA	Univ Leeds	UK	**Counts for Univ Leeds**	**Counts for UK**
Adams, J	Univ Leeds	UK	No gain for Univ Leeds	No gain for UK
Kochalko, D	Univ C San Diego	USA	**Counts for UCSD**	**Counts for USA**
Munshi, S	Gujarat Univ	India	**Counts for Gujarat Univ**	**Counts for India**
Pendlebury, D	Univ Oregon	USA	**Counts for Univ Oregon**	No gain for USA

So this one paper with five authors would be included once in the tallies for each of four universities and once in the tallies for each of three countries.

Work carried out within Thomson Reuters, and research published elsewhere, indicates that fractional weighting based on the balance of authors by organisation and country makes little difference to the conclusions of an analysis at an aggregate level. Such fractional analysis can introduce unforeseen errors in the attempt to create a detailed but uncertain assignment. Partitioning credit would make a greater difference at a detailed, group level but the analysis can then be manually validated.

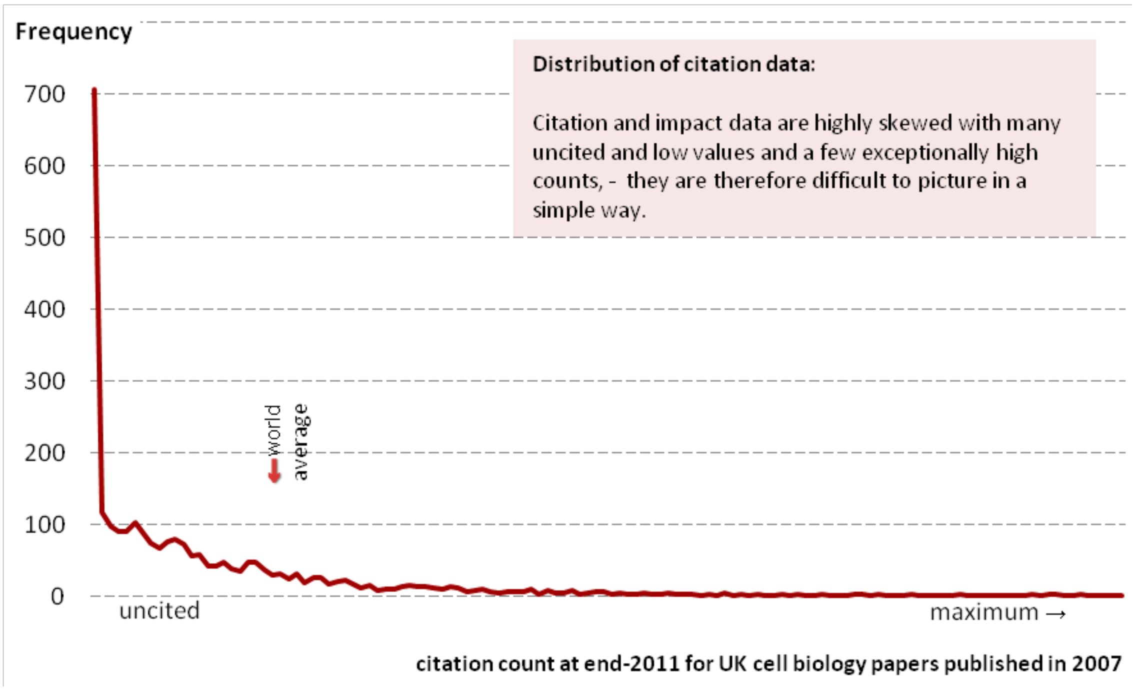
A4.4 CITATION COUNTS

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

Why are 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.

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 publication 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.



The figure shows the skewed distribution of more or less frequently cited papers from a sample of UK authored publications 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.

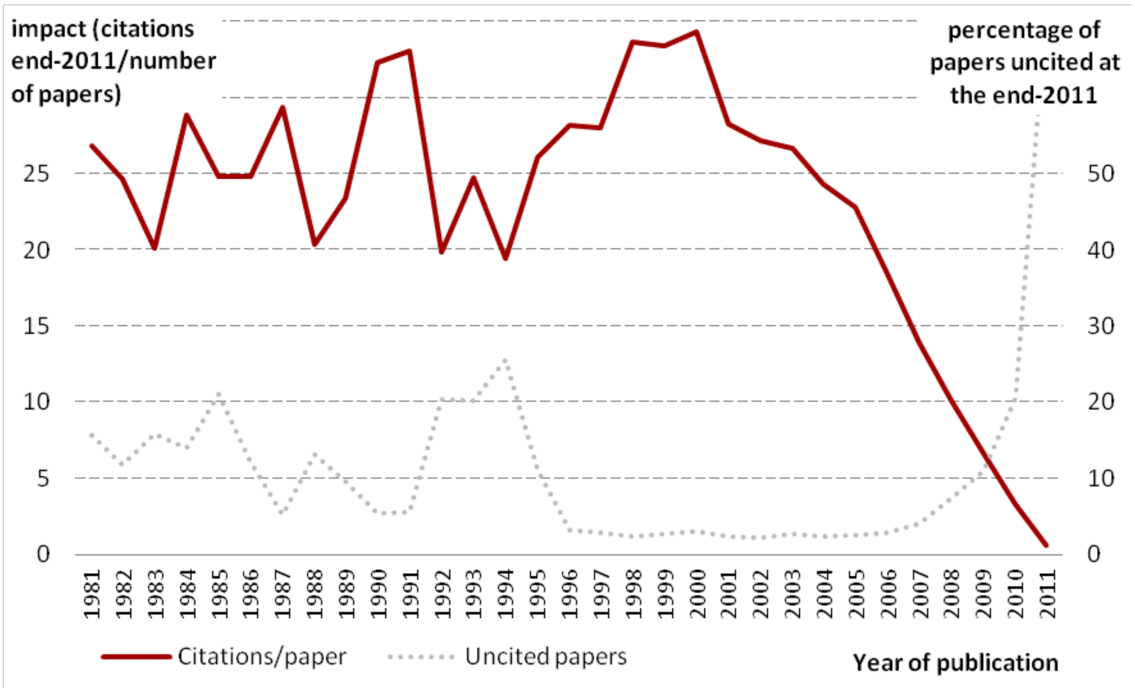
We do not seek to account separately for the effect of self-citation. If the citation count is significantly affected by self-citation then the paper is likely to have been infrequently cited. This is therefore only of consequence for low impact activity. Studies show that for large samples at national and organisational level the effect of self-citation has little or no effect on the analytical outcomes and would not alter interpretation of the results.

A4.5 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.



A4.6 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 252 *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 normalised citation impact data. The journals included in the Thomson Reuters databases and how they are selected are detailed here http://scientific.thomsonreuters.com/mjl/.

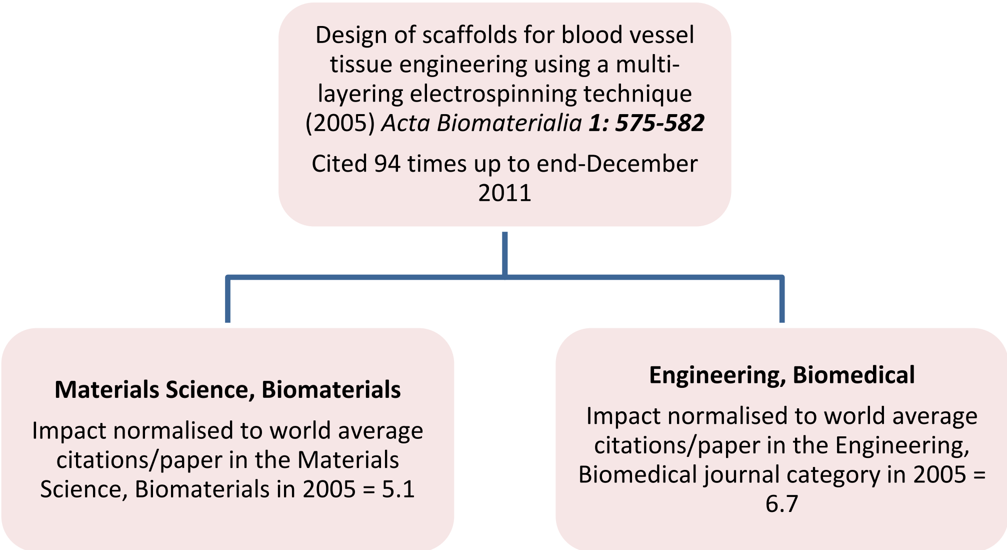
Some journals with a very diverse content, including the prestigious journals *Nature* and *Science* were classified as ***Multidisciplinary*** in databases created prior to 2007. The papers from these ***Multidisciplinary*** journals are now re-assigned to more specific research fields using an algorithm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

A4.7 NORMALISED CITATION IMPACT

Because citations accumulate over time at a rate that is dependent upon the field of research, 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.

We only use citation counts for reviews and articles in calculations of impact, because document type influences the citation count. For example, a review will often be cited more frequently than an article in the same field, but editorials and meeting abstracts are rarely cited and citation rates for conference proceedings are extremely variable. The most common normalisation factors are the average citations per paper for (1) the year and (2) either the field or the 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 'normalised impact', or NCI. The following schematic illustrates how the normalised citation 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 normalised citation impact (in the above example the category-specific NCI_F for ***Materials Science, Biomaterials*** is 5.1 and the category-specific NCI_F for ***Engineering, Biomedical*** is higher at 6.7). Most papers (nearly two-thirds) are assigned to a single journal category whilst a minority are assigned to more than 5.

Citation data provided by Thomson Reuters 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.

World average impact data are sourced from the Thomson Reuters National Science Indicators baseline data for 2011.

A4.8 MEAN NORMALISED CITATION IMPACT

Research performance has historically been indexed by using average citation impact, usually compared to a world average that accounts for time and discipline. As noted, however, the distribution of citations amongst papers is highly skewed because many papers are never cited while a few papers accumulate very large citation counts. That means that an average may be misleading if assumptions are made about the distribution of the underlying data.

In fact, almost all research activity metrics are skewed: for research income, PhD numbers and publications there are many low activity values and a few exceptionally high values. In reality, therefore, the skewed distribution means that average impact tends to be greater than and often significantly different from either the median or mode in the distribution. This should be borne in mind when reviewing analytical outcomes.

The average (normalised) citation 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. In the example above, the average citation impact of the *Acta Biomaterialia* paper can be expressed as ((5.1 + 6.7)/2) = 5.9.

A4.9 IMPACT PROFILES®

We have developed a bibliometric methodology[26] that 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.

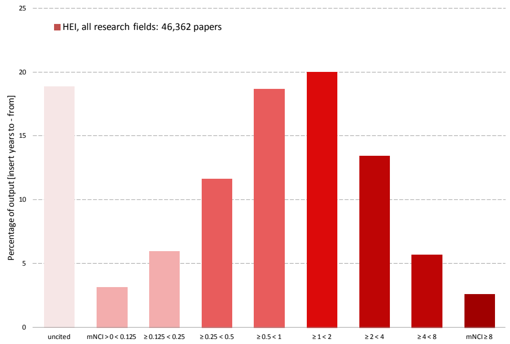
Papers which are "highly-cited" are often defined in our reports as those with an average citation impact (NCI_F) greater than or equal to 4.0, i.e. those papers which have received greater than or equal to four times the world average number of citations for papers in that subject published in that year. This differs from Thomson Reuters database of global highly-cited papers, which are the top 1% most frequently cited for their field and year. The top percentile is a powerful indicator of leading performance but is too stringent a threshold for most management analyses.

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 are, of course, those that are most recently published.

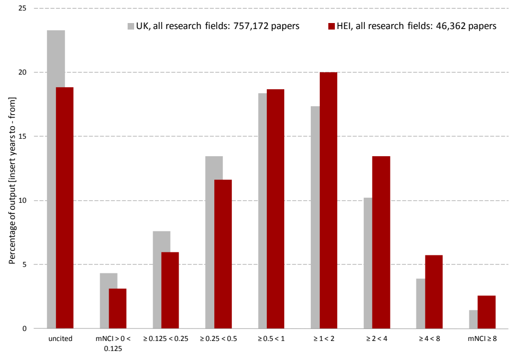
[26] Adams J, Gurney K & Marshall S (2007) Profiling citation impact: A new methodology. *Scientometrics* **72**: 325-344.

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

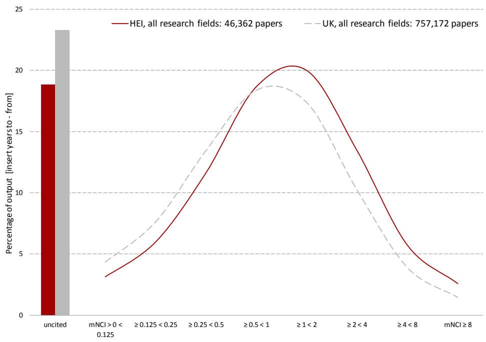
A



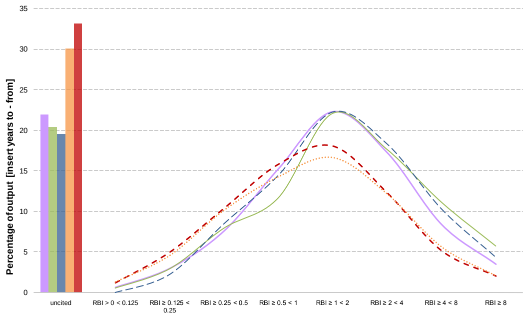
B



C



D



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.

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**.

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 to world average (which becomes 1.0 in this graph). Normalised 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 papers are never subsequently cited after publication, even by their authors. This accounts for about half the total global 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 but not a quality factor. While there is also an obvious increase in the likelihood of citation over time, 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.

A4.10 EVIDENCE QUALITY INDEX

Another bibliometric indicator which can be very useful in small datasets is the *Evidence* quality index. This indicator is calculated from the citation impact relative to the specific journal in which the paper is published.

For the paper on page 63 which has been cited 94 times to the end-December 2011, the expected citation rate for a paper in *Acta Biomaterialia* published in 2005 would be 23.2. Therefore, this paper has been cited more than expected for the journal. For a set of papers, we calculate the quality index as the percentage of papers which are cited more than expected for the relevant journals.

This indicator should be considered alongside that of normalised citation impact as they are complementary. For example, a given set of publications may have a high *Evidence* quality index and relatively low citation impact. This would imply that these papers were well cited in relation to other papers in that journal and that year but when considered in relation to other papers published in more highly-cited journals in the same research field did not perform as well. The interpretation would be that the publications are in relatively low impact journals.

A4.11 *WEB OF SCIENCE* JOURNAL CATEGORIES

Acoustics	Classics	Engineering, multidisciplinary
Agricultural economics & policy	Clinical neurology	Engineering, ocean
Agricultural engineering	Communication	Engineering, petroleum
Agriculture, dairy & animal science	Computer science, artificial intelligence	Entomology
Agriculture, multidisciplinary	Computer science, cybernetics	Environmental sciences
Agriculture, soil science	Computer science, hardware & architecture	Environmental studies
Agronomy	Computer science, information systems	Ergonomics
Allergy	Computer science, interdisciplinary applications	Ethics
Anatomy & morphology	Computer science, software engineering	Ethnic studies
Andrology	Computer science, theory & methods	Evolutionary biology
Anesthesiology	Construction & building technology	Family studies
Anthropology	Criminology & penology	Film, radio, television
Applied linguistics	Critical care medicine	Fisheries
Archaeology	Crystallography	Folklore
Architecture	Dance	Food science & technology
Area studies	Demography	Forestry
Art	Dentistry, oral surgery & medicine	Gastroenterology & hepatology
Asian studies	Dermatology	Genetics & heredity
Astronomy & astrophysics	Developmental biology	Geochemistry & geophysics
Automation & control systems	Ecology	Geography
Behavioral sciences	Economics	Geography, physical
Biochemical research methods	Education & educational research	Geology
Biochemistry & molecular biology	Education, scientific disciplines	Geosciences, multidisciplinary
Biodiversity conservation	Education, special	Geriatrics & gerontology
Biology	Electrochemistry	Health care sciences & services
Biology, miscellaneous	Emergency medicine	Health policy & services
Biophysics	Endocrinology & metabolism	Hematology
Biotechnology & applied microbiology	Energy & fuels	History
Business	Engineering, aerospace	History & philosophy of science
Business, finance	Engineering, biomedical	History of social sciences
Cardiac & cardiovascular systems	Engineering, chemical	Horticulture
Cell biology	Engineering, civil	Humanities, multidisciplinary
Chemistry, analytical	Engineering, electrical & electronic	Imaging science & photographic technology
Chemistry, applied	Engineering, environmental	Immunology
Chemistry, inorganic & nuclear	Engineering, geological	Industrial relations & labor
Chemistry, medicinal	Engineering, industrial	Infectious diseases
Chemistry, multidisciplinary	Engineering, manufacturing	Information & library science
Chemistry, organic	Engineering, marine	Instruments & instrumentation
Chemistry, physical	Engineering, mechanical	Integrative & complementary medicine
International relations	Mining & mineral processing	Psychology

Language & linguistics
Language & linguistics theory
Law
Limnology
Linguistics
Literary reviews
Literary theory & criticism
Literature
Literature, African, Australian, Canadian
Literature, American
Literature, British Isles
Literature, German, Dutch, Scandinavian
Literature, romance
Literature, Slavic
Management
Marine & freshwater biology
Materials science, biomaterials
Materials science, ceramics
Materials science, characterization & testing
Materials science, coatings & films
Materials science, composites
Materials science, multidisciplinary
Materials science, paper & wood
Materials science, textiles
Math & computational biology
Mathematics
Mathematics, applied
Mathematics, interdisciplinary applications
Mechanics
Medical ethics
Medical informatics
Medical laboratory technology
Medicine, general & internal
Medicine, legal
Medicine, research & experimental
Medieval & renaissance studies
Metallurgy & metallurgical engineering
Meteorology & atmospheric sci
Microbiology
Microscopy
Mineralogy

Multidisciplinary sciences
Music
Mycology
Nanoscience & nanotechnology
Neuroimaging
Neurosciences
Nuclear science & technology
Nursing
Nutrition & dietetics
Obstetrics & gynecology
Oceanography
Oncology
Operations research & management science
Ophthalmology
Optics
Ornithology
Orthopedics
Otorhinolaryngology
Paleontology
Parasitology
Pathology
Pediatrics
Peripheral vascular disease
Pharmacology & pharmacy
Philosophy
Physics, applied
Physics, atomic, molecular & chemical
Physics, condensed matter
Physics, fluids & plasmas
Physics, mathematical
Physics, multidisciplinary
Physics, nuclear
Physics, particles & fields
Physiology
Planning & development
Plant sciences
Poetry
Political science
Polymer science
Psychiatry

Psychology, applied
Psychology, biological
Psychology, clinical
Psychology, developmental
Psychology, educational
Psychology, experimental
Psychology, mathematical
Psychology, multidisciplinary
Psychology, psychoanalysis
Psychology, social
Public administration
Public, environmental & occupational health
Radiology, nuclear medicine & medical imaging
Rehabilitation
Religion
Remote sensing
Reproductive biology
Respiratory system
Rheumatology
Robotics
Social issues
Social sciences, biomedical
Social sci, interdisciplinary
Social sci, mathematical methods
Social work
Sociology
Soil science
Spectroscopy
Sport sciences
Statistics & probability
Substance abuse
Surgery
Telecommunications
Theater
Thermodynamics
Toxicology
Transplantation
Transportation
Transportation science & technology
Tropical medicine

Urban studies

Urology & nephrology
Veterinary
Veterinary sciences
Virology
Water resources
Women's studies
Zoology