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

IMI EXECUTIVE OFFICE

Initial report: October 2012

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

1 EXECUTIVE SUMMARY

Before discussing the bibliometric analyses and indicators presented in this report it should be noted that the first IMI project started in May 2009 and currently there are 30 active projects, out 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 JU projects that are analysed here are therefore relatively young, and early bibliometric indicators may not fully reflect their eventual impact.

- A total of 214 publications resulting from IMI projects were identified (up to the cut-off date of the end of August 2012). Most of the publications associated with IMI JU-funded projects are relatively recent. Publication output has increased each year since 2009 with a substantial increase between 2010 and 2011. Also the publication output for 2012 up to the end of August has already exceeded the 2011 total. It is expected that publication output will continue to grow non-linearly as the number of funded projects increases and those projects yield results for publication (Section 4.3).
- Review papers are accounting for both an increasing volume and an increasing proportion of IMI project research over time. This could be taken as indicative of the increasing esteem in which research from IMI project is held as projects accumulate both publications and expertise (Section 4.2).
- IMI project publications have been published in a total of 119 journals, of which 95 are ranked in the top quartile (by Journal Impact Factor) of journals in their specific research fields. A total of 151 publications (82.7% of IMI project publications) have been published in these well-regarded journals, including, Nature, JAMA, PNAS and Nature Genetics (Section 4.4).
- Around one-fifth (20.7%) of IMI project research (especially from the NEWMEDS, EUROPAIN and PharmaCog projects) is published in journals associated with Neurosciences. Publication output in Pharmacology & Pharmacy journals (12.5%) is associated with the PROTECT and NEWMEDS projects. Among the other journal categories, output is more evenly spread (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. 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).
- IMI project research published in Anaesthesiology journals is exceptionally well-cited with average citation impact around four times the European and world benchmarks. This performance is driven partly by two highly-cited papers, one of which is identified as a 'hot paper' in the Thomson Reuters databases (Section 4.6).
- Analysis shows that the majority of IMI project publications are associated with Call 1 with lower output from projects in the newer funding calls. On a project level, the average citation impact of all but one of the Call 1 projects is well above world citation impact (1.0), with the average citation impact of this project (IMIDIA), approaching world average (Section 5).

Patents assigned to IMI-supported researchers were identified using researcher names, projects, and affiliations supplied by IMI. For this initial report, data and analyses are limited to the researchers funded by the first IMI funding call (Call 1) in 2008.

- In total 1 245 inventions were identified and associated with at least one IMI researcher. Of these, around 10% (116 of the 1 245 inventions) were identified as being of high IMI relevance using key word searching in the patent abstract (Section 6.1).
- The IMIDIA project appears to have stimulated the greatest activity in patents and inventions, however, it is the NEWMEDS and EUROPAIN projects that are associated with the greatest number of inventions closely related to IMI.
- The most active patenting entity appears to be the University of Cambridge (and Cambridge Enterprise Ltd) with 24 inventions of high IMI relevance all of which can be associated with Sabine Bahn, a researcher funded by the NEWMEDS project (Table 6.2.4).

- Most of the other patent activity has been initiated by corporate entities including Roche and Sanofi-Aventis. However, most of this activity, even that of high IMI relevance, appears not to be associated with IMI projects, with the exception of Roche and the IMIDIA project.

The productivity, research performance and collaboration of researchers funded by Call 1 IMI projects were assessed by analysing the total publication output these individuals (not limited to publications acknowledging funding from IMI projects). 1470 researchers were included in the analysis and 9716 of their publications were identified for the period January 2007-August 2012.

- Analysis shows that publication output is, not surprisingly, higher for IMI-supported researchers based in academic institutions and research institutions compared to industry and SMEs (Section 7.3).
- Researchers who are based in academic or in other research-active institutions also have the strongest research performance. Of the 385 publishing academic-based researchers, 23% of researchers have published at least one ‘hot paper’, 20% have an h-index of at least 10 and the majority have published most frequently 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. About three quarters of Call 1 researchers that were indexed in the *Web of Science* collaborated (co-authored) with at least one other IMI researcher during the period January 2007-August 2012 (Section 7.5).
- As expected, co-authorship is more common among researchers in the same sector than among researchers in different sectors. However, there are also substantial co-authorship activities among researchers from different sectors, accounting for 40% 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 20% of co-authorship relationships being cross-project (Section 7.6).
- The two entities which span the most communities are Astra Zeneca (corporate) and Imperial College London (academic) with 18 and 15 active researchers respectively (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 first 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 seven Calls for proposals to be funded under the initiative. The first funding call was announced in 2008 and the latest, 7th, funding call was launched on 17th July 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
- Section 4 presents bibliometric indicators for IMI-supported researchers and analyses of collaboration between these individuals
- Sections 5 and 6 present citation analyses of research from IMI projects
- Section 7 presents an analysis of IMI patent data

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.[1] 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

[1] *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 September 2012 with citation counts as at mid-August 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.[2] 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, an 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.[3] 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.

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

---

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

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

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 30 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,[4] 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 disciplines. This creates a problem when we try to define, for example, ‘Parasitology research’. Is

[4] 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

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 7.1 and outlined in Figure 7.2.1.

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. The descriptor 'IMI project research' is also used to refer to this dataset.

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

3.2.1 BACKGROUND

Research funding agencies often choose to measure research output not only in the form of scientific or scholarly articles, but also in the form of patent applications, especially if the mission of the funding agency is to encourage delivery of the benefit of research to the public in the form of commercialised products or services that accrue to public benefit. Some agencies make reporting on patent applications an obligation for the grantee or even make the filing of patent applications per se an obligation.

Some agencies also provide incentives for patent applications in the form of supplemental funding to defray costs of obtaining patent protection. Regardless of which path an agency chooses, it would be necessary to track patent information as part of the outcomes management for the program.

Thomson Reuters is a premier provider of patent monitoring tools and analytic services on patent information, and can assist funding agencies in assessing grantees and in providing public accountability for their use of funds.

3.2.2 PATENT DATA SOURCE

The patent analysis in this report used the Derwent World Patents Index (DWPI), a value added database which covers 47 authorities worldwide with all abstracts in English language to allow ease of searching. The DWPI database allows us to utilise the DWPI editorially enhanced text and detailed Manual Coding. The DWPI database includes patent applications and/or granted patents from the following 47 patent authorities, and utility models (less robust patents) are also available for many countries:

Australia AU, Argentina AR, Austria AT, Brazil BR, Belgium BE, Canada CA, China CN, Czech Republic CZ, Czechoslovakia CS, Denmark DK, Finland FI, France FR, East Germany DD, European Patent Office EP, Germany DE, Gulf Cooperative Council GC, Hong Kong HK, Hungary HU, India IN, Ireland IE, Israel IL, Italy IT, Japan JP, Luxembourg LU, Malaysia MY, Mexico MX, Netherlands NL, New Zealand NZ, Norway NO, Patent Cooperation Treaty Applications WO, Philippines PH, Portugal PT, Romania RO, Russia RU, Singapore SG, Slovakia SK, South Africa ZA, South Korea KR, Soviet Union SU, Spain ES, Sweden SE, Switzerland CH, Taiwan TW, Thailand TH, United Kingdom GB, United States US, Vietnam VN.

Inventions are enumerated using the DWPI patent family count to avoid counting the same invention more than once. A single patent only provides a statutory monopoly for the patented technology within the legal jurisdiction of the authority that granted the patent. This means that inventors must file applications for a patent in each jurisdiction where they foresee a need for protection.

Each related patent application and granted patent is added to the DWPI family record as it is published. This being the case, all counts of inventions in this report refer to patent families or inventions, and not to individual patent documents. For example, the European application, European granted patent and the US granted patent for a single invention family is counted in aggregate as “1” in all the analyses in this report unless otherwise noted.

This provides a more accurate measure of the level of inventive activity from a company within the technical space, and a truer picture of the overall level of innovation across the field as a whole.

3.2.3 PATENT INDICATORS

Thomson Reuters have compiled lists of published applications and granted patents, and have set up an alerting mechanism allowing future patent publications to be collected. These lists will provide IMI with current information and data analyses on technology transfer activities associated with their funded researchers.

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).

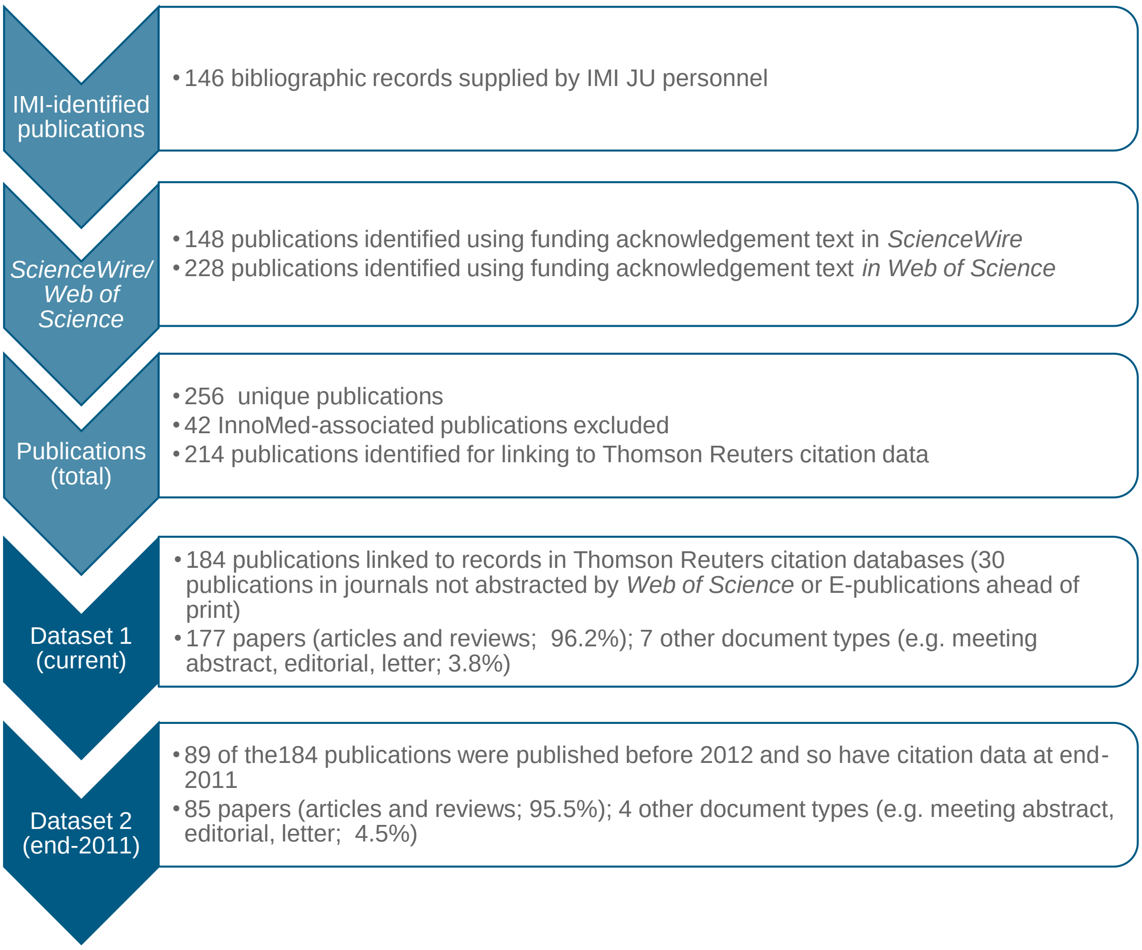
4.1 PUBLICATIONS FROM IMI-SUPPORTED PROJECTS

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 *ScienceWire* or 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. A number of publications associated with a precursor project, InnoMed, were identified by IMI staff and excluded from the analyses. All remaining publications were assigned to specific projects by IMI staff or through the text in the abstract or funding acknowledgements of the publications.

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

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



4.1.1 CITATION DATA FOR PUBLICATIONS FROM IMI-SUPPORTED PROJECTS

A total of 214 publications by IMI-supported researchers were identified and 184 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.* Counts have been extracted at two distinct census points– current (mid-August) and end-2011.

The 'current' census point (Dataset 1) will allow 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 average.

Dataset 2 with a census point of end-2011, where citation counts are collated to the same census point as used in the calculation of global baselines will be used to calculate the normalised citation impact of IMI research (see Section 3.1.3) which can be benchmarked to world averages.

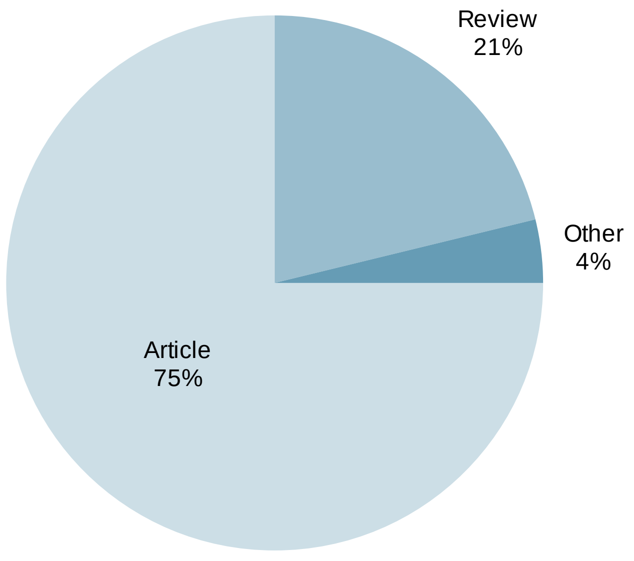
The analyses presented in this Section cover both raw and normalised citation impact data. It is important to note that all publication types will have raw citation impact values while only papers (substantive journal articles and reviews) are used in the analyses with normalised citation impact. Normalised bibliometric indicators for the papers in this report have been 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 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 184 unique *Web of Science* publications linked to citation data (as outlined in Figure 4.1.1). Over 95% of these documents (96.2%) were substantive articles and reviews with only seven documents not falling into this grouping. These seven documents (classified as 'Other') include three editorials, two meeting abstracts, one letter and one news item.

FIGURE 4.2.1 CATEGORISATION OF IMI PROJECT RESEARCH BY DOCUMENT TYPE



4.3 TRENDS IN PUBLICATION OUTPUT

Publication output has increased each year with a substantial increase between 2010 and 2011 (Figure 4.3.1).

The volume of research published to-date in 2012 has already exceeded the 2011 total. It is expected that publication output will continue to grow non-linearly as the number of funded projects increases and those projects funded earliest in the programme yield results for publication.

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

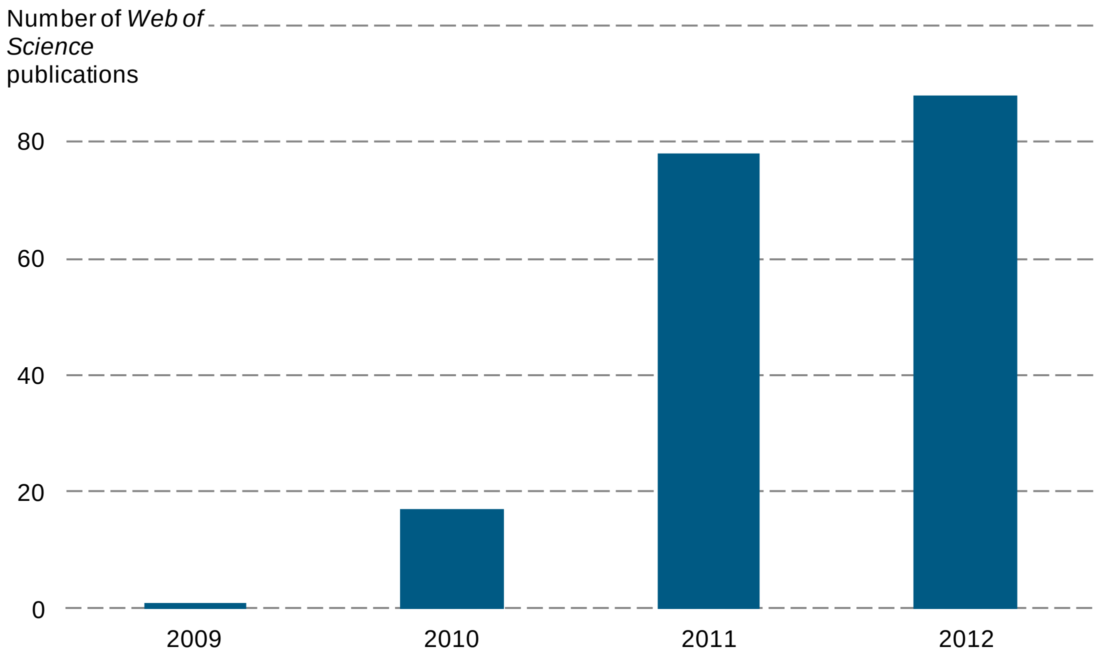
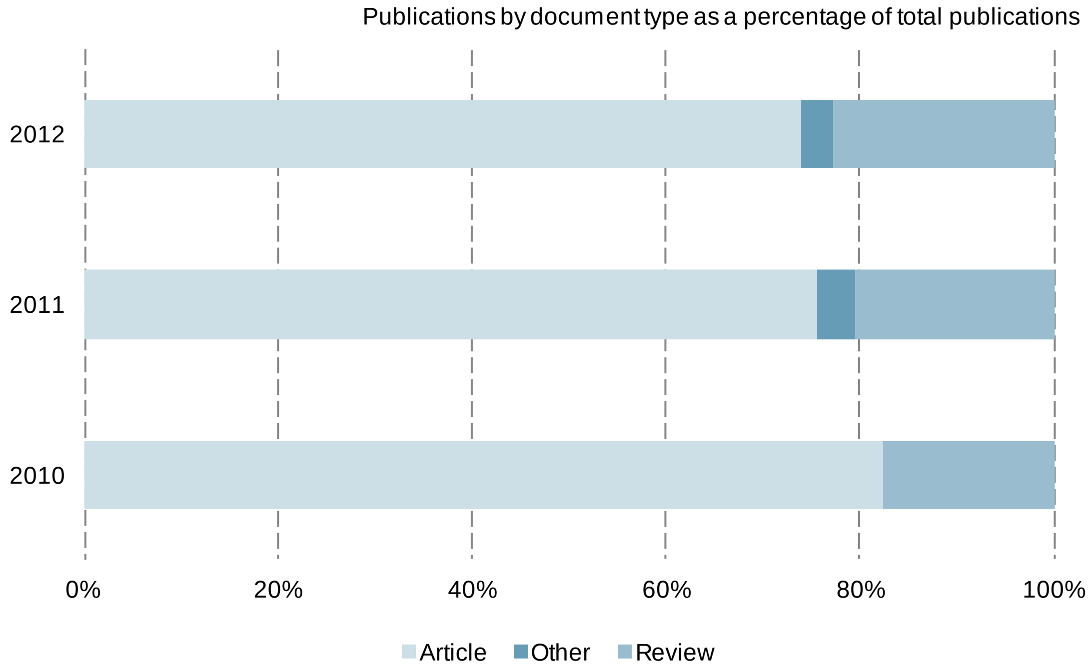


Figure 4.3.2 shows the proportion of papers (articles and reviews) relative to other document types for IMI project research over a 3-year time period.[5]

[5] 2009 publications comprise a single meeting abstract – this has been omitted from Figure 4.3.2 for clarity.

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



These two analyses show that reviews account for both an increasing volume and an increasing proportion of IMI project research over time. This could be taken as indicative of the increasing esteem in which research from IMI project is held as projects accumulate both publications and expertise.

Analysis at journal level has revealed that this trend is not limited to particular journals, journal categories or IMI projects but is generalised.

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

The 18 journals appearing most frequently in the IMI project publications dataset, 2009-2012 are listed in Table 4.4.1 (a total of 32 journal titles are used more than once).

Together, the items in the 18 most frequently used journals comprise 69 *Web of Science* publications, or just over one-third (37.5%) of the total number of items in the dataset.

This core set of journals highlights the range of IMI-supported projects – the top 18 journals include titles focused on rheumatology and endocrinology as well as elite multidisciplinary and field-specific journals (PNAS and Nature Genetics respectively).

All but four of the journals in Table 4.4.1 are ranked in the top quartile (by Journal Impact Factor) of journals in their specific research fields.[6] Journals not meeting this threshold are shaded in grey.

IMI project publications have been published in a total of 119 journals, of which 95 are ranked in the top quartile (by Journal Impact Factor) of journals in their specific research fields. A total of 151 publications (82.7% of IMI project publications) have been published in these well-regarded journals.

[6] Where a particular journal is associated with more than one *Web of Science* journal category that in which it performs best has been used to assign the quartile. If more than one category is listed, this means the journal ranks equally with respect to quartile in these categories.

TABLE 4.4.1 JOURNALS IN WHICH IMI PROJECT PUBLICATIONS HAVE BEEN PUBLISHED MOST FREQUENTLY (2009-2012), 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	9	9	4.092	Biology
Pain	8	8	5.777	Neurosciences; Clinical Neurology; Anaesthesiology
European Journal of Cancer	6	6	5.536	Oncology
Diabetes	4	4	8.286	Endocrinology & Metabolism
Annals of the Rheumatic Diseases	3	3	8.727	Rheumatology
Arthritis and Rheumatism	3	2	7.866	Rheumatology
Diabetologia	3	3	6.814	Endocrinology & Metabolism
Drug Discovery Today	3	3	6.828	Pharmacology & Pharmacy
Expert Opinion on Drug Metabolism & Toxicology	3	3	3.119	Pharmacology & Pharmacy; Biochemistry & Molecular Biology
Health and Quality of Life Outcomes	3	3	2.112	Health Care Sciences & Services
Journal of Alzheimers Disease	3	3	3.745	Neurosciences
Journal of Clinical Investigation	3	3	13.069	Research & Experimental Medicine; Immunology
Nature Genetics	3	1	35.532	Genetics & Heredity
Neuroimage	3	3	5.895	Radiology, Nuclear Medicine & Medical Imaging; Neuroimaging; Neurosciences
Pharmacoepidemiology and Drug Safety	3	3	2.528	Pharmacology & Pharmacy
PLOS Computational Biology	3	3	5.215	Biochemical Research Methods; Mathematical & Computational Biology
Proceedings of the National Academy of Sciences USA	3	3	9.681	Multidisciplinary Sciences
Psychopharmacology	3	3	4.077	Psychiatry; Neurosciences; Pharmacology & Pharmacy

Table 4.4.2 lists the twenty journals with the highest journal impact factor (JIF) used in the IMI project publications dataset. Overall, there are 23 publications in journal titles with an impact factor of 10 or above and 8 publications in journal titles with an impact factor of 20 or above. These publications comprise more than one-tenth (12.5%) of the total IMI project publications dataset and indicates that a substantial percentage of research from IMI projects is being published in highly-regarded journals.

TABLE 4.4.2 JOURNALS IN WHICH IMI PROJECT PUBLICATIONS HAVE BEEN PUBLISHED MOST FREQUENTLY (2009-2012), TOP TWENTY RANKED BY JOURNAL IMPACT FACTOR

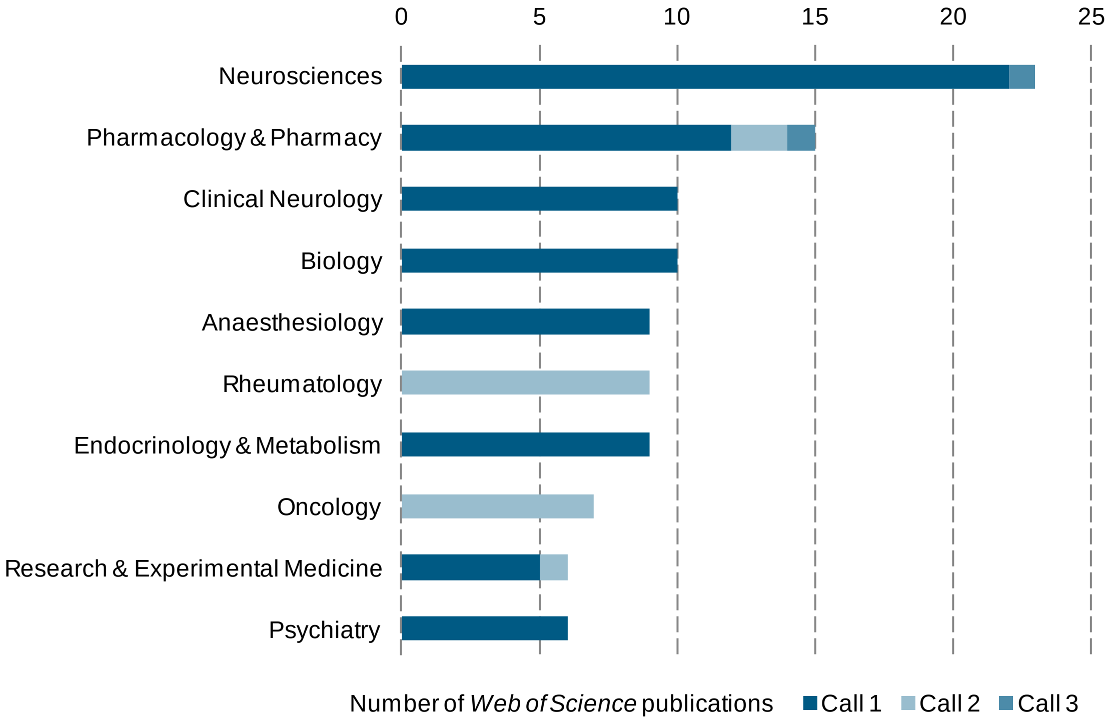
Journal	Number of *Web of Science* publications	Number of papers	Journal Impact Factor (2011)	Journal categories
Nature	1	1	36.28	Multidisciplinary Sciences
Nature Genetics	3	1	35.532	Genetics & Heredity
JAMA-Journal of the American Medical Association	1	1	30.026	General & Internal Medicine
Nature Reviews Drug Discovery	1		29.008	Pharmacology & Pharmacy; Biotechnology & Applied Microbiology
Lancet Neurology	1	1	23.462	Clinical Neurology
Nature Biotechnology	1		23.268	Biotechnology & Applied Microbiology
Nature Neuroscience	1	1	15.531	Neurosciences
British Medical Journal	1	1	14.093	General & Internal Medicine
Journal of Experimental Medicine	1	1	13.853	Research & Experimental Medicine
Molecular Psychiatry	2	2	13.668	Neurosciences; Psychiatry; Biochemistry & Molecular Biology
Journal of Clinical Investigation	3	3	13.069	Research & Experimental Medicine; Immunology
American Journal of Psychiatry	1	1	12.539	Psychiatry
Journal of Allergy and Clinical Immunology	2	2	11.003	Allergy; Immunology
ACS Nano	1	1	10.774	Physical Chemistry; Chemistry, Multidisciplinary; Nanoscience & Nanotechnology; Materials Science, Multidisciplinary
European Heart Journal	1	1	10.478	Cardiac & Cardiovascular Systems
Trends in Immunology	2	2	10.403	Immunology
Nature Protocols	1	1	9.924	Biochemical Research Methods
Journal of the American Chemical Society	1	1	9.907	Chemistry, Multidisciplinary
Proceedings of the National Academy of Sciences USA (PNAS)	3	3	9.681	Multidisciplinary Sciences
Trends in Biotechnology	1	1	9.148	Biotechnology & Applied Microbiology

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[7] into which IMI project research falls

Around one-fifth (20.7%) of IMI project research falls within the journal category of Neurosciences with more than one-tenth published in Pharmacology & Pharmacy journals (12.5%). Among the other journal categories, output is more evenly spread.

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



Publication output from the NEWMEDS, EUROPAIN and PharmaCog projects (11, 5 and 5 publications respectively) is associated mainly with the journal category of Neurosciences.

Publications from the PROTECT and NEWMEDS projects (6 and 3 publications respectively) have been mainly published in journals assigned to the Pharmacology & Pharmacy category.

IMIDIA is associated with 6 of the 9 publications in Endocrinology & Metabolism.

All of the publications in Psychiatry journals are associated with the NEWMEDS project and all of the publications in Rheumatology journals are associated with the BTCure project. The BTCure project was part of the Call 2 portfolio and so has been quick to publish in scientific journals.

Together, the most frequently used journal categories in Figure 4.5.1 reflect the breadth and depth of the IMI project portfolio which contains both cross-cutting and more specialised research.

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

[7] 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) and restricted to categories with six or more publications.

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 project research 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 (mid-August) citation counts (see Section 4.1.1)

The average citation impact for IMI project research is 1.34 (where world average is 1.0) for the 2-year period, 2010-2011. For comparison, the EU's average citation impact[8,9] relative to world baseline for the same 2-year period in similar research fields was 1.14 and for the IMI researchers dataset was 2.09 (Table 4.6.1).

Similarly, using current citation counts and raw citation impact, IMI project research performs less well compared to the IMI-supported researchers dataset (3.39 compared to 5.70; Table 4.6.2).

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 project research is 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).

At this initial stage the IMI project research dataset is very small in comparison with the IMI researchers dataset and these analyses this 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.

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

		Citation impact			
	Number of papers	**Normalised at field level**	**Normalised at journal level**	**Average percentile**	**% Highly-cited papers[10]**
IMI projects, 2010-2011	85	1.34	0.95	66.47	10.6%
IMI researchers, 2010-2011	3 538	2.09	1.26	53.19	19.9%

[8] 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.

[9] 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.

[10] '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 4.6.2 SUMMARY CITATION ANALYSIS FOR IMI RESEARCH – CITATIONS TO CURRENT

	Number of publications	Number of *Web of Science* publications	Total citations	Raw citation impact	% *Web of Science* publications in top quartile journals
IMI projects, 2010-current	210	184	623	3.39	82.1%
IMI researchers, 2010-current	n/a	5 910	34 681	5.70	67.9%

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, Anaesthesiology and Psychiatry has higher citation impact, on average, than similar research by IMI-supported researchers.

IMI project research in Anaesthesiology is exceptionally well-cited with average citation impact over four times the European benchmark and four times the world average citation impact (1.0). This performance is driven partly by two highly-cited papers one of which is identified as a 'hot paper' in the Thomson Reuters databases (Annex 2).

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

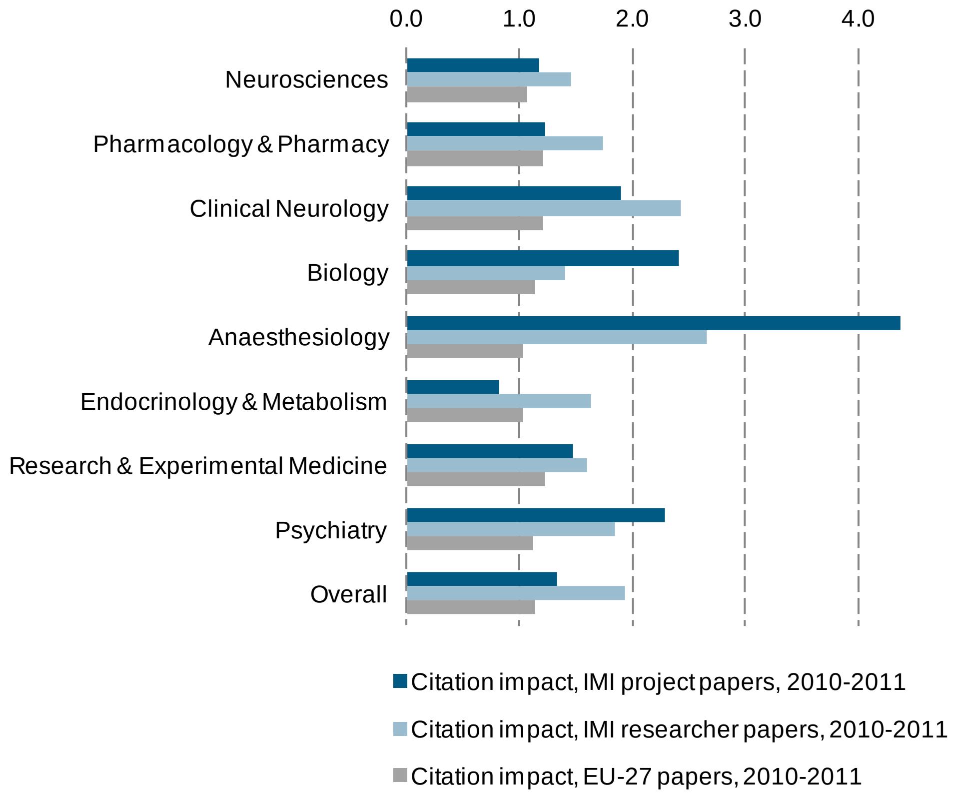


TABLE 4.6.3 SUMMARY OF PUBLICATION OUTPUT AND 2-YEAR AVERAGE CITATION IMPACT FOR IMI PROJECT RESEARCH BY TOP *WEB OF SCIENCE* JOURNAL CATEGORIES, 2010-2011 BENCHMARKED AGAINST IMI RESEARCHERS DATASET AND SIMILAR PUBLICATIONS FROM THE EUROPEAN RESEARCH BASE

	IMI projects		**IMI researchers**		**EU-27**	
***Web of Science* journal category**	**Number of papers**	**Citation impact**	**Number of papers**[11]	**Citation impact**	**Number of papers**[17]	**Citation impact**
Neurosciences	21	1.18	472	1.45	26 862	1.07
Pharmacology & Pharmacy	13	1.22	401	1.74	21 999	1.22
Clinical Neurology	11	1.90	254	2.42	18 856	1.20
Biology	5	2.41	174	1.41	17 551	1.15
Anaesthesiology	6	4.37	34	2.66	3 469	1.03
Endocrinology & Metabolism	7	0.81	253	1.64	13 135	1.03
Research & Experimental Medicine	3	1.48	84	1.61	9 400	1.23
Psychiatry	5	2.30	291	1.84	12 567	1.11
Overall	**85**	**1.34**	**5 483**	**2.09**	**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.

The data in Figure 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 3.

[11] 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.

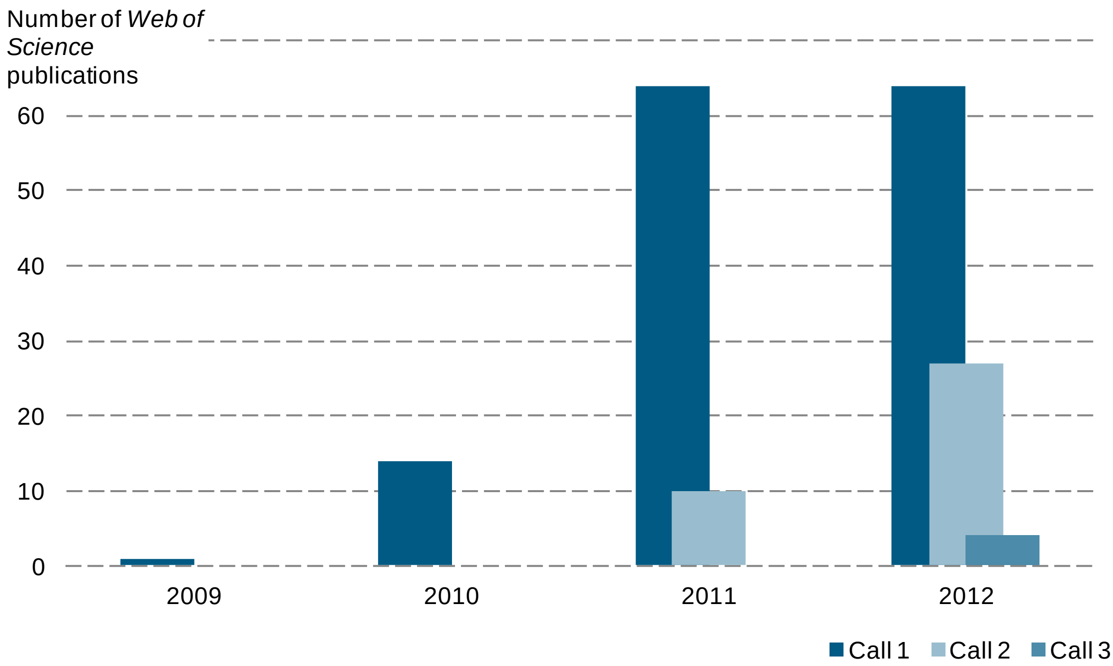
5 CITATION ANALYSIS – AT IMI PROJECT LEVEL

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 shows that the majority of IMI project publications and papers are associated with Call 1 with lower output from projects in the newer funding calls. Most of the research associated with Calls 2 and 3 was published in 2012 and citation data for these publications is sparse in comparison to the first Call.

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

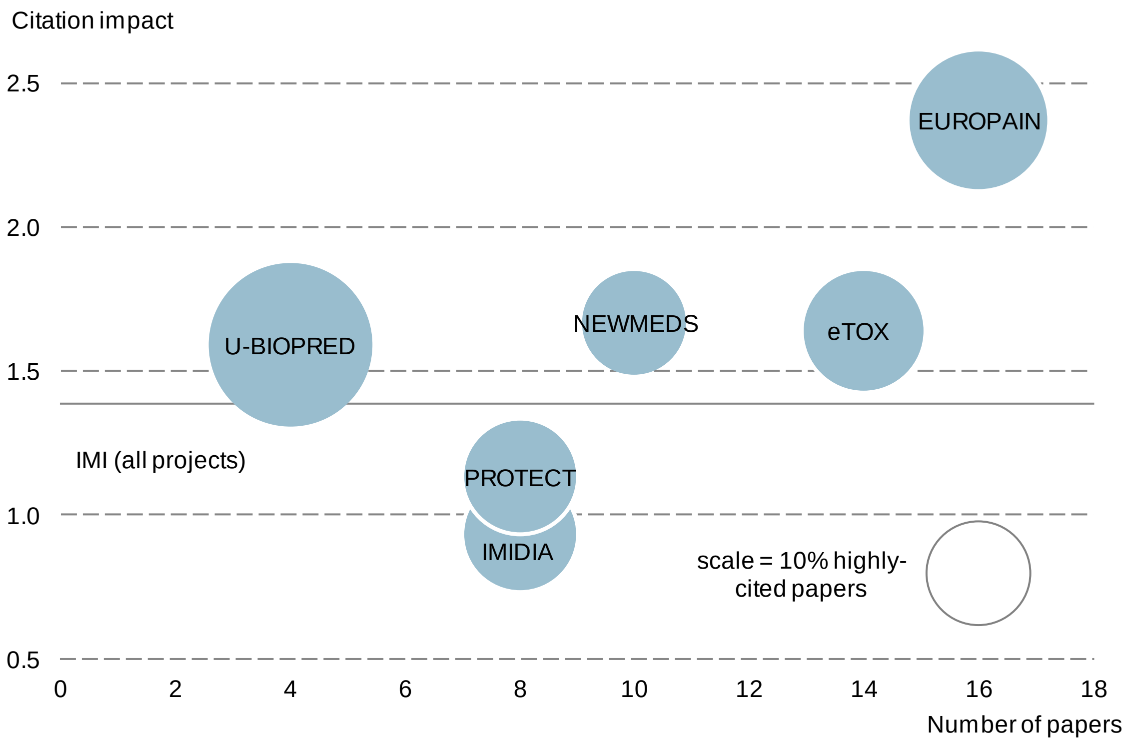


For this reason, summary bibliometric data presented in Section 5.2 will focus on projects funded in Call 1. Summary bibliometric data for projects associated with Calls 2 and 3 are presented in Annex 2.

5.2 SUMMARY BIBLIOMETRIC ANALYSES FOR IMI PROJECTS – CALL 1

Figure 5.2.1 presents a 'bubble-chart' visualisation of IMI project 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



The data in Figure 5.2.1 show that the average citation impact of all but one of the projects is well above world citation impact (1.0) with the average citation impact of this project (IMIDIA), approaching world average.

Research associated with the EUROPAIN project is very well-cited with a mean citation impact almost twice the average for the dataset and well over twice world average. Two years after funding (in 2009) almost one-fifth (18.8%) of EUROPAIN papers are highly-cited.

Though paper numbers are small, early indications are that U-BIOPRED is publishing well-cited research as it has already accumulated one highly-cited paper.

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 (mid-August) 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). All though each of these projects has one publication associated with them, the journals in which the publications appear are not currently abstracted in the *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**[12]
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%
PharmaCog	9	1.03	0.91	57.14	0.0%
PharmaTrain	0	0.00	0.00	0.00	0.0%
PROactive	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)	**85**	**1.34**	**0.95**	**66.47**	**10.6%**

[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.

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

Project	Number of publications	Number of *Web of Science* publications	Total citations	Raw citation impact	% *Web of Science* publications in top quartile journals
EMTRAIN	1	0	0	0.00	0.0%
eTOX	26	23	82	3.09	65.2%
Eu2P	1	0	0	0.00	0.0%
EUROPAIN	34	31	214	6.58	80.6%
IMIDIA	15	13	58	4.15	100.0%
MARCAR	3	3	1	0.33	100.0%
NEWMEDS	28	26	91	3.31	88.5%
PharmaCog	14	11	25	2.09	72.7%
PharmaTrain	1	0	0	0.00	0.0%
PROactive	6	5	2	0.40	20.0%
PROTECT	15	14	15	1.07	71.4%
SafeSciMET	1	0	0	0.00	0.0%
SAFE-T	3	2	6	2.50	100.0%
SUMMIT	11	8	15	1.88	87.5%
U-BIOPRED	7	7	50	7.00	100.0%
Overall (IMI projects)	**210**	**184**	**623**	**3.39**	**82.1%**

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 for all three IMI funding calls are listed in Section 5.2 and Annex 3.

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

Figure 5.3.1 and Figure 5.3.2 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 5.3.1 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 5.3.1a shows that EUROPAIN and NEWMEDS account for the highest output among the group and have increased this output over the time period. As in the analyses of output trends at overall level (Section 4.3) there is a jump in activity between 2010 and 2011 as projects yield results for publication.

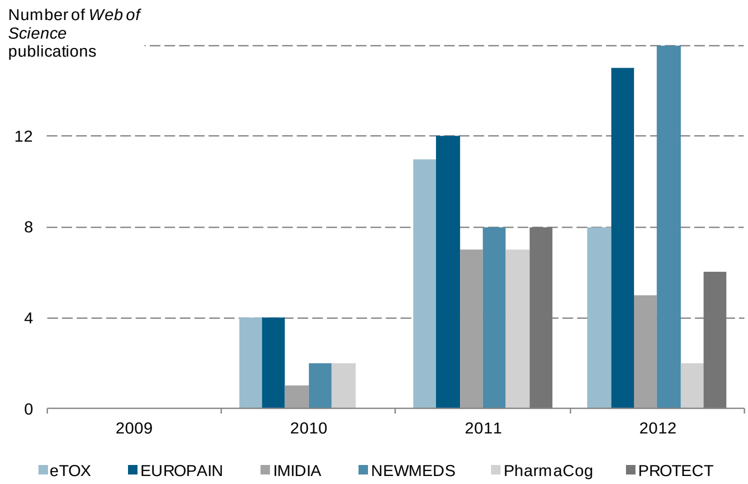


Figure 5.3.1b shows a 'spike' in the raw citation impact of EUROPAIN publications in 2010. As noted in Section 5.2 this is attributable to two highly-cited papers associated with this project, one of which has been awarded 'hot paper' status within Thomson Reuters databases (Annex 2).

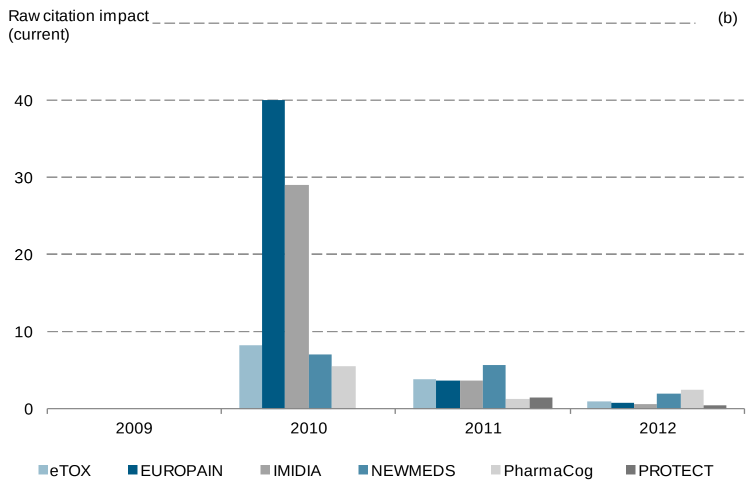


FIGURE 5.3.2 TRENDS IN (A) OUTPUT AND (B) RAW CITATION IMPACT FOR RESEARCH FROM IMI-SUPPORTED PROJECTS IN CALL 1: MARCAR, PROACTIVE, SAFE-T, SUMMIT AND U-BIOPRED.

Figure 5.3.2a shows that, of the projects in the second group, SUMMIT and U-BIOPRED account for the highest share of publications and have similar output in 2011 and 2012.

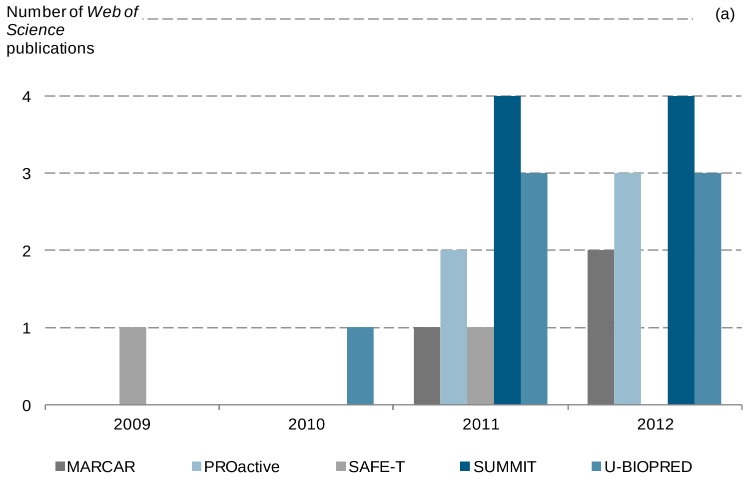
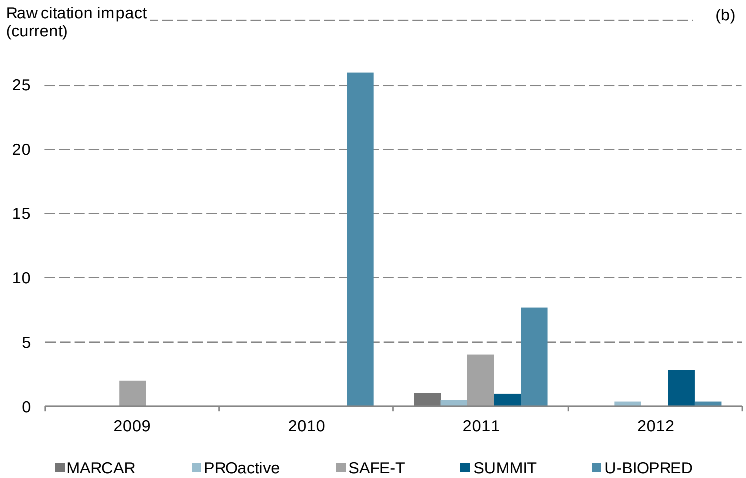


Figure 5.3.2b shows that U-BIOPRED has substantially higher raw citation impact than the other projects in the group with a sharp increase in this indicator for 2010 relative to other years. This increase is attributable to a single publication in 2010 that has accumulated 26 citations to date, 12 of which were made in 2012.



Four of the Call 1 projects (EMTRAIN, EU2P, PharmaTrain, and SafeSciMET) had no *Web of Science* publications. Although each of these projects had one publication associated with them, the journals in which the publications appear are not currently abstracted in *Web of Science*.

6 PATENT ANALYSIS – INNOVATION AND TECHONOLOGY TRANSFER

This Section of the report presents an analysis of patent publications and applications by IMI-supported researchers.

6.1 PATENTS ASSIGNED TO IMI-SUPPORTED RESEARCHERS

Patents assigned to 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 1 470 researchers associated with projects funded by the first IMI funding call (Call 1) in 2008.

Names of researchers associated with Call 1-funded projects were provided by IMI staff along with organisational affiliation. Initial searches for inventor names using the Derwents Index (DWPI) yielded a very large, complex dataset as some of these inventor names were common.

The dataset was subsequently filtered using assignee names and restricted by selecting medically-related patents.

Where possible each invention was associated with a specific IMI project. This association was made using the name of the inventor linked to the IMI researcher list for Call 1.

Inventions were identified of being of high IMI relevance by searching for specific terms in the invention abstract text. The abstract text on DWPI using controlled language which makes such searches more robust. Subsearching was used to identify the inventions with the highest interest to IMI using the following search terms: outcome* OR efficacy OR (clinical ADJ trial*) OR safety OR predict* OR computer* OR modeling OR (knowledge ADJ management) OR education OR training OR (pre ADJ clinical) OR preclinical. The results from this search were reviewed and inventions covering drug compositions or use in treatment were removed from the “high interest list” leaving the documents that were more oriented toward more general aspects of prediction and outcomes.

6.2 NUMBER OF PATENTS ASSOCIATED WITH IMI RESEARCHERS

In total 1 245 inventions were identified and associated with at least one IMI researcher (Table 6.2.1). The data highlight the volume of inventions associated with researchers funded by IMI. Table 6.2.2 shows the inventions identified as having high relevance to IMI. Around 10%, 116 of the 1 245 inventions associated with the IMI researchers were identified as being of high IMI relevance.

Although IMIDIA appears to have the greatest output, it is the NEWMEDS and EUROPAIN projects that are associated with the greatest number of inventions closely related to IMI.

TABLE 6.2.1 NUMBER OF PATENTS AT IMI PROJECT LEVEL

Project	Number of inventions	Project	Number of inventions
none associated	524	PharmaCog; NEWMEDS	7
IMIDIA	228	Marcar	6
EUROPAIN	111	EMTRAIN	4
NEWMEDS	88	EU2P	4
PharmaCog	54	Proactive	4
QUIC-Concept	39	PharmaTrain	3
Predect	31	NEWMEDS; EU2P	2
SAFE-T	31	OPENPHACTS	2
SUMMIT	24	BTCure; EUROPAIN	1

Project	Number of inventions	Project	Number of inventions
RAPP-ID	21	ONCOTRACK; PharmaCog	1
U-BIOPRED	17	QUIC-Concept; PharmaTrain	1
BTCure	15	RAPP-ID; PROactive	1
ONCOTRACK	14	SafeSciMET	1
SUMMIT; IMIDIA	11		

TABLE 6.2.2 PATENTS OF HIGH IMI RELEVANCE AT IMI PROJECT LEVEL

Project	High IMI relevance	Other	Total
none associated	42	482	532
NEWMEDS	28	60	80
EUROPAIN	9	102	111
QUIC-Concept	8	31	39
IMIDIA	7	221	228
SAFE-T	4	27	31
BTCure	3	12	15
PharmaCog	3	51	54
Predect	3	28	31
ONCOTRACK	2	12	14
SUMMIT	2	22	24
Emtrain	1	3	5
PharmaTrain	1	7	9
PROactive	1	3	4
RAPP-ID	1	20	21
U-BIOPRED	1	16	17
BTCure; EUROPAIN		1	1
EU2P		4	4
MARCAR		6	6
NEWMEDS; EU2P		2	2
ONCOTRACK; PharmaCog		1	1
OPENPHACTS		2	2
pharmacog; NEWMEDS		7	7
QUIC-Concept; PharmaTrain		1	1
RAPP-ID; PROactive		1	1
SafeSciMET		1	1
SUMMIT; IMIDIA		11	11
Total	**116**	**1 129**	**1 245**

Table 6.2.3 highlights the number of inventions held by each entity associated with IMI projects. The University of Cambridge appears to have the greatest activity followed by Roche and Sanofi-Aventis.

TABLE 6.2.3 PATENTING ENTITIES WITH AT LEAST ONE PATENT OF HIGH RELEVANCE TO IMI

Patent holder	High IMI relevance	Other	Total
CAMBRIDGE ENTERPRISE LIMITED / CAMBRIDGE UNIV	24	6	30
ROCHE	16	128	144
SANOFI-AVENTIS	9	162	171
SIEMENS	7	1	8
ABBOTT LABS	6	28	34
BOEHRINGER INGELHEIM	4	204	208
CNRS CENT NAT RECH SCI	4	4	8
PFIZER	3	48	51
NOVARTIS	3	63	66
BIOTRIN INTELLECTUAL PROPERTIES LIMITED	3		3
BAYER PHARMA / BAYER SCHERING	2	50	52
INSERM	2	15	17
UNIV LEIDEN	2	6	8
DECODE GENETICS EHF	2	3	5
INSTITUTE OF CANCER RESEARCH ROYAL CANCER HOSPITAL	2	1	3
IMPERIAL INNOVATIONS LIMITED	2	13	15
PSYNOVA NEUROTECH LIMITED	2		2
MAX PLANCK GES	2	5	7
UCL	1		1
VALTION TEKNILLINEN TUTKIMUSKESKUS	1		1
UNIVERSITÄT TÜBINGEN	1	2	3
COMMISSARIAT A L'ENERGIE ATOMIQUE	1	2	3
ROCHE / GENERAL ELECTRIC	1		1
EXONHIT THERAPEUTICS SA	1	4	5
UNIV MEDICAL CENT NIJMEGEN	1		1
HELMHOLTZ ZENTRUM MÜNCHEN	1	2	3
UNIVERSITY OF MANCHESTER	1	3	4
ANTARES PHARMA INC.	1		1
OPTIMATA LTD.	1		1
INNOVATIVE CONCEPTS IN DRUG DEVELOPMENT (ICDD)	1		1
TECHNISCHE UNIVERSITÄT DRESDEN	1	8	9
KATHOLIEKE UNIVERSITEIT LEUVEN	1		1
UNIV AARHUS	1		1
LUNDBECK AS	1	37	38
UNIVERSITÀ DEGLI STUDI DI PADOVA	1	1	2
MEDICAL RESEARCH FUND OF TEL AVIV	1	1	2
MERCK & CO INC	1	41	42
ASTRAZENECA	1	46	47

All 24 of the inventions with high IMI relevance associated with the University of Cambridge (and Cambridge Enterprise Ltd) can be linked to the NEWMEDS project. Within the University of Cambridge, Sabine Bahn appears to be a notable inventor associated with the NEWMEDS project (Table 6.2.4).

In contrast, none of the 9 inventions with high IMI relevance associated with Sanofi-Aventis have been associated with an IMI researcher and therefore cannot be linked to any IMI funding or project.

TABLE 6.2.4 PATENT HOLDERS OF AT LEAST ONE PATENT OF HIGH RELEVANCE TO IMI WITH ASSCOIATED PROJECTS

PATENT HOLDERS OF IMI RELEVANT PATENTS	**NEWMEDS**	**EUROPAIN**	**QUIC-Concept**	**IMIDIA**	**SAFE-T**	**Predect**	**BTCure**	**PharmaCog**	**ONCOTRACK**	**SUMMIT**	**PharmaTrain**	**Proactive**	**U-BIOPRED**	**EMTRAIN**	**RAPP-ID**	**none associated**	**Total**
CAMBRIDGE ENTERPRISE LIMITED / CAMBRIDGE UNIV	24																24
ROCHE				5		1										10	16
SANOFI-AVENTIS																9	9
SIEMENS			6													1	7
ABBOTT LABS		5														1	6
BOEHRINGER INGELHEIM		1														3	4
CNRS CENT NAT RECH SCI			1	2												1	4
BIOTRIN INTELLECTUAL PROPERTIES LIMITED					3												3
NOVARTIS								1							1	1	3
PFIZER		2											1				3
BAYER PHARMA / BAYER SCHERING									1							1	2
DECODE GENETICS EHF	1															1	2
IMPERIAL INNOVATIONS LIMITED																2	2
INSERM							1									1	2
INSTITUTE OF CANCER RESEARCH ROYAL CANCER HOSPITAL			1													1	2
MAX PLANCK GES									1							1	2
PSYNOVA NEUROTECH LIMITED	2																2
UCL		1														1	2
UNIV LEIDEN							2										2
ANTARES PHARMA INC.																1	1
COMMISSARIAT A L'ENERGIE ATOMIQUE																1	1
EXONHIT THERAPEUTICS SA								1									1
HELMHOLTZ ZENTRUM MÜNCHEN										1							1
INNOVATIVE CONCEPTS IN DRUG DEVELOPMENT (ICDD)																1	1
KATHOLIEKE UNIVERSITEIT LEUVEN												1					1
LUNDBECK AS	1																1
MEDICAL RESEARCH FUND OF TEL AVIV																1	1
MERCK & CO INC								1									1
OPTIMATA LTD.																1	1
ROCHE / GENERAL ELECTRIC																1	1
TECHNISCHE UNIVERSITÄT DRESDEN																1	1
UNIV AARHUS																1	1
UNIV GOTHENBURG											1						1
UNIV MEDICAL CENT NIJMEGEN						1											1
UNIVERSITÀ DEGLI STUDI DI PADOVA										1							1
UNIVERSITÄT TÜBINGEN					1												1
UNIVERSITY OF MANCHESTER														1			1
VALTION TEKNILLINEN TUTKIMUSKESKUS						1											1
Total	28	9	8	7	4	3	3	3	2	2	1	1	1	1	1	42	

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 1 470 researchers associated with projects funded by the first IMI funding call (Call 1) in 2008.

Names of researchers associated with Call 1-funded projects were provided by IMI staff 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 through August 2012. These matches were further reviewed and edited by IMI staff.

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 9 716 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 4.2.1.

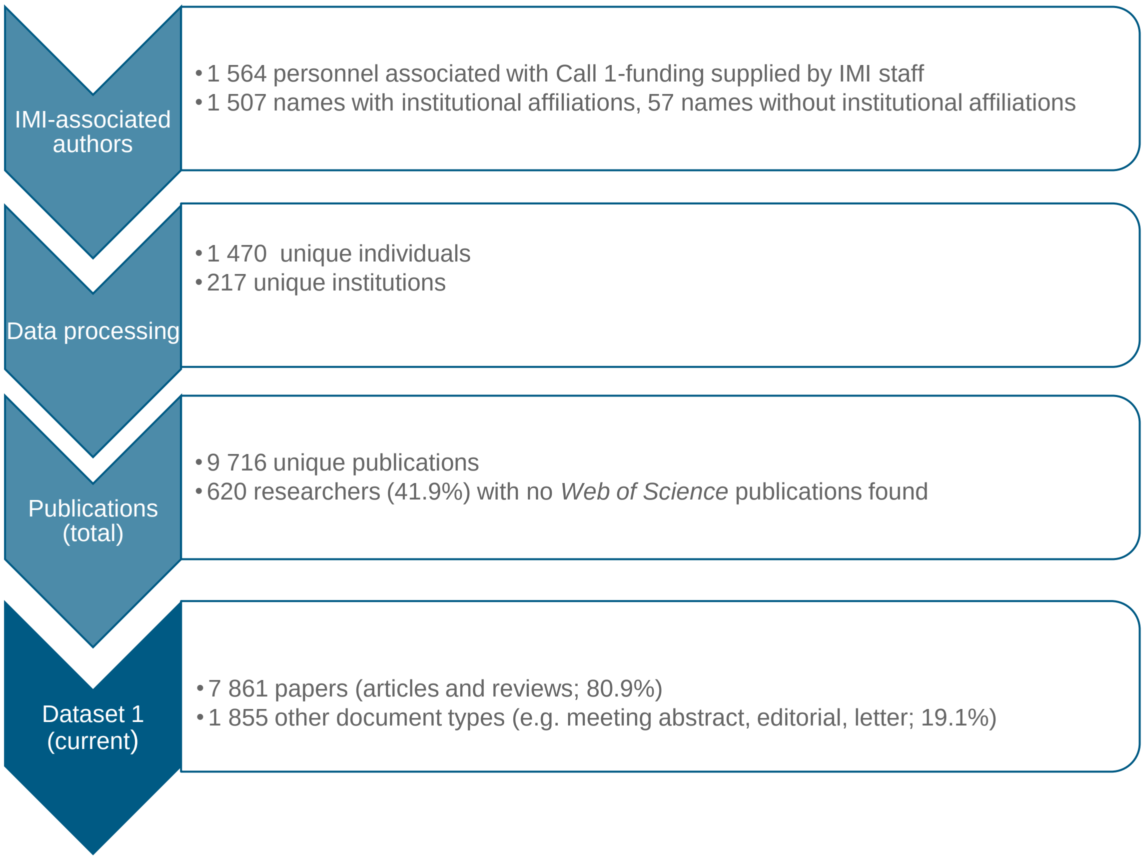
Citation counts for these 9 716 publications have been sourced from the citation databases which underlie Thomson Reuters *Web of Knowledge.* Counts have been extracted at two distinct census points– current (mid-August) and end-2011.

The former census point (Dataset 1) 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 evaluation of IMI research performance against the world average. A second set of citation counts is therefore needed (Dataset 2). These citation counts are taken at the same census point as used in the calculation of global citation baselines and are used to calculate the normalised citation impact of IMI research (see Section 3.1.3).

The analyses presented in this section will cover both raw and normalised citation impact data. It is important to note that all publication types with an expected citation rate[13] with have raw citation impact values while only papers are used in the calculation of normalised citation impact. Normalised bibliometric indicators for the papers in this report have been calculated using standard methodology and the Thomson Reuters *National Science Indicators* (NSI) database for 2011.

[13] 2012 publications will not have an expected citation rate until world citation baselines for this year are calculated.

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 is, not surprisingly, higher for IMI-supported researchers based in academic institutions and other research environments (Table 7.3.1).

The productivity of IMI-supported researchers working in Regulatory Agencies is very high, but this is likely to the due to small numbers of researchers based in that sector.

Overall, 58.5% of researchers had at least one publication.[14]

TABLE 7.3.1 PRODUCTIVITY: PUBLICATION OUTPUT, OVERALL AND BY SECTOR

Sector	Number of researchers		% researchers with publications
	With publications	**Total**	
Academic	385	561	68.6%
Corporate	292	542	53.9%
Patient Organisation	1	16	6.3%
Regulatory Agency	9	12	75.0%
Research (other)	124	210	59.0%
Small Medium Enterprise	31	75	41.3%
No assignment	18	54	33.3%
Total researchers	**860**	**1 470**	**58.5%**

Of the 54 researchers where sector could not be assigned, 5 have multiple different sectors associated with each individual (2 of these have publications). Sixteen have publications but are not associated with any sector because no affiliation was reported in the IMI data.

[14] This proportion, however, is statistically significantly different by sector ($p<0.0001$, Chi-Sq =71.0344 test of the equality of proportions, 6 degrees of freedom).

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.

IMI-supported researchers who are based in academic or in other research-active institutions have the strongest research performance.

Of the 385 publishing academic-based researchers, 88 researchers (22.9%) have published at least one ‘hot paper’ (defined in Section 3.1.3), 79 (19.7%) have an h-index of at least 10 and the majority have published most frequently in top quartile journals.

Similarly, researchers based in other research environments have published research which has performed well. More than 20 of these researchers have published a minimum of one ‘hot paper’, 21 researchers (16.9%) have h-index of at least 10 and most 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 most 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. This is also apparent for researchers who are assigned to the Small Medium Enterprise sector.

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	561	385	88	22.9%	76	19.7%	334	86.8%
Corporate	542	292	17	5.8%	7	2.4%	226	77.4%
Patient Organisation	16	1	0	0.0%	0	0.0%	1	100.0%
Regulatory Agency	12	9	1	11.1%	0	0.0%	5	55.6%
Research (other)	210	124	21	16.9%	21	16.9%	103	83.1%
Small Medium Enterprise	75	31	3	9.7%	1	3.2%	24	77.4%
No assignment	54	18	2	11.1%	1	5.6%	17	94.4%
Total researchers	**1 470**	**860**	**132**	**15.3%**	**106**	**12.3%**	**710**	**82.6%**

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 within 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	561	385	0.521	0.114	0.701	0.115	0.437	0.598	1.581	2.651
Corporate	542	292	0.508	0.144	0.621	0.148	0.365	0.652	0.743	1.424
Patient Organisation*	16	1	0.000	--	0.000	--	--	--	--	--
Regulatory Agency	12	9	0.506	0.096	0.577	0.125	0.325	0.396	0.782	1.091
Research (other)	210	124	0.513	0.121	0.683	0.117	0.424	0.466	1.701	2.740
Small Medium Enterprise	75	31	0.565	0.130	0.635	0.147	0.682	0.909	1.438	2.505
No Assignment	54	18	0.532	0.140	0.683	0.111	0.807	0.998	1.653	2.850
Total researchers	1 470	860	0.517	0.127	0.667	0.135	0.426	0.619	1.332	2.375

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, 860 researchers (58.5% of 1 470 IMI researchers in total) published any documents that were indexed in *Web of Science*. About three quarters of these researchers (N=660, 76.7% of 860) collaborated (co-authored) with at least one other IMI researcher during the period January 2007-August 2012.

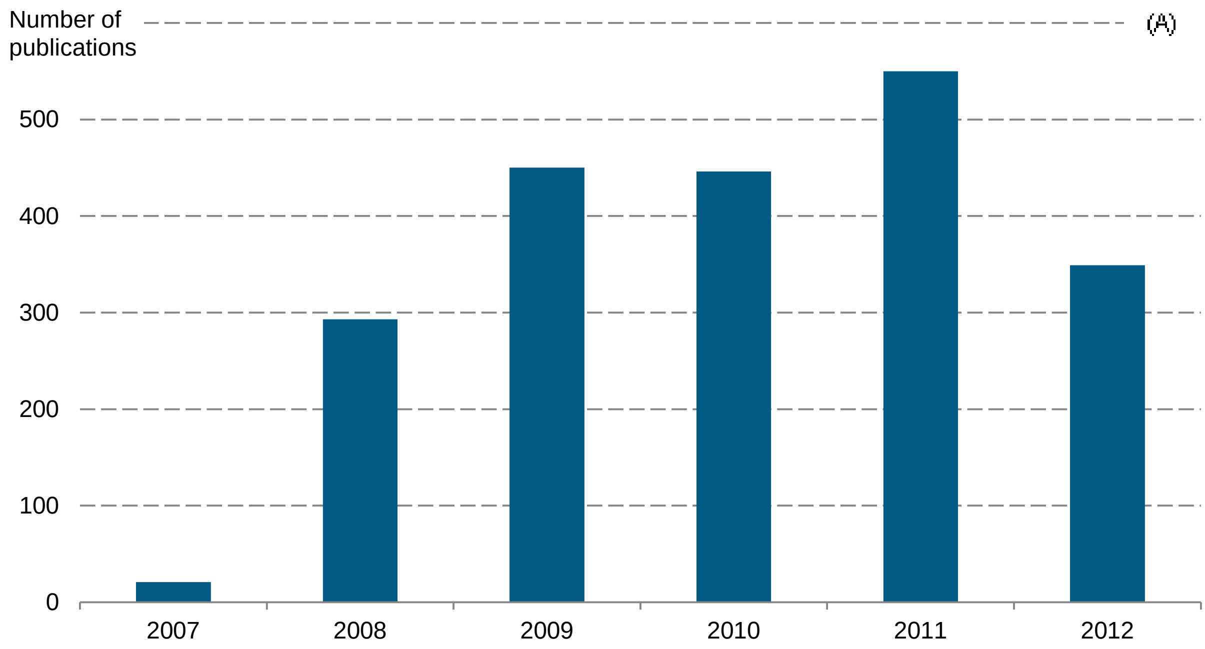
The frequency of collaborative activities are shown over the entire analysis period of January 2007 to August 2012 by year in Figure 7.5.1 and Table 7.5.1 and further illustrated among researchers in Figure 7.5.2.

TABLE 7.5.1 BREAKDOWN OF COLLABORATIVE ACTIVITY BY YEAR

	2007	2008	2009	2010	2011	2012
Publications	21	293	450	446	550	349
Within-Sector Collaborations	28	461	778	1 048	1 050	615
Cross-Sector Collaborations	3	105	408	605	450	228
% Cross-Sector	9.7%	18.6%	34.4%	36.6%	30.0%	27.0%

FIGURE 7.5.1 DISTRIBUTION OF COLLABORATIVE ACTIVITIES BY YEAR

(A) Collaborations defined as distinct researcher dyads within and across sectors appearing on one or more publications during the given publication year. (B) Number of publications co-authored by 2 or more IMI researchers by year



The number of individual researchers with collaborative activity has increased over time from 39 in 2007, to 262 in 2008, 359 in 2009, 365 in 2010, 418 in 2011, 348 in 2012 up to mid-August.

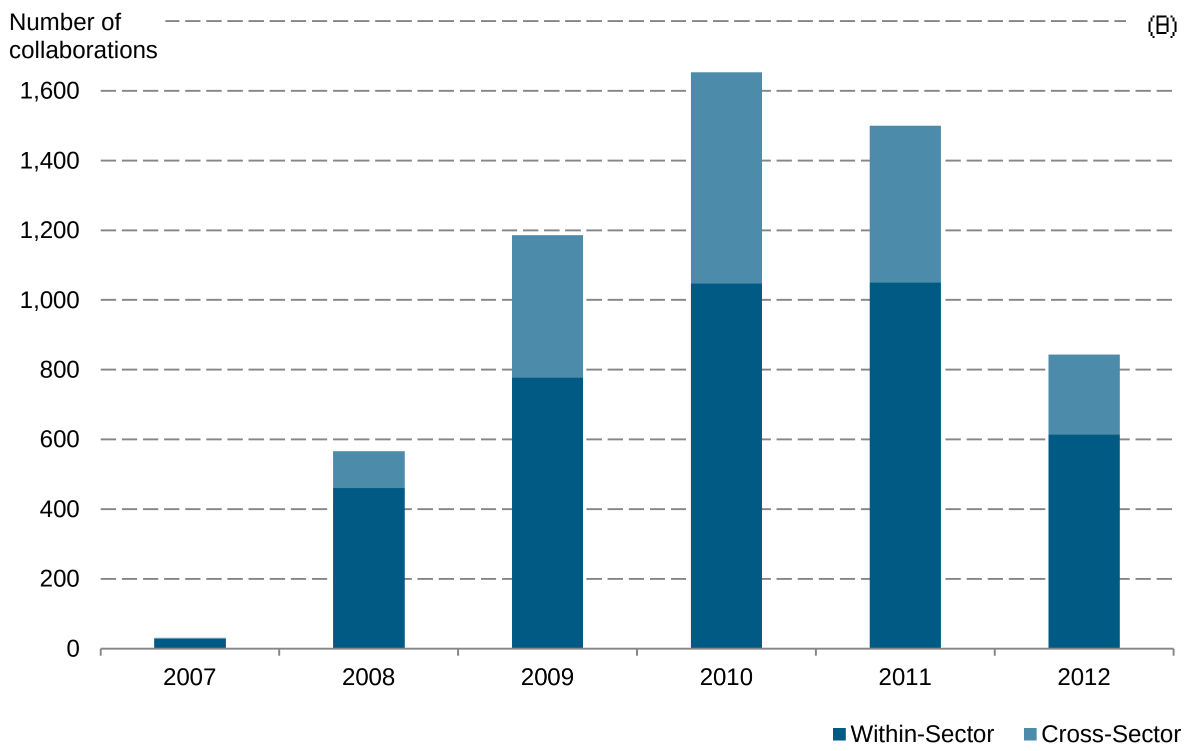
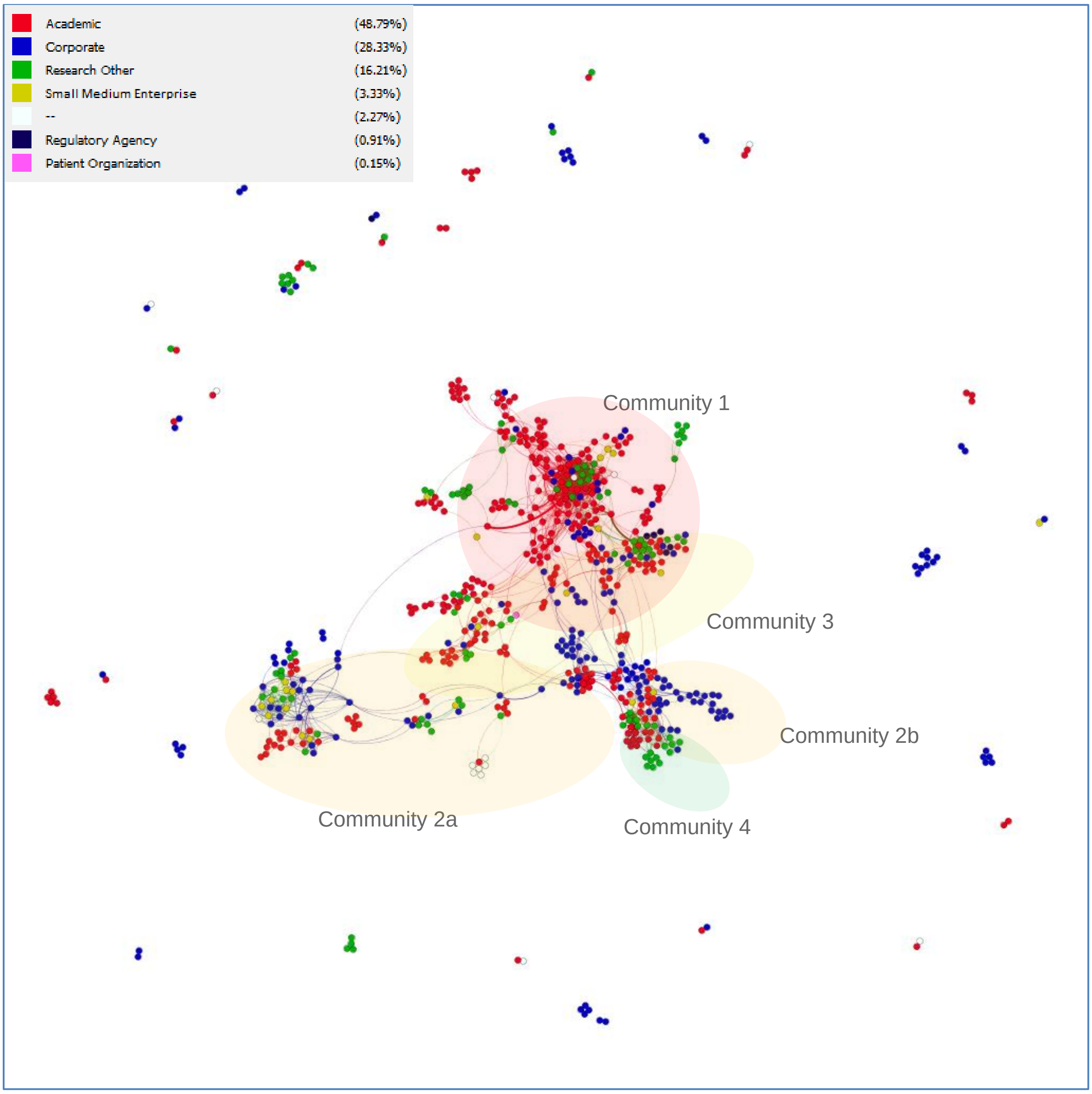


FIGURE 7.5.2 MAP OF 660 IMI PROJECT RESEARCHERS WHO HAVE CO-AUTHORED WITH AT LEAST ONE OTHER RESEARCHER WITHIN THE NETWORK BASED ON CO-AUTHORSHIP ACTIVITIES FROM JANUARY 2007 - AUGUST 2012.

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 8 communities of which the 4 largest are shown in shaded ovals. Graphics produced using Gephi, applying the Force Atlas 2 layout.[15] Communities identified using a resolution of 15.[16]



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

[16] 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 9 researchers each. The largest group of inter-connected researchers (N=554 researchers).

	Isolated Communities		Connected Communities		Connected Community							
					1		2		3		4	
Sector	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
No assignment	5	4.7	10	31.8	3	1.7	1	0.7	-	0.0	-	0.0
Academic	34	32.1	288	24.4	121	68.8	37	27.4	65	63.7	30	51.7
Corporate	49	46.2	138	18.4	29	16.5	63	46.7	8	7.8	7	12.1
Patient Organisation	1	0.9	1	10.5					1	1.0		
Regulatory Agency	1	15.1	5	6.1					5	4.9		
Research (other)	16	0.9	91	6.1	18	10.2	24	17.8	21	20.6	19	32.8
Small Medium Enterprise	1	4.7	21	1.6	5	2.8	10	7.4	2	2.0	2	3.4
Total	**106**		**554**		**176**		**135**		**102**		**58**	

These co-authorships appeared in 21.7% of publications matched to any IMI researcher (Table 7.6.1). The majority of these publications (83.3%) were articles or reviews while 12.3% were meeting abstracts.[17] However, all instances of co-authorship were treated the same regardless of the type of document in which it appeared.

The largest component, shown at the centre of Figure 7.6.1 and defined as groups of researchers where all individuals are connected with one another directly or indirectly via other IMI researchers, consisted of 554 researchers representing all six sectors (Table 7.6.1). Within this set of researchers, eight communities were identified within which there are more frequent and closely inter-related co-authorship activities. The largest four of these eight communities are shown enclosed by coloured oval. A complete depiction of all eight communities can be found in Figure 7.6.1.

The largest community, shown in red in Figure 7.6.1, 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 (68.8%, Table 7.6.1).

The co-authorship activities in group 2 (shown in orange in two ovals in the bottom left of Figure 7.6.1) are more loosely bound which suggest co-authorship among individuals, but not necessarily translation of those collaborative activities into further collaboration among their co-authors. Composition of this group reflects a distribution of researchers more similar to the distribution of IMI researchers overall (Table 7.6.1). While this demonstrates that the collaborative nature of the IMI projects does promote cross-sector co-authorship, the looser relationships in this group may partially be explained by the more broad representation across sectors.

While the majority of publishing researchers are connected to one another and are in the main connected component, 16.1% of publishing researchers (N=106 of 660) collaborate within isolated communities composed of between 2 and 9 researchers. Thirty-five isolated groups exist (shown on the periphery of Figure 7.6.1), of which 18 (51.4%) are composed of researchers from only one sector including the largest such group which includes 9 corporate authors.

This main component includes researchers from 127 distinct organisations, 22.8% (N=29) of which span across communities (Table 7.6.2). Within this set there are 13 academic organisations, 12 corporate organisations, and 4 other research organisations. The two entities which span the most communities are Astra Zeneca (corporate) and Imperial College London (academic). Overall, these organisational affiliations include 40.3% (N=233 or 554) of researchers in the main component.

Co-authorship is more common among researchers in the same sector than among researchers in different sectors (Figure 7.5.1B). This is expected given the principle of homophily which suggests that individuals are more likely to interact with individuals who are like them.[18] However, there are substantial co-authorship activities among researchers from different sectors (Figure 7.5.1B). Of a total 2 048 distinct co-authorship relationships, 816 are cross-sector and involve 343 total researchers from all 6 sectors. This accounts for 39.8% 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 399 of 2 048 of co-authorship relationships (19.5%) 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 – August 2012. 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.

[17] Other document types include biographical materials, editorials, letters, corrections.

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

Researchers with multiple affiliations and/or multiple sectors are not included in sector and affiliation analyses. That is, no choice was made between their affiliations and sectors to arrive at a single assignment for either data element.

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 in the main component are shown coloured by sector. (B) Researchers from the main component are coloured by community. (C) Researchers with any collaborative activity are shown coloured by disease area. (D) Researchers in the main component are shown coloured by disease area.

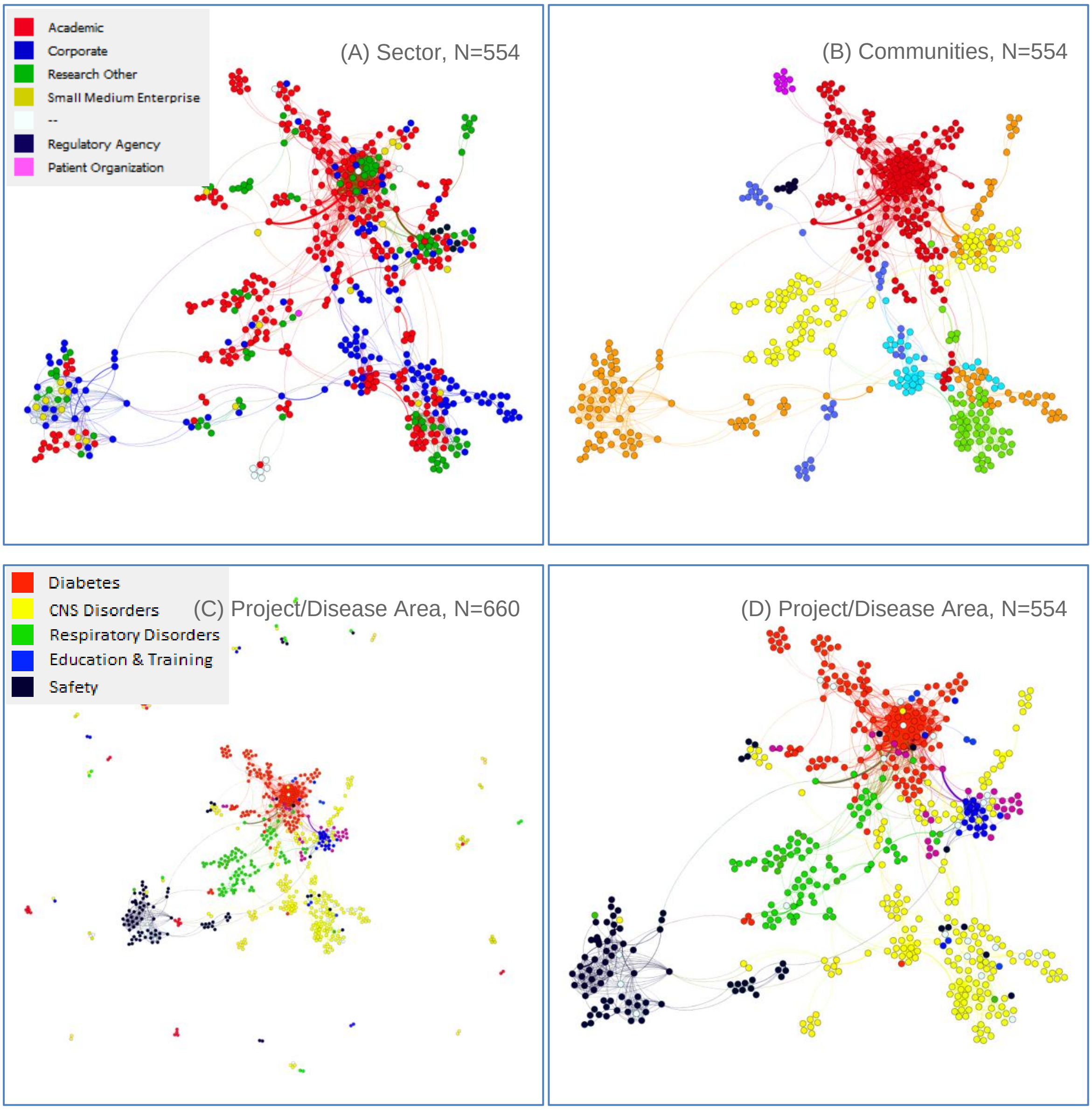


TABLE 7.6.2 ORGANISATIONS AND THE NUMBER OF ASSOCIATED COMMUNITIES AND RESEARCHERS WITHIN THE MAIN INTER-CONNECTED COMPONENT

In all, 127 distinct organisations were identified, of which 29 (22.8%) span communities.

Organisation	Sector	Number of communities	Number of researchers
AstraZeneca	Corporate	4	18
Imperial College London	Academic	3	15
Centre National de la Recherche Scientifique	Research (other)	3	7
GSK	Corporate	3	5
Karolinska Institutet	Academic	3	24
King`s College London	Academic	3	10
LUNDBECK	Corporate	3	13
Pfizer	Corporate	3	15
Roche	Corporate	3	11
Sanofi-Aventis	Corporate	3	6
Servier	Corporate	3	6
UCB Pharma	Corporate	3	4
University of Manchester	Academic	3	6
Boehringer Ingelheim	Corporate	2	2
Eli Lilly	Corporate	2	10
Fundació Institut Mar d'Investigacions Mèdiques IMIM	Research (other)	2	3
Haukeland University Hospital	Academic	2	3
Institut National de la Santé et de la Recherche Medicale	Research (other)	2	2
Laboratorios Almirall S.A	Corporate	2	3
Mario Negri Institute for Pharmacological Research	Research (other)	2	13
Medical University of Vienna	Academic	2	3
Novartis	Corporate	2	9
Universita Cattolica del Sacro Cuore	Academic	2	8
University of Aarhus	Academic	2	7
University of Cambridge	Academic	2	8
University of Copenhagen	Academic	2	2
University of Dundee	Academic	2	11
University of Southampton	Academic	2	6
Utrecht University	Academic	2	3

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

Anaesthesiology includes journals that deal with the administration of anaesthetics, the treatment of pain, and the use of life support systems. This category also covers specific journals on cardiovascular anaesthesia, paediatric anaesthesia, and neurosurgical anaesthesia.

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.

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 nine papers in the IMI project publications dataset that have been identified as highly-cited.

A2.1 HIGHLY-CITED PAPERS ASSOCIATED WITH IMI PROJECTS

(1) 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

(2) EUROPAIN: Aasvang, EK et al. (2010) Predictive Risk Factors for Persistent Postherniotomy Pain, Anesthesiology, 112: 957-969, doi: 10.1097/ALN.0b013e3181d31ff8

(3) 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

(4) 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

(5) 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

(6) 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

(7) eTOX: Taboureau, O et al. (2011) ChemProt: a disease chemical biology database, Nucleic Acids Research, 39: D367-D37210.1093/nar/gkq906

(8) 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

(9) 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

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

Section A2.2 lists the three papers from IMI projects that have been identified as 'hot papers'. The first of these papers is shared with the highly-cited papers dataset. Both the other papers were published much more recently and had not accumulated any citations at end-2011 when the percentile ranking used to define highly-cited papers is calculated.

A2.2 HOT PAPERS' ASSOCIATED WITH IMI PROJECTS

(1) 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

(2) 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

(3) 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

Papers with the highest interdisciplinarity have been defined as those with highest diffusion score as defined Carley and Porter (Section 3.1.3).[19]

Section A2.3 lists the five papers from IMI projects scoring highest on interdisciplinarity.

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

(1) 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

(2) 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

(3) 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

(4) 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

(5) 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

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

ANNEX 3: SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS – CALLS 2 AND 3

TABLE A3.1 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 2 – CITATIONS TO CURRENT

Project	Number of publications	Number of *Web of Science* publications	Total citations	Raw citation impact	% *Web of Science* publications in top quartile journals
BTCure	15	14	16	1.14	92.9%
DDMoRe	1	1	1	1.00	100.0%
EHR4CR	1	0	0	0.00	0.0%
OncoTrack	9	9	17	1.89	77.8%
Open PHACTS	8	6	17	2.83	83.3%
QuIC-ConCePT	6	6	6	1.00	100.0%
RAPP-ID	1	1	0	0.00	100.0%

TABLE A3.2 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[20]
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%
OncoTrack	3	0.51	0.22	74.73	0.0%
Open PHACTS	1	0.00	0.00	100.00	0.0%
QuIC-ConCePT	0	0.00	0.00	0.00	0.0%
RAPP-ID	0	0.00	0.00	0.00	0.0%

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

None of the *Web of Science* publications associated with the DDMoRe, QuIC-ConCePT or RAPP-ID projects were classified as papers (articles and reviews).

None of the papers associated with BTCure or Open PHACTS had been cited at end-2011. No papers from Call 2 projects are, as yet, highly-cited.

[20] '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 A3.3 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 3 – CITATIONS TO CURRENT

Project	Number of publications	Number of *Web of Science* publications	Total citations	Raw citation impact	% *Web of Science* publications in top quartile journals
BioVacSafe	1	1	0	0.00	100.0%
EU-AIMS	5	2	6	3.00	100.0%
MIP-DILI	1	1	0	0.00	100.0%

TABLE A3.4 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[21]
BioVacSafe	0	0	0	0	0.0%
EU-AIMS	0	0	0	0	0.0%
MIP-DILI	0	0	0	0	0.0%

EU-AIMS is the only one of Call 3 project currently with cited *Web of Science* publications.

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.

[21] '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).[22]

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.

[22] 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, our 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.

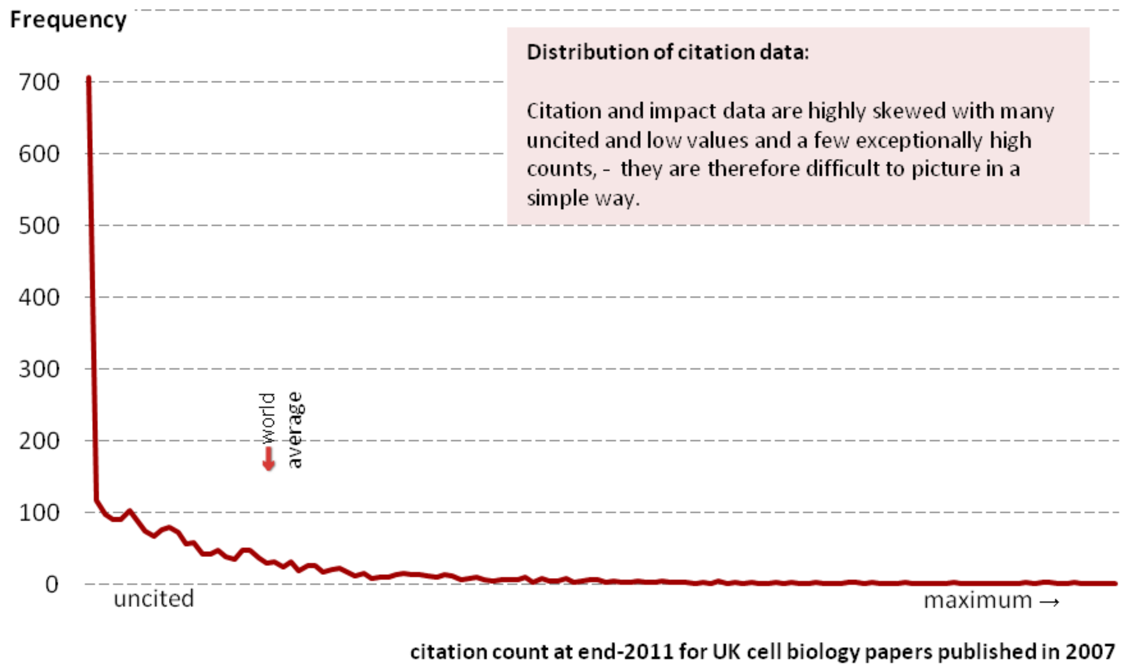
A4.4 CITATION COUNTS

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.

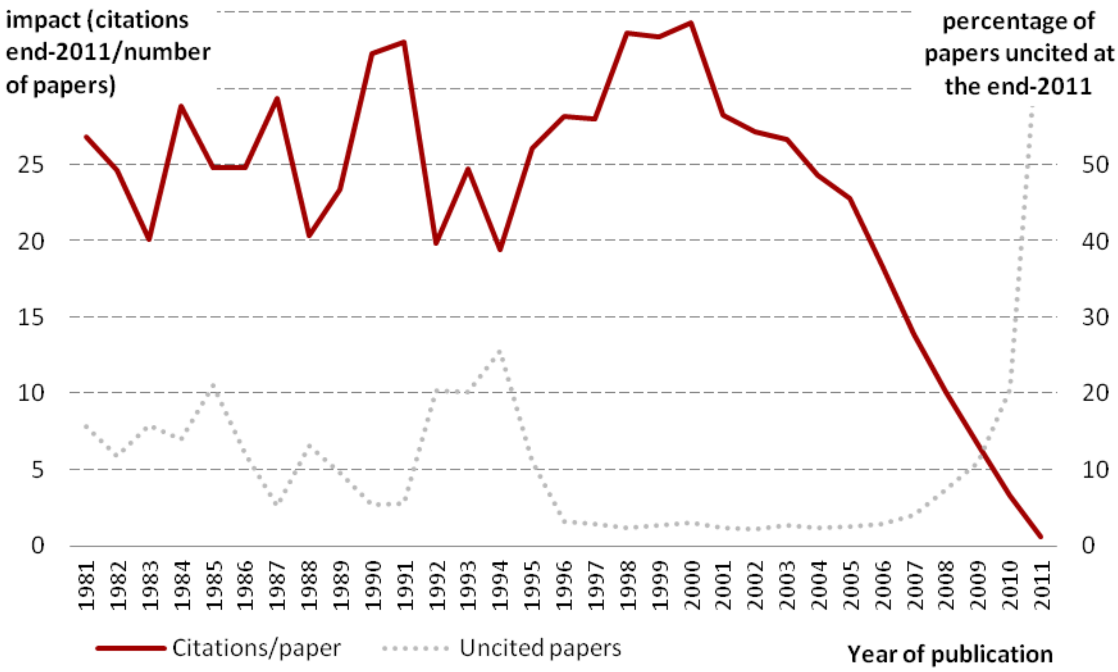
We do not seek to account separately for the effect of self-citation. If the citation count is significantly affected by self-citation then the paper is likely to have been infrequently cited. This is therefore only of consequence for low impact activity. Studies show that for large samples at national 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/.

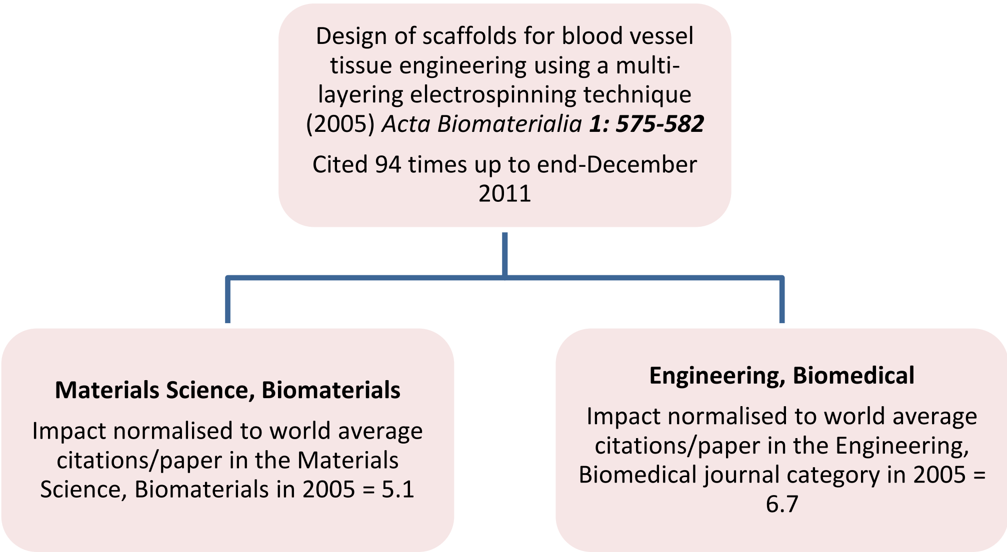
Some journals with a very diverse content, including the prestigious journals *Nature* and *Science* were classified as ***Multidisciplinary*** in databases created prior to 2007. The papers from these ***Multidisciplinary*** journals are now re-assigned to more specific research fields using an algorithm 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[23] 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.

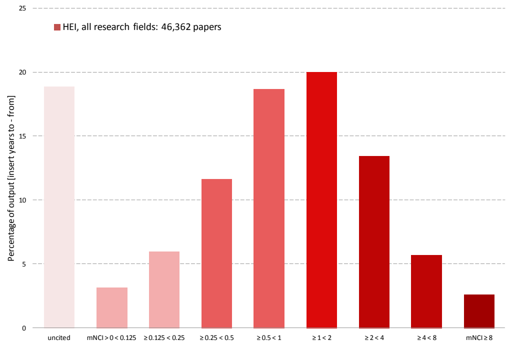
Papers which are "highly-cited" are often defined in our reports as those with an average citation impact (NCI_F) greater than or equal to 4.0, i.e. those papers which have received greater than or equal to four times the world average number of citations for papers in that subject published in that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

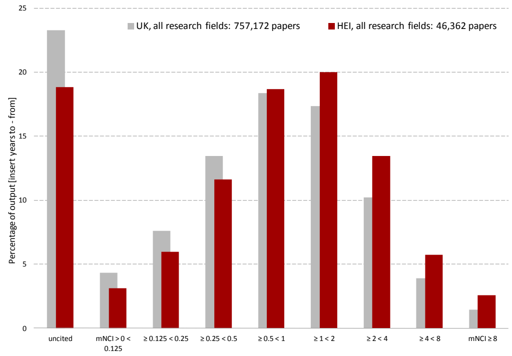
[23] Adams J, Gurney K & Marshall S (2007) Profiling citation impact: A new methodology. *Scientometrics* **72**: 325-344.

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

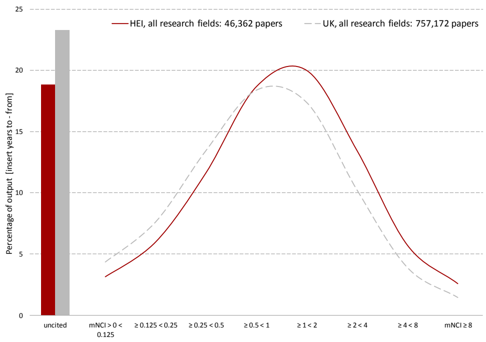
A



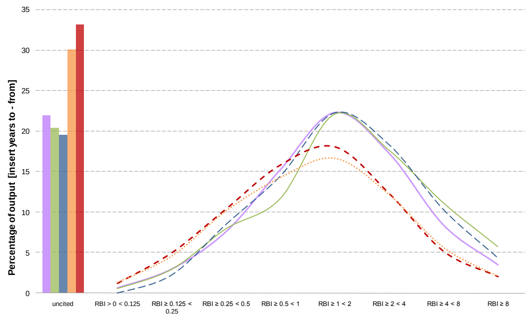
B



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 58 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
Agricultural economics & policy
Agricultural engineering
Agriculture, dairy & animal science
Agriculture, multidisciplinary
Agriculture, soil science
Agronomy
Allergy
Anatomy & morphology
Andrology
Anesthesiology
Anthropology
Applied linguistics
Archaeology
Architecture
Area studies
Art
Asian studies
Astronomy & astrophysics
Automation & control systems
Behavioral sciences
Biochemical research methods
Biochemistry & molecular biology
Biodiversity conservation
Biology
Biology, miscellaneous
Biophysics
Biotechnology & applied microbiology
Business
Business, finance
Cardiac & cardiovascular systems
Cell biology
Chemistry, analytical
Chemistry, applied
Chemistry, inorganic & nuclear
Chemistry, medicinal
Chemistry, multidisciplinary
Chemistry, organic
Chemistry, physical
International relations

Classics
Clinical neurology
Communication
Computer science, artificial intelligence
Computer science, cybernetics
Computer science, hardware & architecture
Computer science, information systems
Computer science, interdisciplinary applications
Computer science, software engineering
Computer science, theory & methods
Construction & building technology
Criminology & penology
Critical care medicine
Crystallography
Dance
Demography
Dentistry, oral surgery & medicine
Dermatology
Developmental biology
Ecology
Economics
Education & educational research
Education, scientific disciplines
Education, special
Electrochemistry
Emergency medicine
Endocrinology & metabolism
Energy & fuels
Engineering, aerospace
Engineering, biomedical
Engineering, chemical
Engineering, civil
Engineering, electrical & electronic
Engineering, environmental
Engineering, geological
Engineering, industrial
Engineering, manufacturing
Engineering, marine
Engineering, mechanical
Mining & mineral processing

Engineering, multidisciplinary
Engineering, ocean
Engineering, petroleum
Entomology
Environmental sciences
Environmental studies
Ergonomics
Ethics
Ethnic studies
Evolutionary biology
Family studies
Film, radio, television
Fisheries
Folklore
Food science & technology
Forestry
Gastroenterology & hepatology
Genetics & heredity
Geochemistry & geophysics
Geography
Geography, physical
Geology
Geosciences, multidisciplinary
Geriatrics & gerontology
Health care sciences & services
Health policy & services
Hematology
History
History & philosophy of science
History of social sciences
Horticulture
Humanities, multidisciplinary
Imaging science & photographic technology
Immunology
Industrial relations & labor
Infectious diseases
Information & library science
Instruments & instrumentation
Integrative & complementary medicine
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