Innovative Medicines Initiative

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

IMI EXECUTIVE OFFICE

Third report: October 2013

1 EXECUTIVE SUMMARY

This report presents a bibliometric analysis of IMI research associated with funding Calls 1 to 4, using citations as an index of research quality and co-authorship as an index of collaboration. The analyses use two sets of research publications, publications from IMI projects (IMI project research) and publications from IMI-supported researchers (IMI researchers).

The overall volume of IMI project research has increased rapidly since 2009 and the initiative continues to show rapid growth. This is partly to be expected as the number of funded projects rises and those projects funded earliest in the program begin to publish. To date, IMI projects have produced 483 publications, over one-third of which have been published in the last six months – a similar expansion to that observed in second report. This research has been published in over 200 journals including JAMA, Science and Nature Publishing Group titles.

The volume of IMI research has also increased at the level of individual projects. BTCure (Call 2) is the most prolific project in any of the four funding Calls with output surpassing EUROPAIN and NEWMEDS (the most prolific projects funded in Call 1) and over 40 publications in 2013 alone. Among more recent projects, EU-AIMS (Call 3) has shown substantial growth in output. Of the 30 projects funded in Calls 1, 2 and 3, all but EUPATI have published at least once to date. Though Call 4 projects were only funded in late 2012, one project (ORBITO) has already generated published research.

IMI project research is wide-ranging – the research portfolio from IMI projects covers diverse research fields. IMI project research has been published most frequently in Rheumatology, Pharmacology & Pharmacy and Endocrinology & Metabolism journals. IMI project research published in Psychiatry journals has become more common since the second report to IMI.

IMI project research is well-regarded – the quality of IMI project research (as indexed by citation impact) has not only been maintained but has increased while output has grown. The citation impact of this research is, at twice world average, internationally influential. Around one-fifth of papers from IMI projects are 'highly-cited', that is, they belong to the world's top ten percent of papers in that journal category and year of publication, when ranked by number of citations received. This is a substantial increase since the second report.

Relative to established funding organisations IMI project research performs well – trends in the quality of this research are generally comparable to the Wellcome Trust, a leading UK funder of biomedical research.

Researchers funded by IMI are well-regarded by their peers. The total research published by IMI researchers funded in Calls 1 to 4 (as opposed to that directly associated with IMI funding) is well cited with an average citation impact over twice world average and about one-fifth of papers being 'highly-cited'. This is similar performance to IMI project research indicating that IMI funds both research and individuals performing at a high overall level.

Researchers and projects funded by IMI are highly collaborative. About two-fifths of all publications by IMI researchers were cross-sector (for example, between academic institutions and small medium enterprises - SMEs) and over half of all papers from IMI projects were cross-sector.

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

Summary of key findings – IMI project research

Currently there are 40 IMI projects, of which 17 were launched since 1 January 2012. It may take several years for a project to progress from inception to the point where it has generated sufficient data for a publication. It may take further years until it has produced its most valuable results. The IMI projects that are analysed here are still relatively young, and early bibliometric indicators may not fully reflect their eventual impact.

- A total of 483 publications from IMI-supported projects were identified for inclusion in this report on IMI research activity up to mid-August 2013. Some 453 of these publications were linked to Thomson Reuters citation data and over 95% of these documents were substantive articles and reviews (Section 4.1, Section 4.2).

- Over one-third (34.2%) of the IMI project research portfolio was published in the last six months (Section 4.1, Annex 1).

- IMI project research continues to be published in highly-regarded journals (Science, Nature, JAMA and PNAS) with output in PNAS doubling since the second report. The core set of journals used by IMI projects continues to highlight the diversity of IMI-supported research with titles focused on bioinformatics, genetics and psychiatry as well as disease areas such as arthritis and diabetes (Section 4.4, Annex 1).

- IMI project research is most frequently published in Rheumatology journals and in Pharmacology & Pharmacy journals with output in Rheumatology (primarily from BTCure) more than doubling since the last report. Output in Biochemistry & Molecular Biology, Genetics & Heredity and Immunology journals has also expanded since the second report and these fields are now among the top ten most frequently used (Section 4.5).

- The average citation impact for IMI project research is 2.04 for the 3-year period, 2010-2012, (where world average is 1.0) and just under one-fifth (19.3%) of papers are highly-cited. For comparison, the EU's average citation impact relative to world baselines for the same 2-year period in similar research fields was 1.13. This is a substantial increase over the second report (average citation impact, IMI project research = 1.55) and highlights the rapid uptake of research published by IMI projects – more than half of IMI's highly-cited papers were published in 2012 (Section 4.6).

- IMI project research shows year-on-year growth and has comparable citation impact to research acknowledging Wellcome Trust funding; research from both funders is internationally influential with citation impact twice world average and just under one-fifth of papers being highly-cited (Section 4.7).

- IMI project research is collaborative at sector, institution and country level. Well over half (61.3%) of all IMI project papers have been published by researchers affiliated with different sectors and half (49.8%) of all IMI project papers have an international co-author (Section 4.8).

- Output has increased for all three IMI funding Calls, though Call 1 continues to account for the highest share of IMI research. EUROPAIN and NEWMEDS continue to be the most prolific projects funded in Call 1, however, output from the Call 2 project, BTCure, has more than doubled in the last six months, overtaking both EUROPAIN and NEWMEDS. Two Call 3 projects, ABIRISK and PreDiCT-TB did not have output in the second report but have accumulated publications in the last six months (Section 5).

Summary of key findings – IMI researchers

The productivity, research performance and collaboration of researchers funded by IMI through Calls 1 to 4 were assessed by analysing their total publication output (not limited to publications acknowledging IMI funding). Some 4 603 researchers were included in the analysis and 23 915 of their publications were identified for the period January 2007-August 2013 (Section 6.1).

- Publication output, as previously, is higher for IMI-supported researchers based in academic institutions and other research environments compared to industry and SMEs (Section 6.3).

- Since the second report, researchers associated with patient organisations have published well-regarded research and continue to show the strongest research performance. Six (30.0%) have published at least one 'hot' paper, 5 (25.0%) have an h-index of at least 10 and 18 (90.0%) have published exclusively in top-quartile journals (Section 6.4).

- Collaboration analysis was performed on the basis of co-authorship between IMI-supported researchers. Over two-thirds (68.9%) of IMI researchers collaborated (co-authored) with at least one other IMI researcher during the period of analysis (Section 6.5).

- Again, co-authorship is more common between researchers in the same sector than among researchers in different sectors. Cross-sector co-authorship accounts for around two-fifths (41.7%) of all co-authorship activities during the analysis period (Section 6.6).

- The same is true of co-authorship activities by project. The majority of collaborative relationships are among researchers associated with the same project with only around one-third (36.9%) being cross-project. The share of cross-project activity between researchers has, however, increased relative to the second report (Section 6.6).

- Geographical maps of collaboration among IMI researchers comparing activity before and after IMI funding award show that there is a small increase in the share of cross-sector output after IMI funding. Though data are presented cumulatively, these analyses have used a 'staggered' time window in which the pre-IMI and post-IMI periods are defined for each project based on funding award date (Section 6.7).

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 third evaluation in the series. Since the number of applications and awards specifically generated by IMI projects to date is small, IMI personnel have advised that patent analyses are not required for this third evaluation.

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 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 nine Calls for proposals to be funded under the initiative. The first funding call was announced in 2008 and the latest, 9th, funding call was launched on 7th July 2013.

This report will cover the research outputs (publications and papers) from the first four calls which have resulted in 37 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 15 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 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.

Outline of report

- Section 3 describes the data sources and methodology used in this report along with definitions of the indicators and guidelines to interpretation.

- Sections 4 and 5 present citation analyses of IMI project publications overall (Annex 1 provides summary analyses of IMI project publications identified since the last report to IMI).

- Section 6 presents bibliometric indicators for IMI-supported researchers and analyses of collaboration between these individuals.

- Previous reports (October 2012 and February 2013) contained patent data and analyses for IMI research. This component is not required by IMI at present but could be taken up again in future reports if IMI wish to revisit this option.

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 Knowledgeℳ*, 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 all data and search terms together and therefore provides a level of comparability not found in other

---

[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 .

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 August 2013 with citation counts as at mid-August for all 'current' indicators and citation counts as at end-2012 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 2008 will typically have more citations than papers published in 2011).

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, a researcher who has published 30 papers, of which 17 have received 17 or more citations while the remaining 13 have received fewer than 17 citations, has an h-index of 17. Irrespective of research impact, older researchers in more prolific fields tend to have a higher h-index.

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

Interdisciplinarity/diffusion score: This is indicated by the number and disparateness of the fields from which publications citing an IMI publication originate, summarised in a diffusion score developed by Carley and Porter.[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 115 times to end-December 2012, the expected citation rate for a publication in Acta Biomaterialia published in 2005 would be 28.7 and the NCI$_J$ would be 4.01. 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 2012 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 (2010 and 2011) were cited in 2012. Thus, a JIF of 2.0 means that, on average, the articles published in 2010 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, *Fertility and Sterility*, the 2012 journal Impact Factor would be calculated as follows:

Cites in 2012 to items published in 2010 =   3 259   Number of items published in 2010 =   744

---

[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 2012 to items published in 2011 =   2 556   Number of items published in 2011 =   649
Total   **5 815**   **1 393**

$$\frac{Number\ of\ citations}{Number\ of\ \text{items}} = \frac{5\,815}{1\,393} = \textbf{4.174}$$

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 4.174 ranks the journal *Fertility and Sterility* 4th out of 77 journals in the Obstetrics & Gynaecology journal category and therefore in the top quartile. However, the journal *Journal of Alzheimer's Disease* with the same JIF of 4.174 is ranked in the second quartile (64th out of 251 journals) in the journal category Neurosciences.

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

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

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

For clarity, in this report:

- **Publication** is used inclusively to refer to all IMI publications whether linked to Thomson Reuters citation data or not.

- *Web of Science* **publication** is used exclusively to refer to those IMI publications which have been linked to Thomson Reuters citation data.

- **Paper** is used exclusively to refer only to substantive *Web of Science* publications (journal articles, reviews and some proceedings papers) that have been linked to Thomson Reuters citation data. This definition excludes editorials, meeting abstracts or other types of publication. Papers are the subset of publications for which citation data are most informative and which are used in calculations of citation impact.

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

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

3.1.4 INTERPRETATION OF BIBLIOMETRIC INDICATORS AND CITATION ANALYSES

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

- IMI JU only started to fund projects in May 2009. Of the 40 active projects, 17 were launched since 1 January 2012. 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.

- 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 many 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 example, 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 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

---

[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

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

IMI researcher publications/papers: This dataset comprises publications by IMI-supported researchers as described in Section 6 and outlined in Figure 6.2.1.

Similar European research: this benchmark dataset has been created using the EU-27 grouping of countries: Thomson Reuters *National Science Indicators* 2012 database and only research falling into the same journal categories as in the IMI project dataset.

Wellcome Trust publications/papers: this benchmark dataset has been created using specific keyword searches on funding acknowledgment data in Thomson Reuters *Web of Science* to define those publications where the Wellcome Trust has been acknowledged as a funder. This is the same process by which IMI project publications have been identified.

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 6.1). IMI project research is also benchmarked against similar European research (see footnote on page 25) and research associated with Wellcome Trust funding (Section 4.7).

Publications for analyses include all IMI-supported publications identified in Thomson Reuters *Web of Science℠* to date – that is, publications new to this report (Annex 1) as well as publications identified in the previous reports. The census point for inclusion of publications in the second report to IMI was end-January 2013. The census point for inclusion of publications in this third report was mid-August 2013. The third report therefore reflects changes in IMI activity between these points.

Citation counts for all publications included previously have been updated from the end-January census point used in the second report.

When considering the analyses in this Section, earlier caveats regarding paper numbers should be borne in mind (Section 3.1.4).

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 *Web of Science*. The number of funding calls for analysis has been extended from the second report to IMI and includes projects from Calls 4 to 6 as well as Calls 1 to 3.

The aggregated list of publications was reviewed by Thomson Reuters (*Evidence*) and supplied to IMI for further verification prior to inclusion in the analyses. Ten publications have not been assigned to specific projects despite review by IMI personnel. Only one publication was found for Call 4 and no publications were found for projects in Calls 5 to 6. As these projects were only awarded funding recently (late 2012 or early 2013) and it may take years for a project to progress towards published research this should not be taken as evidence of low productivity.

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

Two datasets are used in the citation analyses:

- Dataset 5: This is an expansion of Dataset 3 in the second report and comprises 453 *Web of Science* publications. Of these 453, 165 are new to this report. All citation counts are at mid-August 2013.
- Dataset 6: This is an expansion of Dataset 4 in the second report and comprises 279 *Web of Science* publications. All citation counts are at end-2012.
    - This dataset has changed relative to Dataset 4 in the second report as the normalised citation impact of papers at end-2012 can now be calculated (world citation baselines for 2012 used to calculate normalised citation impact were not available at the time of the second report) and 2012 papers can now be included.

To benchmark IMI project research performance, these cumulative datasets are also compared to publications collated for IMI-supported researchers (Section 6.1), similar European research (see footnote on page 25) and research acknowledging Wellcome Trust funding (Section 4.7).

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

IMI-identified publications
- 24 publication records supplied by IMI personnel
- 3 publications overlapping with Report 2 dataset
- 21 publications for addition to dataset

Web of Science
- 156 publications identified as IMI-associated through *Web of Science* funding text

Publications (total)
- 483 IMI publications overall
- 319 included in Report 2 (298 linked to Thomson Reuters citation databases)
- 164 newly identified for Report 3 (155 linked to Thomson Reuters citation databases)

Dataset 5 (current)
- 453 unique *Web of Science* publications linked to Thomson Reuters citation databases (33 publications in journals not abstracted by *Web of Science* or e-publications ahead of print, 3 publications assigned to more than one IMI project)
- 442 papers (articles and reviews; 97.6%); 11 other document types (e.g. meeting abstract, editorial, letter; 2.4%)

Dataset 6 (end-2012)
- 279 of the 453 *Web of Science* publications were published before 2013 and so have citation data at end-2012
- 269 papers (articles and reviews; 96.4%); 10 other document types (e.g. meeting abstract, editorial, letter; 3.6%)

Datasets 5 and 6 are expansions of Datasets 3 and 4 used in the second report, as described above. Dataset 6 contains fewer papers than Dataset 5 as the citation impact of 2013 papers cannot yet be calculated (the world citation baselines for 2013 are not yet available).

World citation baselines for 2013 will be available in February/March 2014. Normalised citation impact indicators for IMI-supported papers would therefore be updated in the fourth report to IMI, depending on the timing of that report.

4.1.1   CITATION DATA FOR PUBLICATIONS FROM IMI-SUPPORTED PROJECTS

A total of 483 publications resulting from IMI-supported projects were identified and 453 of these publications linked to records in *Web of Science*. Citation counts have been sourced from the citation databases which underlie Thomson Reuters *Web of Knowledge* and have been extracted at two distinct census points – current (mid-August 2013) and end-2012.

The 'current' census point (Dataset 5) allows assessment of the performance of IMI research from as up-to-date a viewpoint as possible through calculation of 'raw' citation impact (see Section 3.1.3). This, however, does not allow benchmarking of IMI research performance against the world and European average.

Dataset 6 is used to evaluate the citation impact of IMI-supported research relative to world average (normalised citation impact) and has same end-2012 census point used in the calculation of global citation baselines (see Section 3.1.3).

Normalised bibliometric indicators for Dataset 6 were calculated using standard methodology and the Thomson Reuters *National Science Indicators* (NSI) database for 2012.

4.2   SHARE OF PAPERS RELATIVE TO OTHER PUBLICATION TYPES

FIGURE 4.2.1 CATEGORISATION OF IMI PROJECT RESEARCH BY DOCUMENT TYPE

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

IMI project research comprises 453 unique *Web of Science* publications linked to citation data (as outlined in Figure 4.1.1). Over 95% of these documents (97.6%) were substantive articles and reviews with only 11 documents not falling into this grouping. These documents (classified as 'Other') include six editorials, two meeting abstracts, two letters and one news item.

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

Review 19.4%

Article 78.1%

Other 2.4%
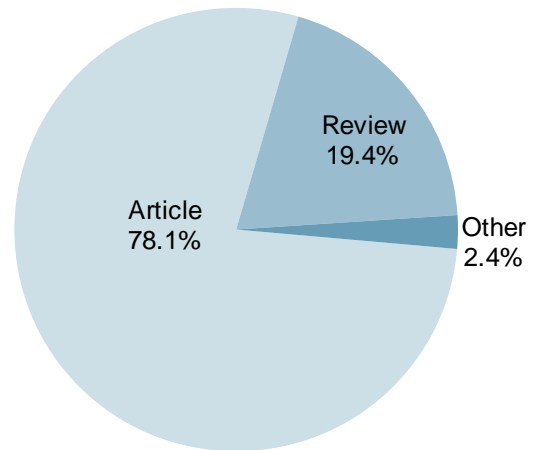
4.3   TRENDS IN PUBLICATION OUTPUT

Figure 4.3.1 presents the annual numbers of *Web of Science* publications in this third report to IMI.

- IMI project research continues to show substantial growth with just over one-third (34.2%) of publications new to this report – a similar expansion to that observed between the first and second reports.

- IMI projects have generated more than 200 publications in 2012 and growth looks set to continue in 2013 with some 150 publications to date.

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

Number of *Web of Science* publications

Figure 4.3.2 shows the proportion of papers (articles and reviews) relative to other document types for IMI project research from 2010 to date.[5]

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

---

[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

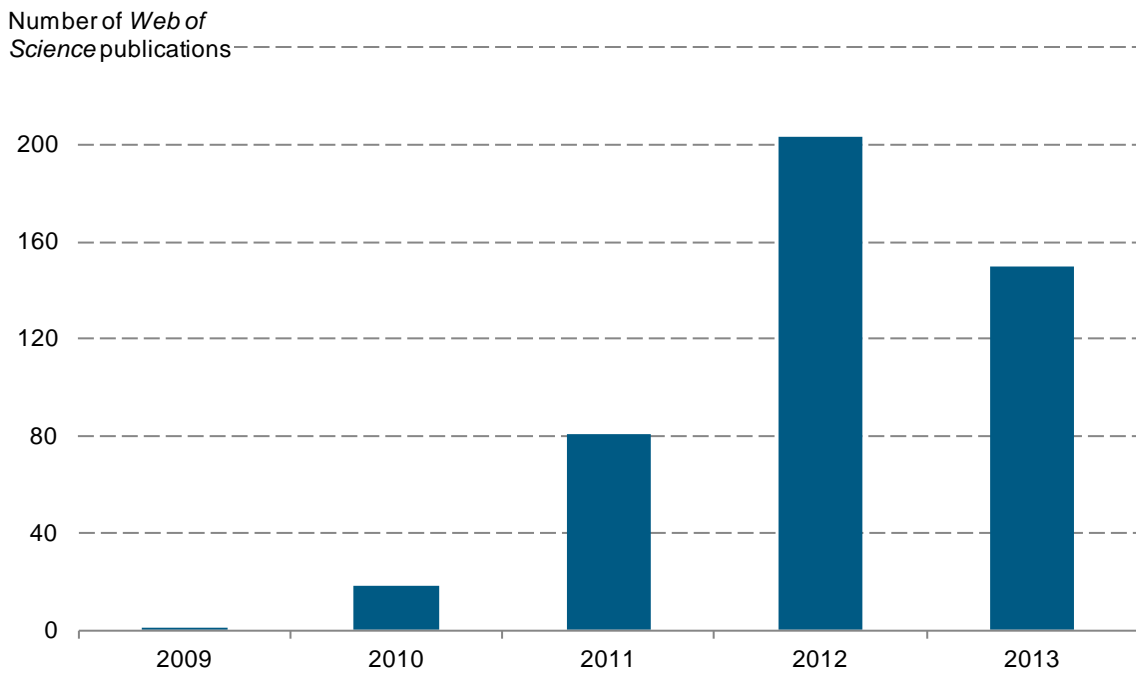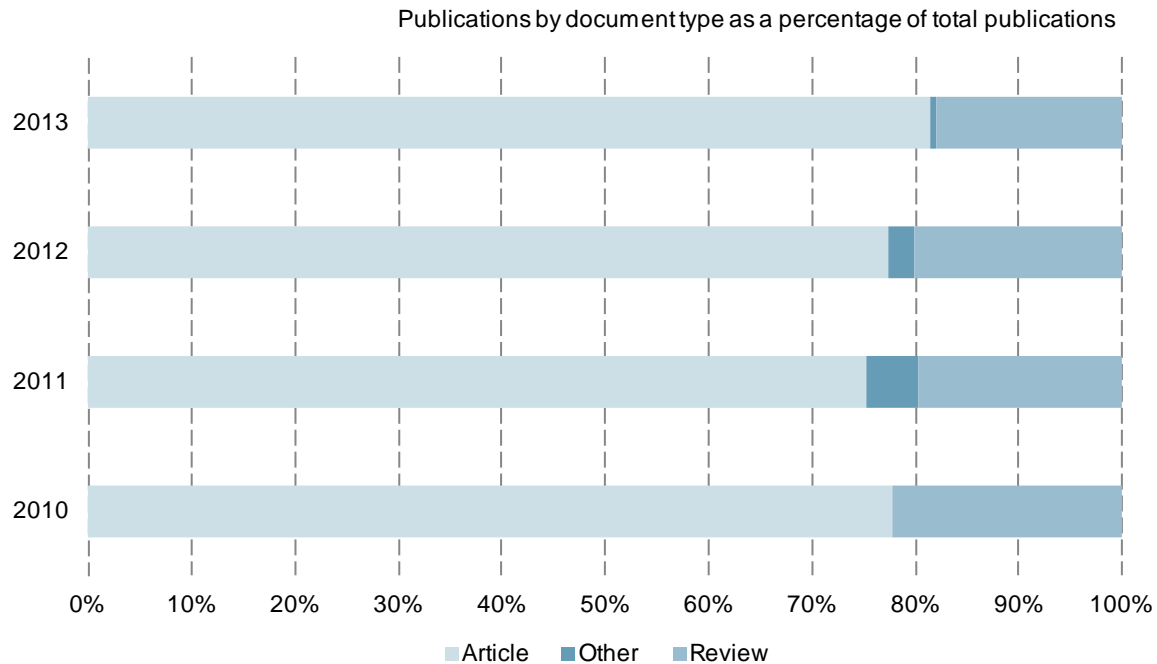Publications by document type as a percentage of total publications

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

The 19 journals appearing most frequently[6] in the IMI project publications dataset, 2009-2013, are listed in Table 4.4.1. A total of 86 journal titles are used more than once.

Together, the 19 most frequently used journals cover 150 *Web of Science* publications, one-third (33.1%) of the total number of items in the dataset.

Annals of the Rheumatic Diseases has overtaken PLoS One as the journal in which IMI project publications appear most frequently. This is driven by growth in output from the Call 2 project BTCure; all 21 publications in this journal are associated with this project. BTCure is also associated with increased output in the journal Arthritis and Rheumatism and the appearance of Arthritis Research & Therapy in the list of most frequently used titles.

The elite multidisciplinary journal PNAS, was not among the most frequently used journal titles in second report but is now the sixth most frequent journal in which IMI project publications appear.

Though there is an increased focus on arthritis and rheumatology in comparison with the second report, the core set of journals for IMI projects continues to highlight the diversity of IMI-supported research with titles in bioinformatics, genetics and psychiatry as well as other disease areas such as diabetes.

All but two of the journals (Molecular Informatics, and Pharmacology Epidemiology and Drug Safety) in Table 4.4.1 are in the top quartile when ranked by Journal Impact Factor among all journals in that category. These exceptions are both ranked in the second quartile.

IMI project publications have been published in a total of 237 journals, of which 155 are ranked in the top quartile (by Journal Impact Factor) of journals in their specific research fields. A total of 333 publications (73.5% of IMI project publications) have been published in these well regarded journals.

---

[6] Table 4.4.1 uses a frequency threshold of at least four publications. This is a change from second reports where this threshold was at least three publications. Below this threshold a further 17 journal titles would be included.

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

Journal	Number of *Web of Science* publications	Number of papers	Journal Impact Factor (2012)	*Web of Science* journal categories
Annals of the Rheumatic Diseases	21	20	9.111	Rheumatology
PLoS One	20	20	3.73	Multidisciplinary Sciences[7]
Molecular Informatics	13	13	2.338	Medicinal Chemistry; Mathematical & Computational Biology
Arthritis and Rheumatism	13	12	7.477	Rheumatology
Pain	12	12	5.644	Anaesthesiology; Clinical Neurology; Neurosciences
Proceedings of the National Academy of Sciences USA (PNAS)	7	7	9.737	Multidisciplinary Sciences
Arthritis Research & Therapy	6	6	4.302	Rheumatology
Diabetologia	6	6	6.487	Endocrinology & Metabolism
European Journal of Cancer	6	6	5.061	Oncology
Journal of Clinical Investigation	6	6	12.812	Research & Experimental Medicine
Pharmacogenomics	6	6	3.857	Pharmacology & Pharmacy
Diabetes	5	5	7.895	Endocrinology & Metabolism
Molecular Psychiatry	5	5	14.897	Biochemistry & Molecular Biology; Neurosciences; Psychiatry
Bioinformatics	4	4	5.323	Biochemical Research Methods; Biotechnology & Applied Microbiology; Mathematical & Computational Biology
BMC Bioinformatics	4	4	3.024	Biochemical Research Methods; Biotechnology & Applied Microbiology; Mathematical & Computational Biology
Drug Discovery Today	4	4	6.551	Pharmacology & Pharmacy
Nucleic Acids Research	4	4	8.278	Biochemistry & Molecular Biology
Pharmacoepidemiology and Drug Safety	4	4	2.897	Pharmacology & Pharmacy
Psychopharmacology	4	4	4.061	Neurosciences; Pharmacology & Pharmacy; Psychiatry

---

[7] PLoS One has been reassigned from the journal category 'Biology' to 'Multidisciplinary Sciences'. Publications in 'Multidisciplinary Sciences' journals are assigned to more specific journal categories based on the references cited by the publication. Where a publication cannot be assigned to a specific journal category, the 'Multidisciplinary Sciences' assignment is retained.

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

Overall, there are 56 publications in journals with an impact factor of 10 or above, an increase of 12 publications over the second report. Of these, some 16 publications appear in journals with an impact factor of 20 or above, 3 more than in the second report. These new publications include elite multidisciplinary journals, such as Nature Medicine (Annex 1).

Together the top 20 journals by Journal Impact Factor account for just under one-tenth (7.9%) of all IMI-supported publications.

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

Journal	Number of *Web of Science* publications	Number of papers	Journal Impact Factor (2012)	Journal categories
Nature	2	2	38.597	Multidisciplinary Sciences
Nature Genetics	3	1	35.209	Genetics & Heredity
Nature Reviews Drug Discovery	2	0	33.078	Biotechnology & Applied Microbiology; Pharmacology & Pharmacy
Nature Biotechnology	1	0	32.438	Biotechnology & Applied Microbiology
Science	1	1	31.027	Multidisciplinary Sciences
JAMA-Journal of the American Medical Association	1	1	29.978	General & Internal Medicine
Nature Immunology	1	1	26.199	Immunology
Lancet Neurology	2	2	23.917	Clinical Neurology
Nature Medicine	2	2	22.864	Biochemistry & Molecular Biology; Cell Biology; Research & Experimental Medicine
Pharmacological Reviews	1	1	22.345	Pharmacology & Pharmacy
British Medical Journal	3	3	17.215	General & Internal Medicine
Nature Reviews Neurology	1	1	15.518	Clinical Neurology
PLoS Medicine	2	2	15.253	General & Internal Medicine
Nature Neuroscience	1	1	15.251	Neurosciences
Nature Reviews Clinical Oncology	1	1	15.031	Oncology
Molecular Psychiatry	5	5	14.897	Biochemistry & Molecular Biology; Neurosciences; Psychiatry
American Journal of Psychiatry	2	2	14.721	Psychiatry
Alzheimer's & Dementia	2	2	14.483	Clinical Neurology
European Heart Journal	1	1	14.097	Cardiac & Cardiovascular Systems
Journal of Experimental Medicine	2	2	13.214	Immunology; Research & Experimental Medicine

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

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

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

Number of *Web of Science* publications   ■Call 1 ■Call 2 ■Call 3 ■Call 4 □Unassigned
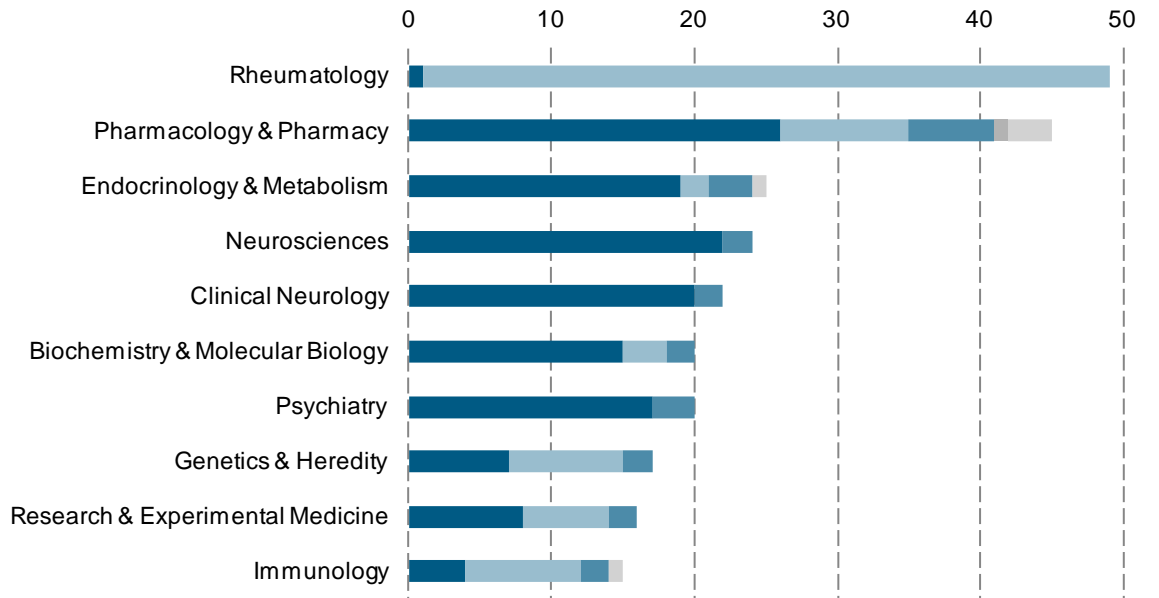
- Just over one-tenth (10.8%) of IMI project research is assigned to the journal category of Rheumatology which has overtaken Pharmacology & Pharmacy (9.9%) as the most frequently used journal category in this dataset.

- Biochemistry & Molecular Biology, Genetics & Heredity and Immunology were not in the most frequently used journal categories in the second report to IMI.

- Output in Rheumatology journals has again more than doubled since the second report; all but one of the publications in this category are associated with the BTCure project (Call 2) with one publication from PROTECT (Call 1). This highlights the productivity of the BTCure project which has accumulated 26 *Web of Science* publications since the second report.

- Oncology and Mathematics & Computational Biology are no longer among the top ten most frequently used journal categories.

- The most frequently used journal categories in Figure 4.5.1 continue to reflect breadth and depth in the IMI research portfolio which contains cross-cutting and more specialised research.

The analysis presented in Figure 4.5.1 includes all publication types and spans the full time period of IMI-supported publications (2009-2013).

Standard descriptions of the scope of these journal categories are given in Annex 2.

---

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

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 (Section 6.1). Table 4.6.1 presents a viewpoint of IMI-supported papers at the end of 2012 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).

SUMMARY OF KEY FINDINGS

The citation impact of papers associated with IMI projects or papers by IMI-supported researchers is internationally influential with citation impact at or above twice world average and around three-quarters of output in top quartile journals (Table 4.6.1).

- The citation impact for IMI project papers is 2.04 (where world average is 1.0) for the 3-year period, 2010-2012. This is a substantial increase from the second report (average citation impact = 1.55) and indicates that the quality of this research (as indexed by citation impact) has not only been maintained but has increased while output has grown (Section 4.3).

- The citation impact for IMI project papers is lower, on average, than the citation impact for the IMI researchers dataset (2.26) but the margin of difference between the two datasets has decreased relative to previous reports.

- The citation impact for IMI project papers remains well above the EU's average citation impact[9,10] relative to the world baseline for the same 3-year period in similar research fields (average citation impact = 1.13).

More than one-tenth of publications from IMI projects or IMI-supported researchers are published in open access journals (12.2% and 11.6%, respectively – Table 4.6.2). These values are above the global average reported by a 2011 study reviewing accessibility in the journal literature between 2003 and 2009[11] though it should be noted that the majority of IMI research has been published since 2009 and the global average share of publications that are openly accessible is likely to have grown.

Overall, though the IMI projects dataset is small in comparison with the IMI researchers dataset, these data show that the performance of the two groups is similar.

---

[9] EU-27 grouping of countries: Thomson Reuters *National Science Indicators* 2012 database; similar research has been defined as including the same journal categories as in the IMI project dataset.

[10] For this analysis, only papers are considered since only these publication types have normalised citation impact data (see Section 3.1.3).

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

TABLE 4.6.1 SUMMARY CITATION ANALYSIS FOR IMI RESEARCH – CITATIONS TO END-2012

2010-2011	Citation impact				
	Number of papers	Normalised at field level	Normalised at journal level	Average percentile	% Highly-cited papers[12]
IMI projects	269	2.04	1.24	54.0	19.3%
IMI researchers	10 949	2.26	1.30	44.7	23.8%

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

2010 to current	IMI publications		*Web of Science* publications			
	Total	% Open access journals[13]	Total	Citations	Raw citation impact	% Top quartile journals[14]
IMI projects	483	12.2%	453	2 301	5.08	73.5%
IMI researchers	n/a	11.6%	17 063	155 054	9.09	71.9%

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 all the top ten journal categories is, on average, well above the citation impact of similar European research and in the majority of categories is well above the citation impact for the IMI researchers dataset.

IMI project research in Pharmacology & Pharmacy, Neurosciences, Psychiatry and Immunology has substantially higher citation impact, on average, than similar research by IMI-supported researchers.

IMI project research in Rheumatology and Endocrinology & Metabolism has substantially lower citation impact, on average, than similar research by IMI-supported researchers.

---

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

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

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

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

■ Citation impact, IMI project papers, 2010-2012

■ Citation impact, IMI researcher papers, 2010-2012

■ Citation impact, EU-27 papers, 2010-2012
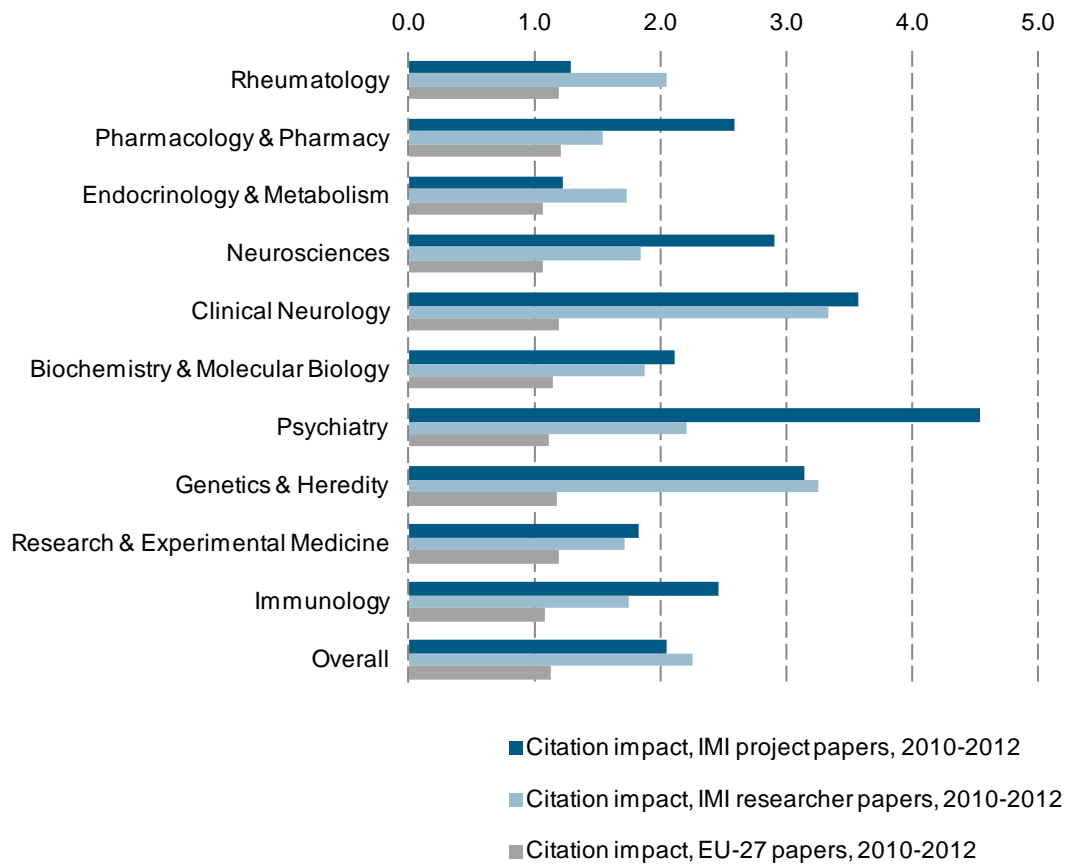
TABLE 4.6.3 SUMMARY OF PUBLICATION OUTPUT AND 3-YEAR AVERAGE CITATION IMPACT FOR IMI-SUPPORTED RESEARCH BY TOP *WEB OF SCIENCE* JOURNAL CATEGORIES, 2010-2012 BENCHMARKED AGAINST THE IMI RESEARCHERS DATASET AND SIMILAR PUBLICATIONS FROM THE EUROPEAN RESEARCH BASE

	IMI projects		IMI researchers		EU-27	
Web of Science journal category	Number of papers[15]	Citation impact	Number of papers	Citation impact	Number of papers	Citation impact
Rheumatology	21	1.29	783	2.05	6 755	1.18
Pharmacology & Pharmacy	34	2.58	1 067	1.54	34 037	1.20
Endocrinology & Metabolism	21	1.23	644	1.74	20 411	1.07
Neurosciences	48	2.90	1 222	1.85	42 577	1.07
Clinical Neurology	26	3.57	740	3.33	28 945	1.19
Biochemistry & Molecular Biology	26	2.12	967	1.88	57 923	1.15
Psychiatry	19	4.54	569	2.20	19 798	1.12
Genetics & Heredity	11	3.14	878	3.26	22 021	1.17

---

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

Web of Science journal category	IMI projects		IMI researchers		EU-27	
	Number of papers[15]	Citation impact	Number of papers	Citation impact	Number of papers	Citation impact
Research & Experimental Medicine	6	1.83	266	1.71	14 396	1.18
Immunology	12	2.46	650	1.75	24 610	1.08
Overall	**269**	**2.04**	**10 949**	**2.26**	**1 167 910**	**1.13**

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

These analyses therefore give a useful 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 further.

Standard definitions of the scope of the journal categories in Figure 4.6.1 and Table 4.6.3 are given in Annex 2.

4.7   HOW DOES IMI PROJECT RESEARCH COMPARE WITH RESEARCH FROM OTHER BIOMEDICAL FUNDING ORGANISATIONS?

This new Section of the report evaluates trends in the performance of IMI project research and benchmarks this performance against the Wellcome Trust, a leading UK funder of biomedical research.

Wellcome Trust publications were identified by specific keyword searches using funding acknowledgment data in *Web of Science*. This is the same process by which IMI project publications have been identified (Section 4.1). This dataset is referred to as 'Wellcome Trust research' in the analyses below. Papers resulting from Wellcome Trust funding where the authors have not specifically acknowledged this funding are not covered by this dataset.

Data are presented for a 3-year time period (2010-2012) for papers where normalised citation impact can be calculated. This provides a reference for both datasets relative to world citation baselines ensuring an even basis for comparison.

4.7.1   TRENDS IN THE OUTPUT AND CITATION IMPACT OF IMI PROJECT RESEARCH COMPARED WITH WELLCOME TRUST RESEARCH

Figures 4.7.1 to 4.7.3 show trends in the output and performance of IMI project research compared with Wellcome Trust research. When considering these analyses, earlier caveats regarding paper numbers should be borne in mind (Section 3.1.4).

In summary:

- IMI project research shows year-on-year growth. This is to be expected as the number of projects funded by IMI increases and this growing body of projects yields results for publication (Figure 4.7.1).

- The citation impact of IMI project research in the most recent year of analysis is just over twice world average and is comparable to the citation impact of Wellcome Trust research which has remained stable at about twice world average over the time period (Figure 4.7.1).

- Though paper numbers are low (n=15) the citation impact of IMI project research published in 2010 is exceptional at over four times world average. This performance is driven by several highly-cited papers rather than a single paper with exceptional citation impact (Figure 4.7.1).

- IMI project research and Wellcome Trust research show a similar profile of uncited research over the time period. No IMI project papers published in 2010 are uncited (Figure 4.7.2).

- Just under one-fifth (17.6%) of IMI project papers in the most recent year of analysis are already highly-cited (ranked in the world's top decile relative to journal category and year). This is a similar percentage to the Wellcome Trust (17.3%) and indicates performance well above world average (Figure 4.7.3).

- Though IMI is a 'young' funding initiative in comparison to the Wellcome Trust, the performance of IMI project research and Wellcome Trust is comparable across the majority of indicators assessed. An exception is the percentage of uncited papers at overall level, which is substantially higher for IMI than for the Wellcome Trust, though the underlying trends are similar. This is attributable to the differences in output trends observed above – more than two-thirds (67.7%) of IMI project papers are from 2012 in comparison with just over one-third (34.7%) of Wellcome Trust papers. As more recent research is less likely to be cited than older research this should not be taken as evidence that IMI project research is more likely to remain uncited (Table 4.7.1, Figure 4.7.1).

FIGURE 4.7.1 TRENDS IN OUTPUT AND CITATION IMPACT – IMI PROJECT RESEARCH COMPARED WITH WELLCOME TRUST RESEARCH

■ Wellcome Trust research - share of output   ■ IMI project research - share of output
◆ Wellcome Trust research - citation impact   ▲ IMI project research - citation impact
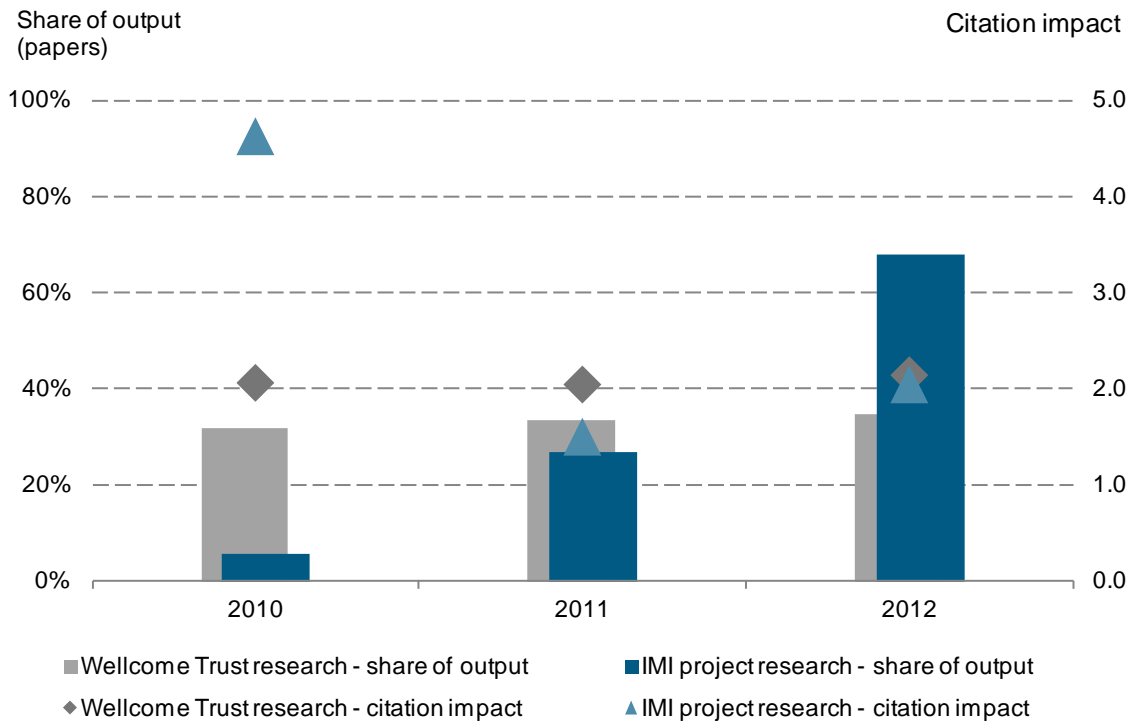
FIGURE 4.7.2 TRENDS IN UNCITED RESEARCH – IMI PROJECT RESEARCH COMPARED WITH WELLCOME TRUST RESEARCH
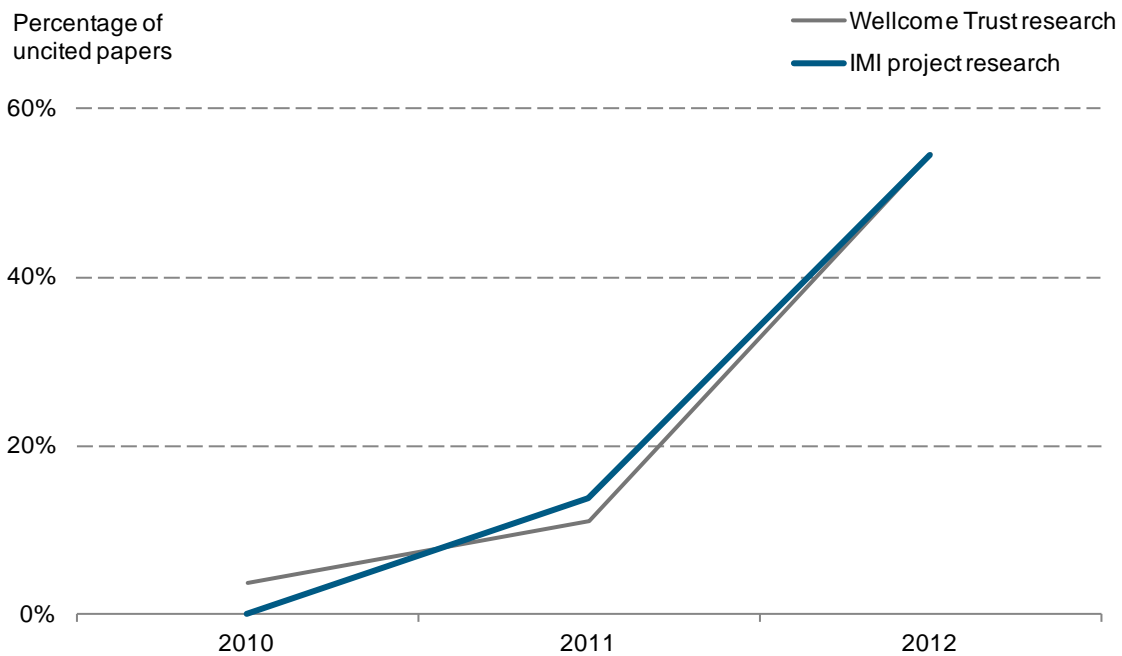
FIGURE 4.7.3 TRENDS IN HIGHLY-CITED RESEARCH – IMI PROJECT RESEARCH COMPARED WITH WELLCOME TRUST RESEARCH

TABLE 4.7.1 SUMMARY BIBLIOMETRIC INDICATORS – IMI PROJECT RESEARCH COMPARED WITH WELLCOME TRUST RESEARCH

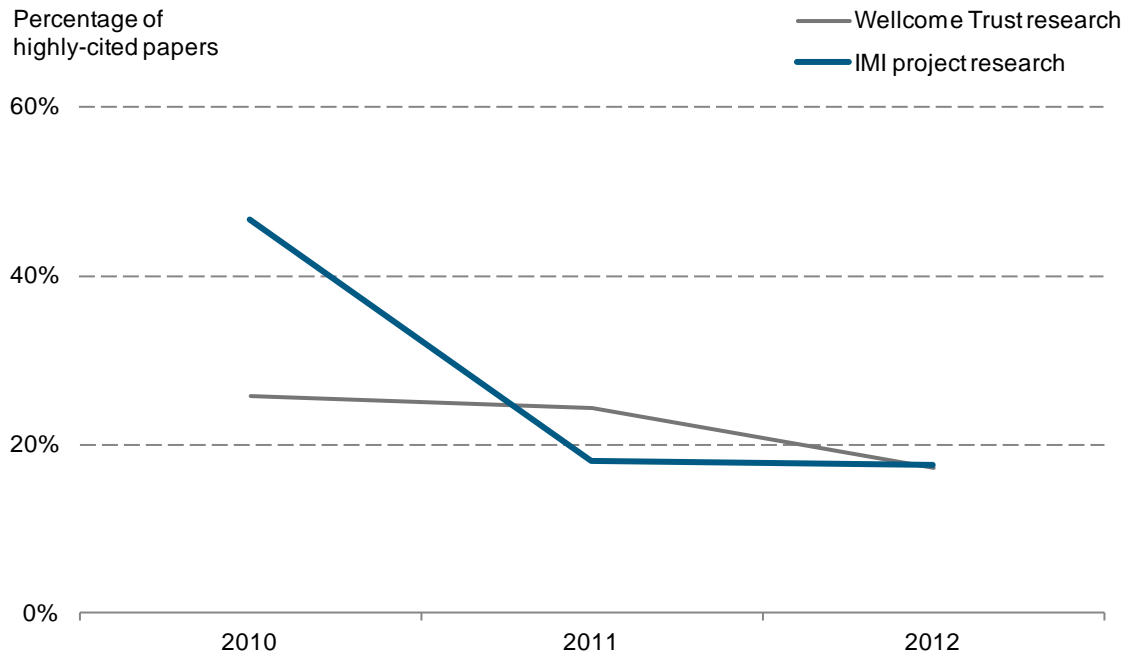|	Number of papers	Citation impact (normalised at field level)	Percentage of uncited papers	Percentage of highly-cited papers
IMI project research	269	2.04	40.5%	19.3%
Wellcome Trust research	15 483	2.08	23.9%	22.3%

4.8 HOW COLLABORATIVE IS IMI PROJECT RESEARCH?

International research collaboration is a rapidly growing element of research activity.[16] The reasons for this have not been fully clarified but include increasing access to facilities and resources, increasing access to knowledge and increasing access to people and expertise. In addition, international collaboration has been shown to be associated with an increase in the number of citations received by research papers, although this does depends on the partner countries involved.[17] Co-authorship is likely to be a good indicator of collaboration, although there will be collaborations that do not result in co-authored papers, and co-authored papers which may have required limited collaboration. Alternative data-based approaches, for example using information about co-funding or international exchanges, have limitations in terms of both comprehensiveness and validity.

In this report, co-authorship is used as a measure of collaboration. Table 4.8.1 compares the output and citation impact of IMI project papers that are co-authored between different sectors, institutions and countries. Sectors are those used in the IMI researchers dataset (Section 6.1).

The data in Table 4.8.1 show that IMI project research is collaborative at sector, institution and country level.

- Well over half (61.3%) of all IMI project papers have been published by researchers affiliated with different sectors.

- Three-quarters (74.7%) of IMI project papers are collaborative between institutions.

- Half (49.8%) of all IMI project papers have are internationally collaborative.

- Collaborative IMI project research is internationally influential with citation impact well over twice world average (1.0) and a clear margin over non-collaborative IMI project research.

TABLE 4.8.1 CROSS-SECTOR, CROSS-INSTITUTION AND INTERNATIONAL OUTPUT– IMI PROJECT RESEARCH

	Number of papers	Percentage of papers	Citation impact (normalised at field level)
Cross-sector	165	61.3%	2.24
Single-sector	104	38.7%	1.73
Cross-institution	201	74.7%	2.26
Single-institution	68	25.3%	1.39
International	134	49.8%	2.57
Domestic	135	50.2%	1.51

A paper is defined as cross-sector if the listed addresses are from more than one sector. For example, if a paper has two addresses – University of Copenhagen and Novartis – it would be classified as cross-sector. If a paper has only two addresses –University of Cambridge and Utrecht University – it would be classified as single-sector.

A paper is defined as cross-institution if more than one institution is listed in the addresses.

A paper is defined as international if more than one country is listed in the addresses or domestic if a single country is listed.

---

[16] Adams J (2013). Collaborations: the fourth age of research. Nature, 497, 557-560.

[17] Adams, J., Gurney, K., & Marshall, S. (2007). Patterns of international collaboration for the UK and leading partners. A report by *Evidence* Ltd to the UK Office of Science and Innovation. 27pp.

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. Data are presented for projects in Calls 1 to 3 as there is only a single publication from a Call 4 project (ORBITO) and this has not yet been cited.

When considering the analyses in this Section, earlier caveats regarding paper numbers should be borne in mind (Section 3.1.4).

5.1 TRENDS IN PUBLICATION OUTPUT BY IMI FUNDING CALL

The data in Figure 5.1.1 show that the majority of IMI-supported publications and papers are associated with Calls 1 and 2, however, output has also increased for Call 3 since February 2013. Among Call 4 projects, only ORBITO has published (1 paper, as yet uncited) and there are no publications associated with the three projects in Calls 5 to 6. This is not unexpected as these projects are very recent (funded between October 2012 and January 2013) and it may take time for a particular project to generate results for publication.

FIGURE 5.1.1 NUMBER OF *WEB OF SCIENCE* PUBLICATIONS BY YEAR AND FUNDING CALL
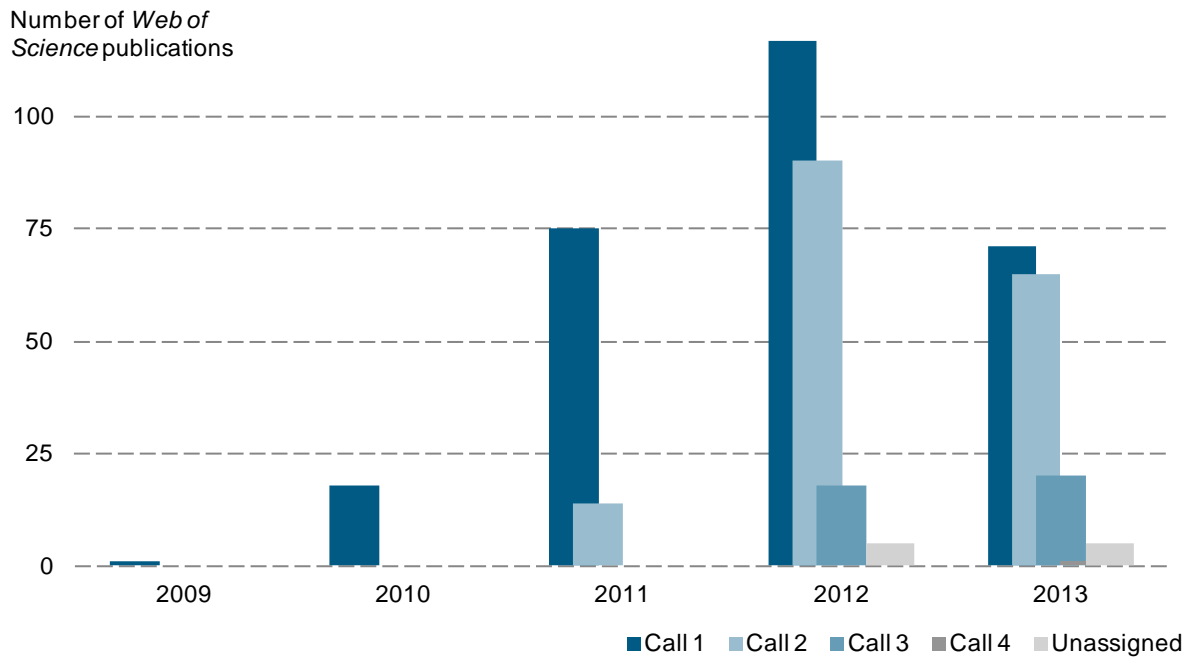
Growth in output is particularly pronounced for Call 2 with 2012 and 2013 output now approaching Call 1 totals. Output associated with Call 3 has more than doubled compared to the second report (16 publications – second report; 38 publications – third report) and Call 4 has accumulated its first publication. Though projects from Calls 5 and 6 were also included in the search terms used to define the dataset of IMI project publications, no publications from these calls have yet been found.

5.2   SUMMARY BIBLIOMETRIC ANALYSES FOR IMI PROJECTS – CALL 1

Figure 5.2.1 presents a 'bubble-chart' visualisation of IMI-supported research for those projects with at least 10 papers – one of which is highly-cited – over the time period (2010-2012). The number of papers, 3-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, 3-YEAR AVERAGE CITATION IMPACT AND SHARE OF HIGHLY-CITED RESEARCH FOR SELECTED IMI PROJECTS – CALL 1
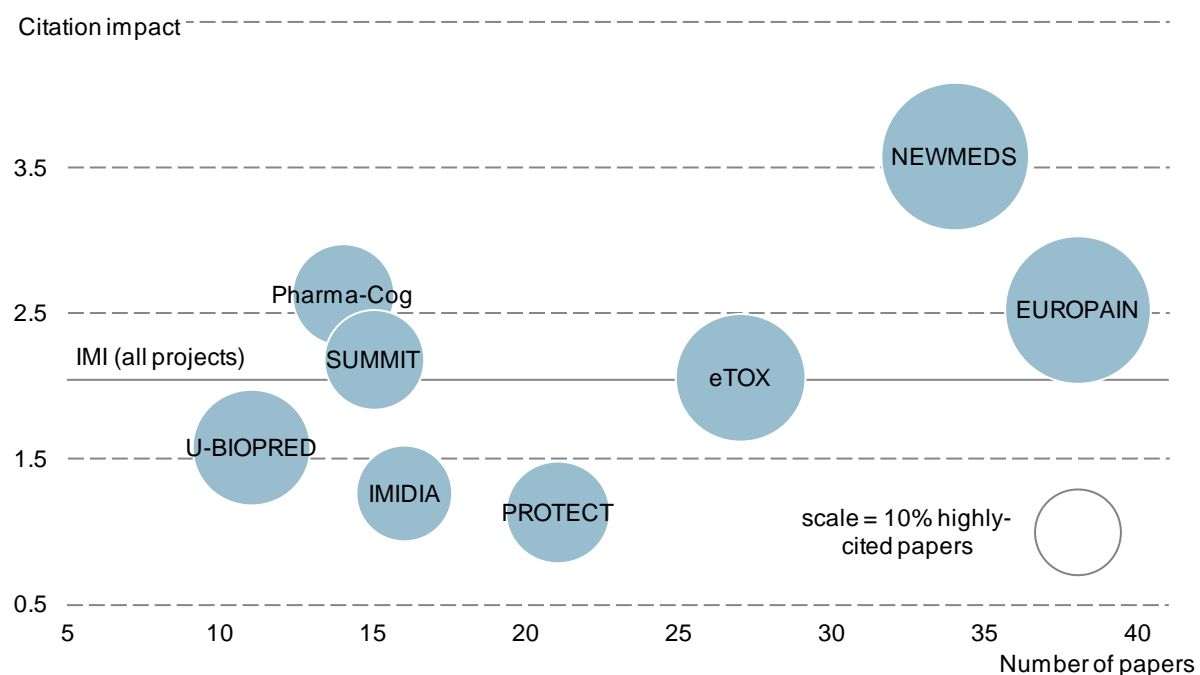

Figure 5.2.1 has been updated from the second report to cover a 3-year rather than 2-year time period; world citation baselines for 2012 became available subsequent to the second report, enabling the calculation of normalised citation impact for 2012 papers (Section 4.1.1).

The data in Figure 5.2.1 show that:

- The average citation impact of all of these eight projects is above world average (1.0) and in all but one case has increased or been maintained since the second report. The citation impact of research associated with Pharma-Cog has decreased slightly since the second report but this research remains very well-cited.

- Research associated with five of these eight projects (eTOX, EUROPAIN, NEWMEDS, Pharma-Cog, SUMMIT) is very well-cited with citation impact at or above twice world average.

- Research associated with the NEWMEDS project is exceptionally well-cited with citation impact over three times world average and well above the average for all IMI projects. Three years after funding, almost one-third (29.4%) of NEWMEDS papers are highly-cited.

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).

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

TABLE 5.2.1 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 1 – CITATIONS TO END-2012

Project	Citation impact				% Highly-cited papers[18]
	Number of papers	Normalised at field level	Normalised at journal level	Average percentile	
EMTRAIN	0	0.00	0.00	0.00	0.0%
eTOX	27	2.06	1.81	42.09	22.2%
Eu2P	0	0.00	0.00	0.00	0.0%
EUROPAIN	38	2.53	1.93	46.80	28.9%
IMIDIA	16	1.26	0.67	52.53	12.5%
MARCAR	3	1.78	1.59	44.75	0.0%
NEWMEDS	34	3.58	1.56	46.11	29.4%
Pharma-Cog	14	2.63	1.22	49.43	14.3%
Pharmatrain	0	0.00	0.00	0.00	0.0%
PRO-active	6	0.46	1.87	81.03	0.0%
PROTECT	21	1.13	0.70	65.30	14.3%
SafeSciMET	1	0.00	0.00	100.00	0.0%
SAFE-T	2	0.91	0.44	59.32	0.0%
SUMMIT	15	2.18	0.72	59.71	13.3%
U-BIOPRED	11	1.58	0.93	52.86	18.2%
Overall (IMI projects)	**269**	**2.04**	**1.24**	**54.00**	**19.3%**

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

Project	IMI publications		*Web of Science* publications			
	Total	% Open access journals	Total	Citations	Raw citation impact	% Top quartile journals
EMTRAIN	1	0.0%	0	0	0.00	0.0%
eTOX	36	27.8%	35	207	5.91	0.0%
Eu2P	1	0.0%	0	0	0.00	0.0%
EUROPAIN	55	7.3%	52	450	8.65	0.0%
IMIDIA	22	4.5%	20	142	7.10	85.0%
MARCAR	14	21.4%	13	27	2.08	92.3%
NEWMEDS	61	8.2%	53	309	5.83	79.2%
Pharma-Cog	20	5.0%	18	245	13.61	61.1%
Pharmatrain	1	0.0%	0	0	0.00	0.0%
PRO-active	6	83.3%	6	8	1.33	0.0%
PROTECT	25	8.0%	24	70	2.92	75.0%
SafeSciMET	4	0.0%	2	2	1.00	0.0%

---

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

	IMI publications		*Web of Science* publications			
Project	Total	% Open access journals	Total	Citations	Raw citation impact	% Top quartile journals
SAFE-T	4	25.0%	3	13	4.33	0.0%
SUMMIT	23	39.1%	21	88	4.19	90.5%
U-BIOPRED	14	0.0%	14	121	8.64	78.6%
Overall (IMI projects)	**483**	**12.2%**	**453**	**2 301**	**5.08**	**73.5%**

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 3.

5.3   SUMMARY BIBLIOMETRIC ANALYSES FOR IMI PROJECTS – CALL 2

The numbers of papers from Call 2 projects at end-2012 are now sufficient for a 'bubble chart' visualisation of IMI-supported research at project level.

Figure 5.3.1 presents an analysis of IMI-supported research for those projects with at least 5 papers – one of which is highly-cited – over the time period (2010-2012). The number of papers, 3-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 5.3.1 PAPER NUMBERS, 3-YEAR AVERAGE CITATION IMPACT AND SHARE OF HIGHLY-CITED RESEARCH FOR SELECTED IMI PROJECTS – CALL 2
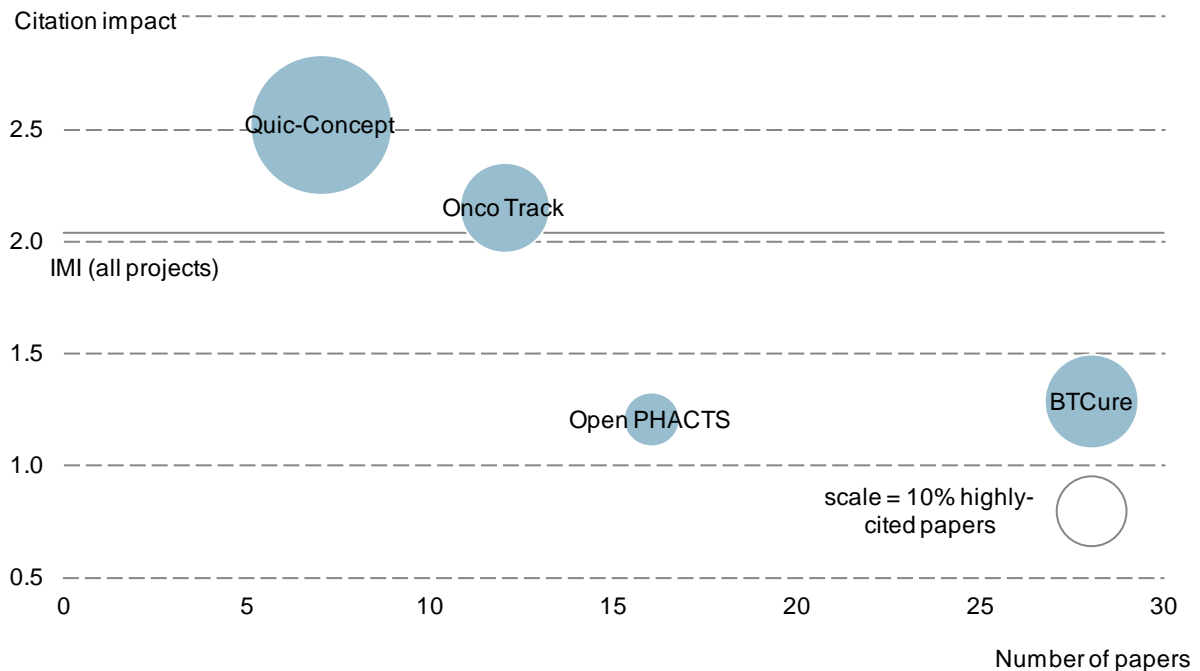
The data in Figure 5.3.1 show that:

- The average citation impact of all four of these projects is above world average (1.0).
- Research associated with Onco Track or Quic-Concept is very well-cited with citation impact above twice world average.
- BTCure has been the most prolific Call 2 project, with more than twice the number of papers of Open PHACTs, the next most prolific project in this call.

Tables 5.3.1 and 5.3.2 compare bibliometric indicators for all projects in Call 2. Table 5.3.1 presents indicators where citation impact has been normalised against world average values and is an expansion of the data used in Figure 5.3.1. Table 5.3.2 presents a more recent (but also more descriptive) viewpoint using indicators based on current (mid-August) citation counts (see Section 4.1.1).

TABLE 5.3.1 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 2 – CITATIONS TO END-2012

Project	Citation impact				
	Number of papers	Normalised at field level	Normalised at journal level	Average percentile	% Highly-cited papers
BTCure	28	1.29	0.57	66.81	17.9%
DDMoRe	2	1.20	0.75	58.17	0.0%
EHR4CR	0	0.00	0.00	0.00	0.0%
Onco Track	12	2.15	1.39	43.15	16.7%
Open PHACTS	16	1.20	1.09	68.58	6.3%
Predect	1	0.00	0.00	100.00	0.0%
Quic-Concept	7	2.52	2.50	28.09	42.9%
RAPP-ID	2	1.99	2.03	53.30	50.0%
Overall (IMI projects)	**269**	**2.04**	**1.24**	**54.00**	**19.3%**

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

Project	IMI publications		*Web of Science* publications			
	Total	% Open access journals	Total	Citations	Raw citation impact	% Top quartile journals
BTCure	77	5.2%	76	191	2.51	89.5%
DDMoRe	9	22.2%	8	5	0.63	62.5%
EHR4CR	1	0.0%	0	0	0.00	0.0%
Onco Track	18	27.8%	17	101	5.94	64.7%
Open PHACTS	25	8.0%	23	98	4.26	52.2%
Predect	3	0.0%	3	0	0.00	100.0%
Quic-Concept	11	9.1%	11	43	3.91	81.8%
RAPP-ID	9	22.2%	9	11	1.22	88.9%
Overall (IMI projects)	**483**	**12.2%**	**453**	**2 301**	**5.08**	**73.5%**

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

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 3.

5.4 SUMMARY BIBLIOMETRIC ANALYSES FOR IMI PROJECTS – CALL 3

The numbers of papers from Call 3 projects at end-2012 are generally too few to allow a 'bubble-chart' visualisation of IMI-supported research at project level. An exception to this would be for EU-AIMS with six papers, two of which are highly-cited. Providing paper numbers are sufficient, a 'bubble chart' visualisation of Call 3 projects will be included in the next report.

The number of publications from Call 3 projects has increased since the second report and Tables 5.4.1 and 5.4.2 compare bibliometric indicators for all projects in Call 3. Table 5.4.1 presents indicators where citation impact has been normalised against world average. Table 5.4.2 presents a more recent (but also more descriptive) viewpoint using indicators based on current (mid-August) citation counts (see Section 4.1.1).

TABLE 5.4.1 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 3 – CITATIONS TO END-2012

	Citation impact				
Project	Number of papers	Normalised at field level	Normalised at journal level	Average percentile	% Highly-cited papers
ABIRISK	0	0.00	0.00	0.00	0.0%
BioVacSafe	3	0.55	0.32	75.12	0.0%
DIRECT	0	0.00	0.00	0.00	0.0%
EU-AIMS	6	4.44	0.71	45.05	33.3%
EUPATI	0	0.00	0.00	0.00	0.0%
MIP-DILI	2	0.91	0.72	64.41	0.0%
PreDiCT-TB	0	0.00	0.00	0.00	0.0%
Overall (IMI projects)	**269**	**2.04**	**1.24**	**54.00**	**19.3%**

TABLE 5.4.2 SUMMARY BIBLIOMETRIC INDICATORS FOR IMI PROJECTS IN CALL 3 – CITATIONS TO CURRENT

	IMI publications		*Web of Science* publications			
Project	Total	% Open access journals	Total	Citations	Raw citation impact	% Top quartile journals
ABIRISK	6	33.3%	6	1	0.17	50.0%
BioVacSafe	7	0.0%	6	13	2.17	33.3%
DIRECT	3	0.0%	3	0	0.00	0.0%
EU-AIMS	15	0.0%	15	141	9.40	80.0%
EUPATI	0	0.0%	0	0	0.00	0.0%
MIP-DILI	4	25.0%	4	3	0.75	100.0%
PreDiCT-TB	3	0.0%	3	0	0.00	66.7%
Overall (IMI projects)	**483**	**12.2%**	**453**	**2 301**	**5.08**	**73.5%**

There are no IMI publications associated with EUPATI. Among all Call 3 projects, only BioVacSafe, EU-AIMS and MIP-DILI had published cited research at end-2012 though the remaining projects have published cited research subsequently.

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 3.

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

Figure 5.5.1 and Figure 5.5.2 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 based on total publication volume.

FIGURE 5.5.1 TRENDS IN (A) OUTPUT AND (B) RAW CITATION IMPACT FOR RESEARCH FROM IMI-SUPPORTED PROJECTS IN CALL 1: ETOX, EUROPAIN, IMIDIA, NEWMEDS, PROTECT AND SUMMIT

Figure 5.5.1A shows that EUROPAIN and NEWMEDS continue to be the most prolific Call 1 projects. NEWMEDS has overtaken EUROPAIN in terms of publication output in 2012 and the margin of difference has increased in 2013.
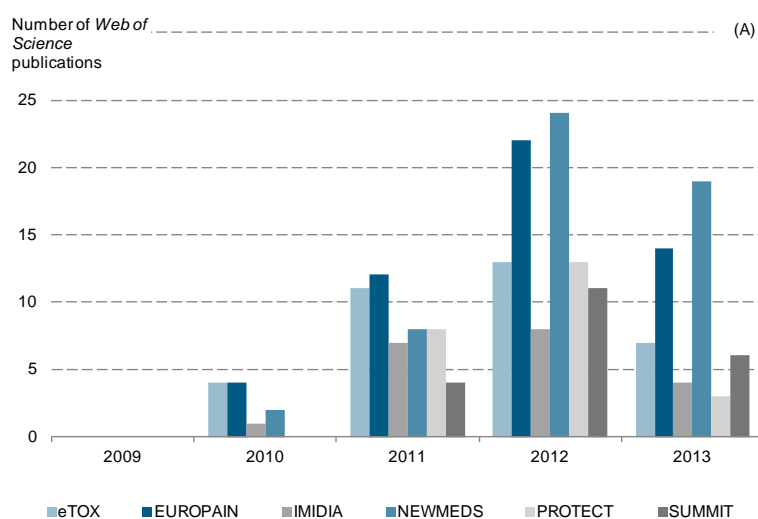
Figure 5.5.1B continues to show 'spikes' in the raw citation impact of EUROPAIN and IMIDIA publications in 2010. For EUROPAIN this is attributable to three highly-cited papers one of which is defined as a 'hot paper' (Annex 2). For IMIDIA this is attributable to a single highly-cited paper.
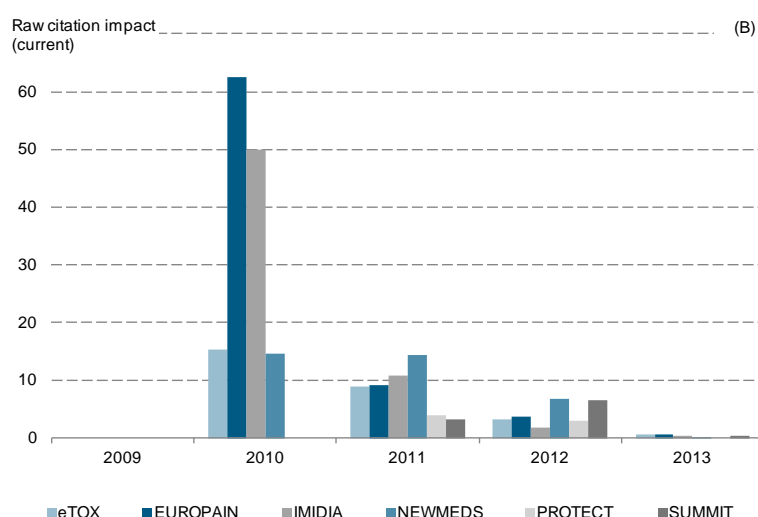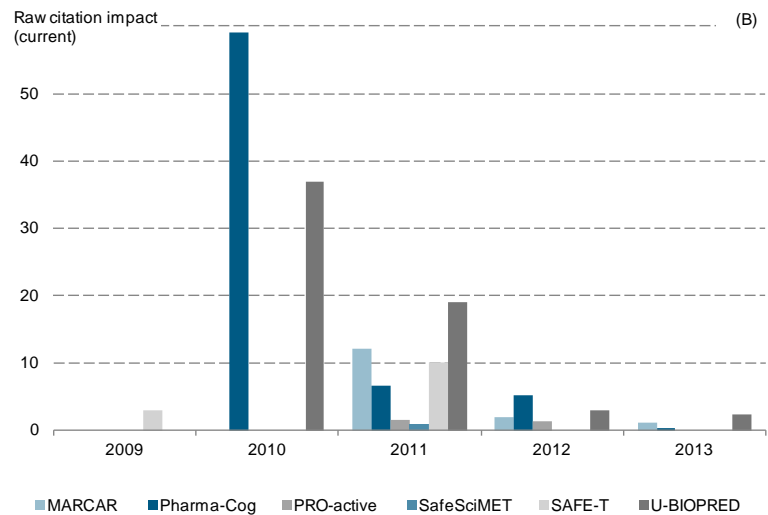
39

FIGURE 5.5.2 TRENDS IN (A) OUTPUT AND (B) RAW CITATION IMPACT FOR RESEARCH FROM IMI-SUPPORTED PROJECTS IN CALL 1: MARCAR, PHARMA-COG, PRO-ACTIVE, SAFE-T AND U-BIOPRED.

Figure 5.5.2A shows that Pharma-Cog accounts for the highest share of publications in this second group.   In previous reports, this project was included in the first group of projects but has since been overtaken by SUMMIT in terms of publication volume.

BIOPRED and MARCAR account for the second and third highest shares of publications.   Output from MARCAR has increased substantially since the second report (from 4 to 13 publications) while output from U-BIOPRED has remained relatively stable (1 additional publication).

Figure 5.5.2B shows that Pharma-Cog has the highest raw citation impact of projects in this group, with U-BIOPRED second highest. MARCAR research is concentrated in 2013 and has therefore had less time to accumulate citations than other projects in this group.

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

Figure 5.6.1 and Figure 5.6.2 show the publication output and raw citation impact of *Web of Science* publications associated with projects in Call 2. For clarity, the projects are split into two groups based on total publication volume.
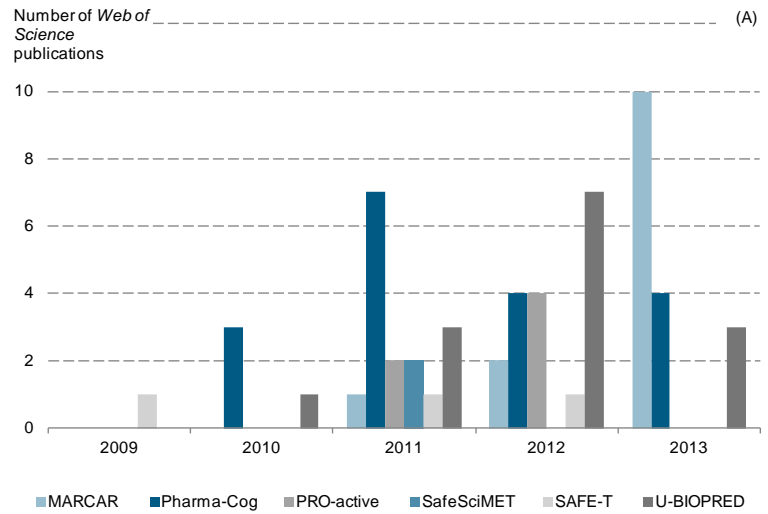
FIGURE 5.6.1 TRENDS IN (A) OUTPUT AND (B) RAW CITATION IMPACT FOR RESEARCH FROM IMI-SUPPORTED PROJECTS IN CALL 2: BTCURE, OPEN PHACTS, ONCO TRACK AND QUIC-CONCEPT

Figure 5.6.1A shows that BTCure continues to be the most prolific Call 2 project with a substantially higher volume of publications compared to other projects in this call. BTCure is the most prolific IMI project in any IMI funding call with output eclipsing the most prolific Call 1 projects and three times that of Open PHACTS, the next most prolific Call 2 project.
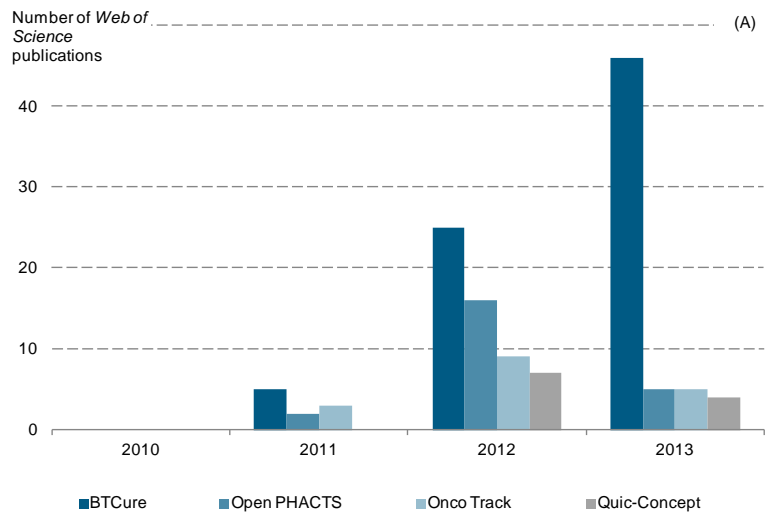
BTCure and Onco Track have the highest raw citation impact of projects in this group (Figure 5.6.1B).

41
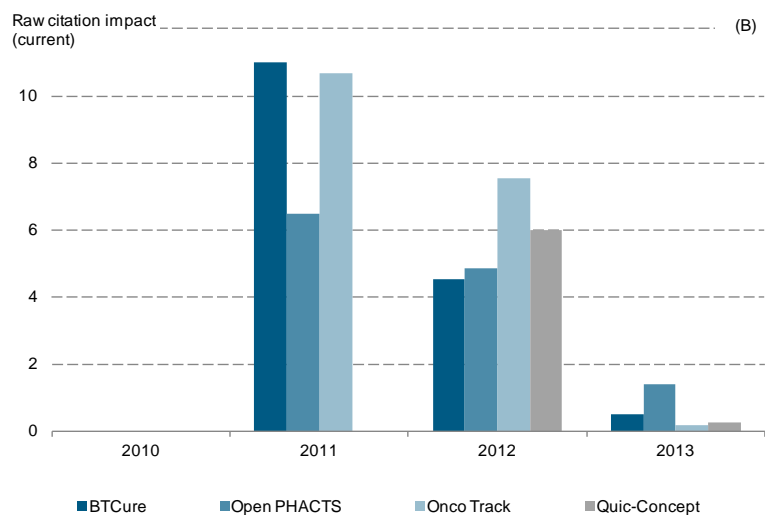
FIGURE 5.6.2 TRENDS IN (A) OUTPUT AND (B) RAW CITATION IMPACT FOR RESEARCH FROM IMI-SUPPORTED PROJECTS IN CALL 2: DDMORE, PREDECT AND RAPP-ID

Figure 5.6.2A shows that RAPP-ID is the most prolific of the Call 2 projects in this group, closely followed by DDMoRe.

Predect only had a single *Web of Science* publication at the time of the second report but has subsequently accumulated two further publications.
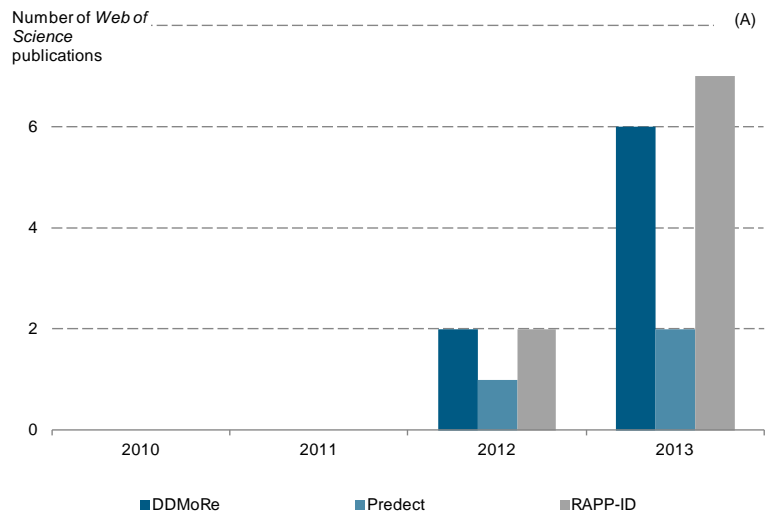
Figure 5.6.2B shows that RAPP-ID has the highest raw citation impact of projects in this group. None of the three Predect publications have yet been cited.
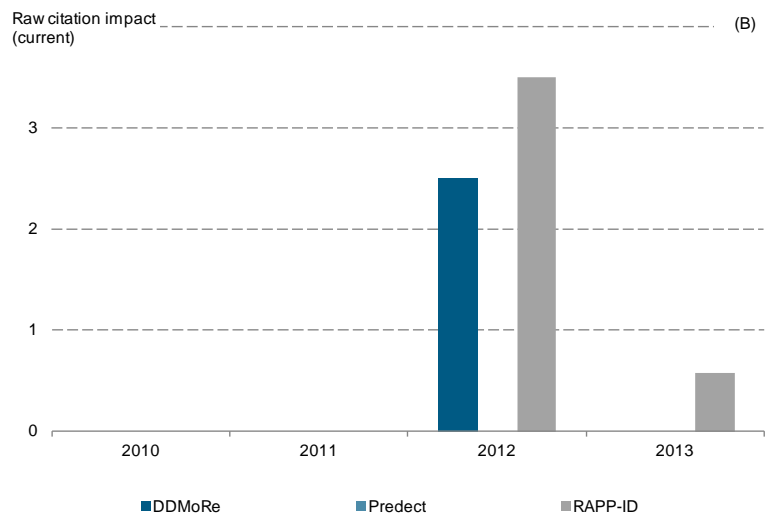
EHR4CR is not included in the above analyses as it has not yet generated publications.

As there has been an increase in the number of Call 3 publications compared to the second report, Section 5.7 presents similar trends analyses for Call 3 projects.

5.7   TRENDS IN PUBLICATION OUTPUT AND RAW CITATION IMPACT FOR IMI PROJECTS – CALL 3

Figure 5.7.1 shows the publication output and raw citation impact of *Web of Science* publications associated with projects in Call 3.

42

FIGURE 5.7.1 TRENDS IN (A) OUTPUT AND (B) RAW CITATION IMPACT FOR RESEARCH FROM IMI-SUPPORTED PROJECTS IN CALL 3: ABIRISK, BIOVACSAFE, DIRECT, EU-AIMS, MIP-DILI AND PREDICT-TB

Figures 5.7.1A shows that EU-AIMS is the most prolific of the Call 3 projects, however, some six Call 3 projects have accumulated publications in 2013 so far. Of these, ABIRISK, and PreDiCT-TB had no *Web of Science* publications in the second report to IMI.
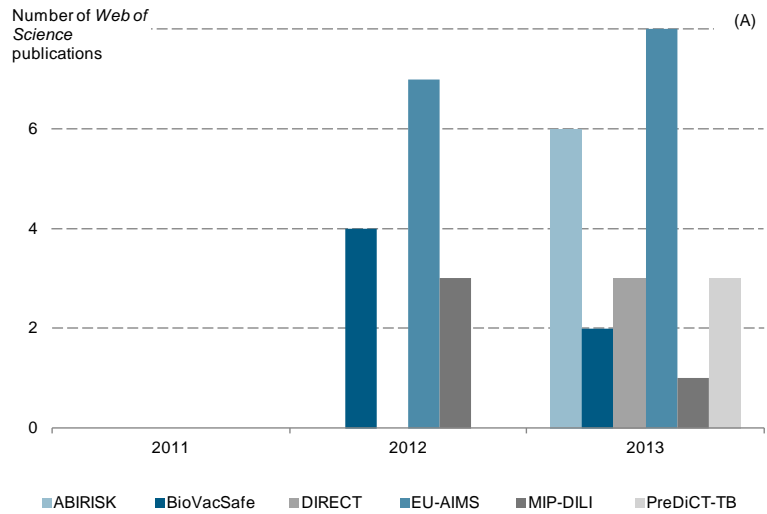
Early indications are that EU-AIMS research is being particularly well-cited (Figure 5.7.1B). This is attributable to two highly-cited papers one of which is defined as a 'hot paper' (Annex 3)
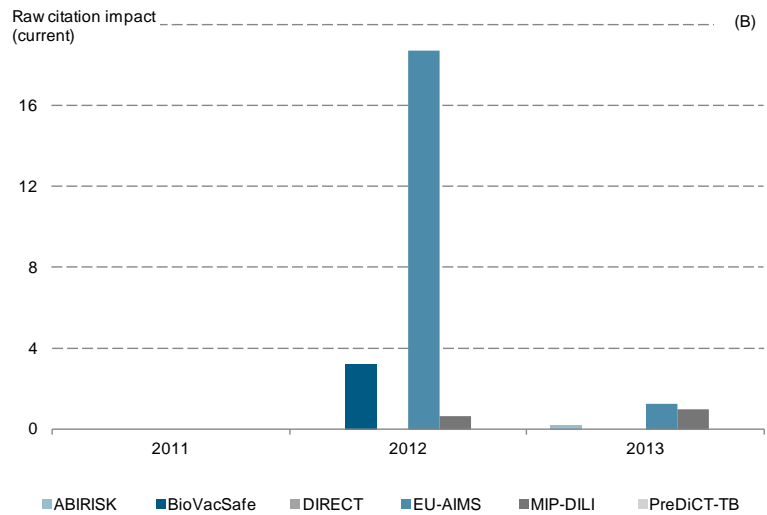
EUPATI is not included in the above analyses as it has not yet generated publications.

6 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.

6.1 PUBLICATIONS BY IMI-SUPPORTED RESEARCHERS

Publications by IMI-supported researchers were identified using researcher names, projects and affiliations supplied by IMI. For this report, data and analyses are limited to those 4 603 researchers associated with 36 of the projects funded by the first four IMI funding calls (Calls 1 to 4).[19]

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

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

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

6.2 CITATION DATA FOR PUBLICATIONS BY IMI-SUPPORTED RESEARCHERS

A total of 23 915 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 6.2.1.

Citation counts for these 23 915 publications have been sourced from the citation databases which underlie Thomson Reuters *Web of Science* and were extracted at end-August 2013.

---

[19] No researcher names were provided for COMPACT (Call 4) or projects in Calls 5 and 6 (ELF, TRANSLOCATION and COMBACTE).

FIGURE 6.2.1 IDENTIFYING PUBLICATIONS BY IMI-SUPPORTED RESEARCHERS WITH THOMSON REUTERS CITATION DATA

IMI-associated authors	• 4 982 names associated with Calls 1 to 4-funding supplied by IMI staff
Data processing	• 4 603 unique individuals • 617 unique institutions
Publications (total)	• 23 915 unique publications • Attributed to 2 438 individuals • 2 165 researchers (47.0%) with no publications found
Dataset for analyses	• 18 793 papers (articles and reviews; 78.6%) • 5 122 other document types (e.g. meeting abstract, editorial, letter; 21.4%)
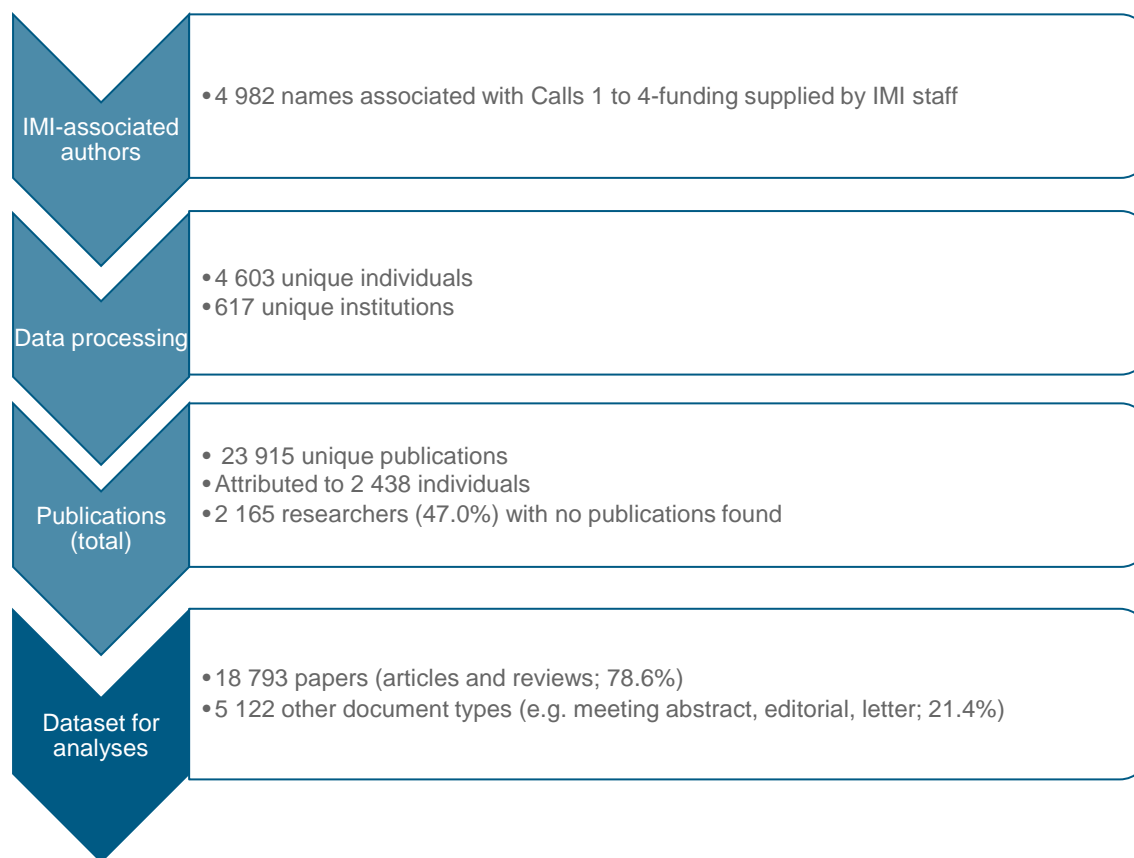
6.3   BIBLIOMETRIC INDICATORS FOR IMI-SUPPORTED RESEARCHERS: PRODUCTIVITY

Publication output, as previously, is higher for IMI-supported researchers based in academic institutions and other research environments (Table 6.3.1).

Overall, 58.3% of researchers have at least one publication.

TABLE 6.3.1 PRODUCTIVITY: PUBLICATION OUTPUT, OVERALL AND BY SECTOR

	Number of researchers		% researchers with publications
Sector	**Total**	**With publications**	
Academia	1 857	1 172	63.1%
Corporate	1 544	880	57.0%
Patient Organisation	62	20	32.3%
Regulatory Agency	65	25	38.5%
Research (other)	761	457	60.1%
Small Medium Enterprise	326	122	37.4%
No assignment	54	23	42.6%
Total researchers	**4 603**	**2 685**	**58.3%**

6.4   BIBLIOMETRIC INDICATORS FOR IMI-SUPPORTED RESEARCHERS: RESEARCH PERFORMANCE

The bibliometric indicators presented in Table 6.4.1 have been calculated for each individual IMI-supported researcher and aggregated by sector.

Since the second report, researchers associated with patient organisations have continued to publish well-regarded research. Six (30.0%) have published at least one 'hot' paper (defined in Section 3.1.3), five (25.0%) have an h-index of at least 10 and 18 (90.0%) have published exclusively in top-quartile journals.

Of the 1 104 publishing academic-based researchers, 212 researchers (19.2%) have published at least one 'hot paper', 217 (19.7%) have an h-index of at least 10 and 949 (86.0%) have published most frequently in top quartile journals.

Similarly, researchers based in other research environments have also published research which has performed well. Ninety of these researchers (20.6%) have published a minimum of one 'hot paper', 68 researchers (15.6%) have h-index of at least 10 and 373 (85.4%) have published in top quartile journals more frequently than in less well-regarded journals.

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

TABLE 6.4.1 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	%
Academia	1 857	1104	212	19.2%	217	19.7%	949	86.0%
Corporate	1 544	764	55	7.2%	18	2.4%	568	74.3%
Patient Organisation	62	20	6	30.0%	5	25.0%	18	90.0%
Regulatory Agency	65	23	1	4.3%	0	0.0%	12	52.2%
Research (other)	761	437	90	20.6%	68	15.6%	373	85.4%
Small Medium Enterprise	326	114	17	14.9%	8	7.0%	89	78.1%
No assignment	54	23	3	13.0%	1	4.3%	22	95.7%
Total researchers	**4 603**	**2 438**	**375**	**15.4%**	**310**	**12.7%**	**1 992**	**81.7%**

TABLE 6.4.2 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.

Sector	Researchers		Mean Diffusion Index		Maximum Diffusion Index		Mean Citation Velocity		Maximum Citation Velocity	
	Total	Publishing	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Academia	1 857	1 104	0.535	0.130	0.684	0.133	0.011	0.016	0.054	0.104
Corporate	1 544	764	0.510	0.172	0.609	0.183	0.009	0.017	0.022	0.042
Patient Organisation	62	20	0.585	0.082	0.690	0.095	0.056	0.122	0.121	0.234
Regulatory Agency	65	23	0.491	0.197	0.560	0.195	0.008	0.010	0.017	0.025
Research (other)	761	437	0.550	0.120	0.694	0.107	0.014	0.029	0.062	0.123
Small Medium Enterprise	326	114	0.564	0.138	0.659	0.130	0.016	0.028	0.057	0.132
No assignment	54	23	0.537	0.141	0.640	0.139	0.015	0.016	0.038	0.063
Total researchers	**4 603**	**2 438**	**0.531**	**0.145**	**0.660**	**0.152**	**0.012**	**0.023**	**0.045**	**0.097**

6.5 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 to identify opportunities to foster collaboration.

Overall, 2 685 researchers (58.3% of 4 603 IMI researchers in total) published any documents that were indexed in *Web of Science*. Over two-thirds of these researchers (N=1 851, 68.9% of 2 685) collaborated (co-authored) with at least one other IMI researcher during the period January 2007-August 2013.

The frequency of collaborative activities are shown over the period of January 2007 to August 2013 by year in Figure 6.5.1 and Table 6.5.1 and further illustrated among researchers in Figure 6.5.2.

TABLE 6.5.1 COLLABORATIVE ACTIVITY BY YEAR – PUBLICATIONS FROM IMI-SUPPORTED RESEARCHERS

	2007	2008	2009	2010	2011	2012	2013
Publications	64	743	859	1 004	1 310	1 519	764
Within-Sector Collaborations	69	562	834	1 160	1 533	1 595	1 058
Cross-Sector Collaborations	22	246	599	882	1 133	1 102	666
% Cross-Sector	24.2%	30.3%	41.5%	43.0%	41.6%	40.6%	38.6%

FIGURE 6.5.1 DISTRIBUTION OF COLLABORATIVE ACTIVITIES BY YEAR

(A) Number of publications co-authored by two or more IMI researchers by year. (B) Collaborations defined as distinct researcher dyads within and across sectors appearing on one or more publications during the given publication year.
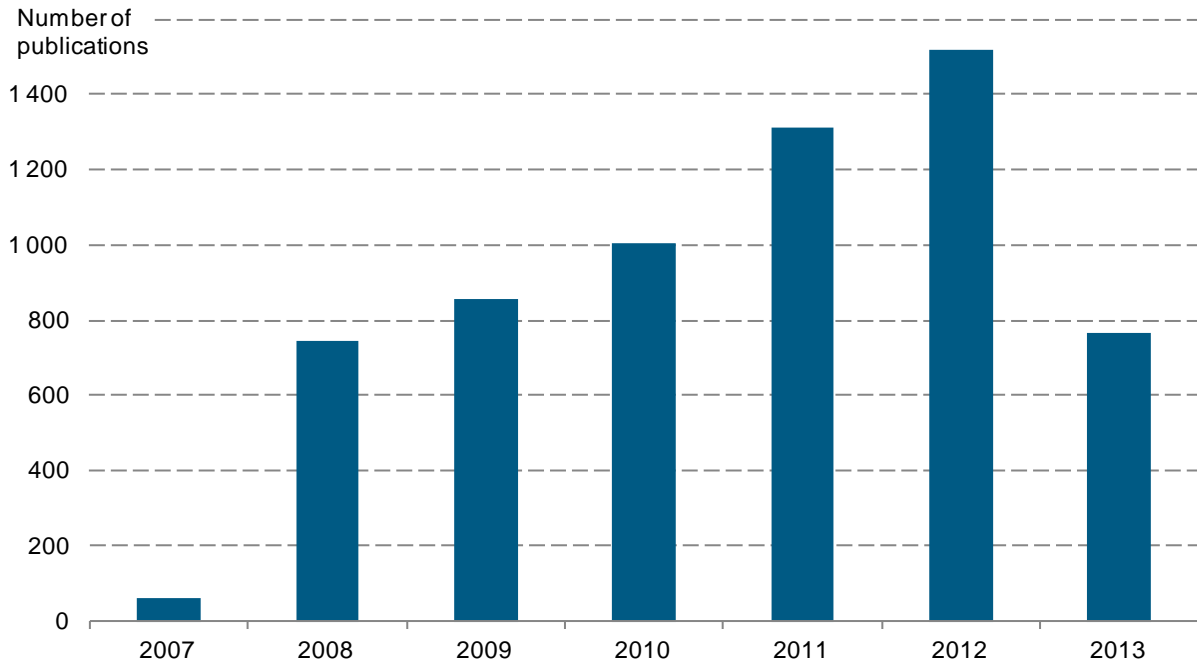
The number of individual researchers with collaborative activity has increased over time from 115 in 2007, to 610 in 2008, 772 in 2009, 841 in 2010, 1 056 in 2011, 1 125 in 2012, and 817 in 2013 up to mid-August.
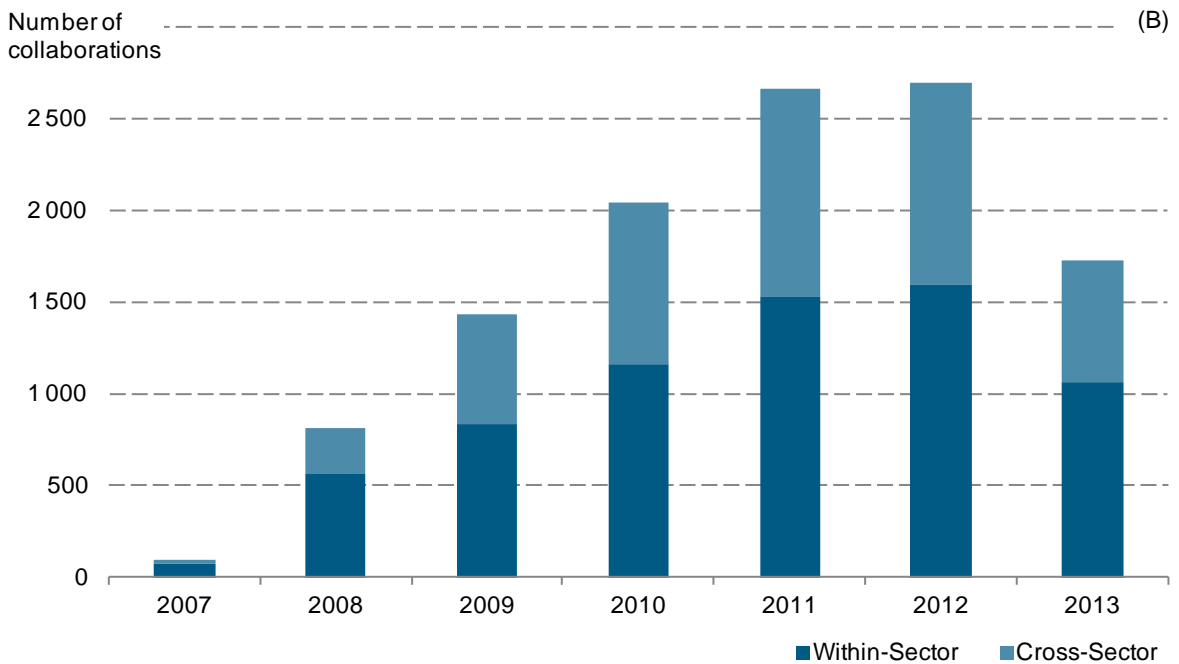
FIGURE 6.5.2 MAP OF 1 828 IMI PROJECT RESEARCHERS WHO HAVE CO-AUTHORED WITH AT LEAST ONE OTHER RESEARCHER WITHIN THE NETWORK BASED ON CO-AUTHORSHIP ACTIVITIES FROM JANUARY 2007 TO AUGUST 2013.

Each individual is represented as a single node coloured with respect to the community of researchers they are based in. Ties between researchers are instances where co-authorship has occurred in a published work. The largest group of inter-connected researchers is composed of 29 communities of which the 4 largest are shown in shaded ovals. Graphics produced using Gephi, applying the Force Atlas 2 layout.[20] Communities identified using a resolution of 1.[21] Isolated communities not connected to the main group of researchers are shown in black.
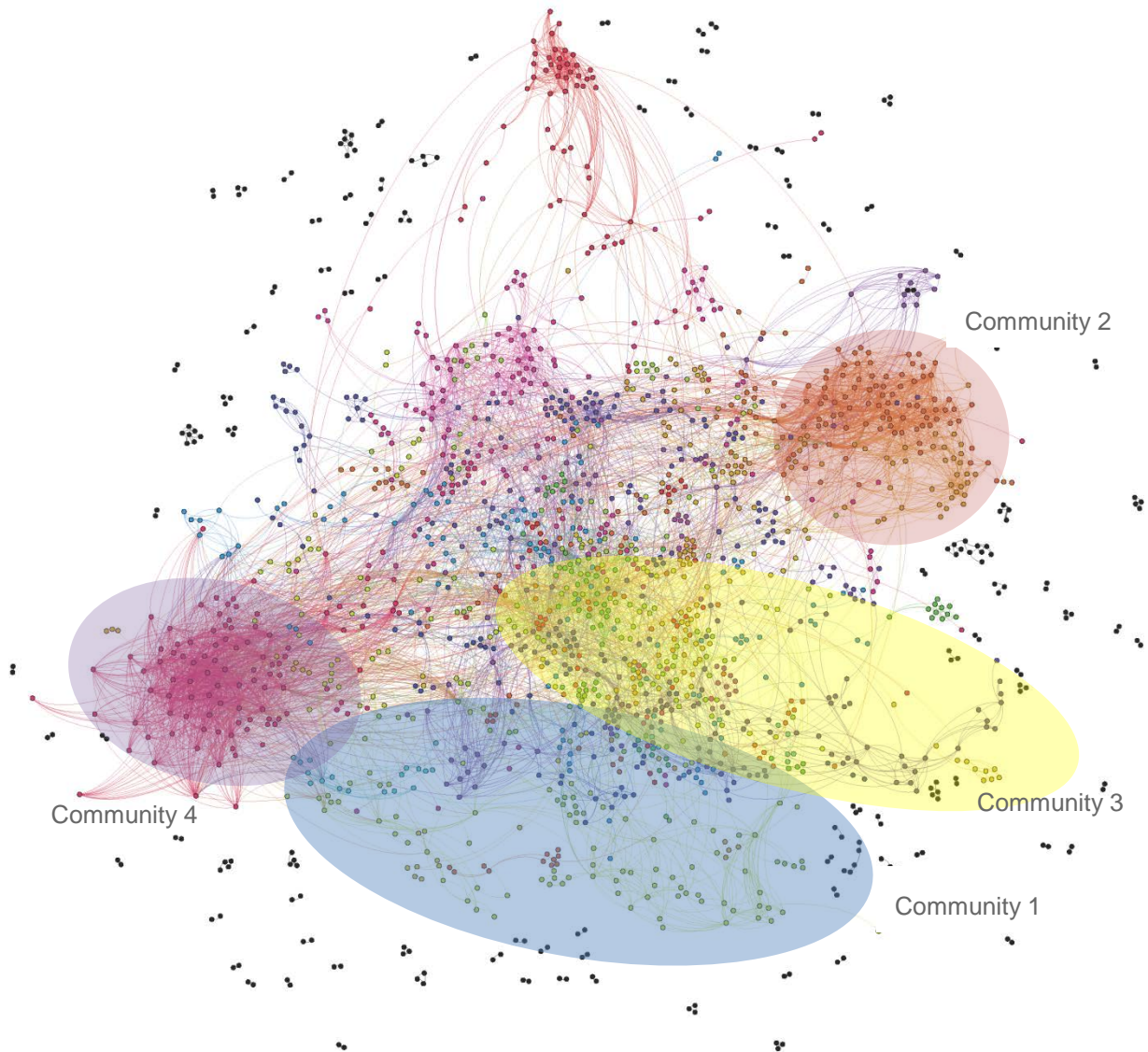
[20] Jacomy, M. (2009). Force-Atlas Graph Layout Algorithm. *URL: http://gephi.org/2011/forceatlas2-the-new-version-of-our-home-brew-layout/*

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

6.6 COLLABORATION BETWEEN IMI-SUPPORTED RESEARCHERS AT SECTOR LEVEL

TABLE 6.6.1 DISTRIBUTION OF SECTORS BY NUMBER OF RESEARCHERS WITHIN SELECT COMMUNITIES BASED ON CO-AUTHORSHIP ACTIVITIES FROM JANUARY 2007 - AUGUST 2013.

Ninety-eight isolated communities exist composed of between 2 and 11 researchers each. The largest group of inter-connected researchers (N=1 581 researchers) is composed of 29 distinct communities the largest four of which are described at below.

Sector	Isolated Communities		Connected Communities		Connected Community							
					1		2		3		4	
	N	%	N	%	N	%	N	%	N	%	N	%
No assignment	1	0.4%	13	0.8%	1	0.7%	--	--	2	1.6%	3	2.6%
Academic	86	35.4%	797	51.5%	72	51.1%	94	66.7%	37	29.6%	83	70.9%
Corporate	106	43.6%	342	22.1%	33	23.4%	22	15.6%	51	40.8%	13	11.1%
Patient Organisation	--	--	13	0.8%	2	1.4%	1	0.7%	--	--	1	0.9%
Regulatory Agency	2	0.8%	10	0.6%	--	--	--	--	1	0.8%	--	--
Research (other)	33	13.6%	314	20.3%	28	19.9%	24	17.0%	31	24.8%	15	12.8%
Small Medium Enterprise	15	6.2%	58	3.7%	5	3.5%	--	--	3	2.4%	2	1.7%
Multiple sectors	4	1.6%	34	2.2%	11	7.2%	1	0.7%	1	0.8%	1	0.8%
Total	**247**		**1 581**		**152**		**142**		**126**		**118**	

The largest component, shown at the centre of Figure 6.5.2 and defined as groups of researchers where all individuals are connected with one another directly or indirectly via other IMI researchers, consisted of 1 581 researchers representing all six sectors (Table 6.6.1). Within this set of researchers, 29 communities were identified within which there are more frequent and closely inter-related co-authorship activities. The largest four of these 29 communities are shown enclosed by coloured ovals. A complete depiction of all 29 communities can be found in Figure 6.6.1D.

The largest community, shown in the blue oval in Figure 6.5.2, is composed of individuals who are all disparately positioned suggesting that while there might be high collaborative activity between researchers, collaborative activity between co-authors of a given researcher's co-authors is less strong. This group is largely composed of academic researchers (51.1%, Table 6.6.1).

Many of the co-authorship activities in Community 2 (shown in red in Figure 6.5.2) are associated with Rheumatoid Arthritis research (Figure 6.6.1C) with many researchers involved in the BTCure project. Other projects associated with this group include ABIRISK and RAPP-ID. Overall, the composition of the group reflects a distribution of researchers more skewed towards academics than in Community 1 (Table 6.6.1).

Community 3 (shown in yellow in Figure 6.5.2) while fairly diffuse, is dominated by researchers working on the Central Nervous System, especially those involved with the Pharma-Cog project. In contrast, Community 4 (highlighted in purple in the same Figure) is more tightly focussed in Diabetes but less constrained to any one project. The most common projects associated with this Community are SUMMIT and DIRECT, but these still only comprise 43.2% (N=51 or 118) of its researchers.

While the majority of publishing researchers are connected to one another and are in the main connected component, 13.5% of collaborating researchers (N=247 of 1 828) collaborate within isolated communities composed of between 2 and 11 researchers. Ninety-eight isolated groups exist (shown on the periphery of Figure 6.5.2), of which 80 (81.6%) are composed of researchers from only one sector.

The main component includes researchers from 371 distinct organisations, 34.0% (N=126) of which span across communities. Within this set there are 167 academic organisations, 34 corporate organisations, 7 patient organisations, 3 regulatory agencies, 94 research/other entities, and 32 small medium enterprise organisations. The two entities which span the most communities in this main component are Astra Zeneca (corporate) and Karolinska Institutet (academia), a situation unchanged from the second report. Overall, these two organisational affiliations include only 6.9% (N=109 of 1 581) of researchers in the main component, and both cover 14 communities.

Co-authorship is more common among researchers in the same sector than among researchers in different sectors (Figure 6.6.1B). This is expected given the principle of homophily which suggests that individuals are more likely to interact with individuals who are like them.[22] However, there are substantial co-authorship activities among researchers from different sectors (Figure 6.6.1B). Of a total of 6 760 distinct co-authorship relationships, 2 820 are cross-sector and involve 1 059 total researchers from all 6 sectors. This accounts for 41.7% of all co-authorship activities during the analysis period, a small increase from 39.9% in the second report.

The same is true of co-authorship activities by project. The majority of collaborative relationships are among researchers associated with the same project with 2 497 of 6 760 of co-authorship relationships (36.9%) being cross-project, an increase from 25.5% in the second report.

6.6.1 STRENGTHS AND LIMITATIONS

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

Researchers with multiple organisational affiliations and/or multiple sectors were assigned to a single organisational affiliation, and sector based on the judgment of IMI personnel. No similar assignment

---

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

was made for investigators affiliated with multiple projects – that is, multiple project affiliations were preserved for all investigators who were involved with multiple projects.

FIGURE 6.6.1 THE LARGEST GROUP OF INTER-CONNECTED IMI RESEARCHERS IS COMPOSED OF 1 581 RESEARCHERS FROM 36 PROJECTS AND ALL SIX SECTORS.

(A) Researchers with any collaborative activity are shown coloured by sector. (B) Researchers are coloured by community. Isolated communities outside of the main connected component are independent of one another and all shown in black. (C) Researchers are shown coloured by disease area. (D) Researchers in the main component are shown coloured by community.
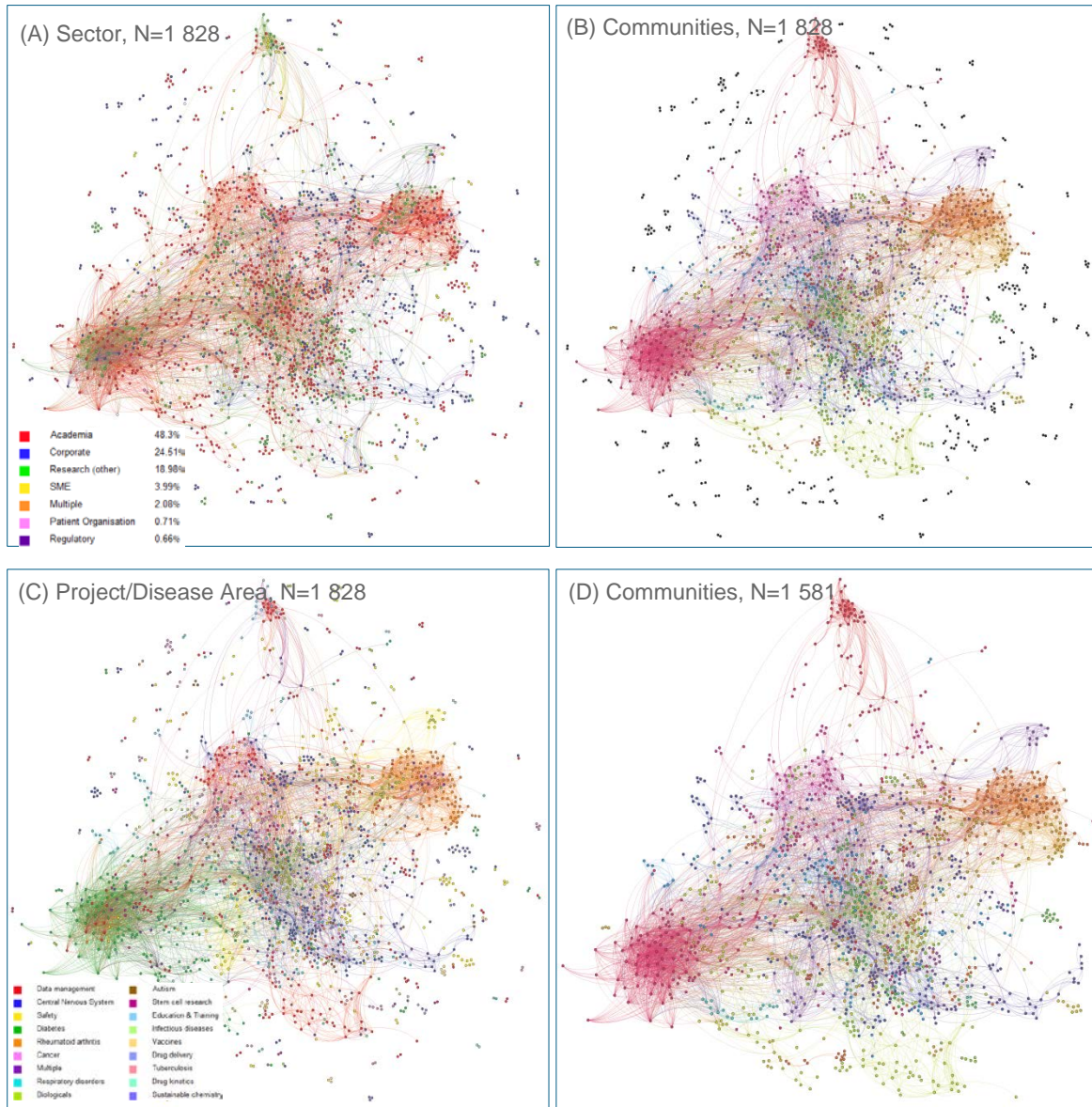
TABLE 6.6.2 ORGANISATIONS AND THE NUMBER OF ASSOCIATED COMMUNITIES AND RESEARCHERS WITHIN THE MAIN INTER-CONNECTED COMPONENT

In all, 570 organisations with collaborative co-authorship activity were identified, of which 421 (73.9%) span two or more communities and 327 (57.4%) span three or more communities. The top 25 organisations by number of communities which they span, is shown below.

Organisation	Sector	Number of communities	Number of researchers
GSK	Corporate	129	144
Sanofi-Aventis	Corporate	118	134
Pfizer	Corporate	110	145
Astra Zeneca	Corporate	106	146
Janssen	Corporate	94	129
Roche	Corporate	93	132
Novartis	Corporate	68	88
Eli Lilly	Corporate	64	82
Merck	Corporate	56	59
Bayer	Corporate	55	57
INSERM	Research (other)	47	57
Boehringer Ingelheim	Corporate	46	57
Karolinska Institutet	Academic	45	93
Novo Nordisk	Corporate	36	37
University of Oxford	Academic	34	51
Lundbeck	Corporate	34	37
Max-Planck-Gesellschaft	Research (other)	31	44
Amgen	Corporate	31	31
Servier	Corporate	31	34
Universiteit Utrecht	Academic	26	33
Orion Pharma	Corporate	25	25
King's College London	Academic	25	45
UCB	Corporate	24	25
University College London	Academic	23	39
University of Lille	Academic	23	28

6.7   MAPPING COLLABORATION AMONG IMI-SUPPORTED RESEARCHERS

6.7.1   PUBLICATION DATASETS

The analysis in this section uses subsets of publications from the dataset of IMI researcher publications described in Figure 6.2.1. 'Pre-IMI' (before funding award) and 'post-IMI' (after funding award) are defined for each project by the date on which IMI funding was formally awarded.

- 'pre-IMI' publications have been defined as all publications from the associated researchers, published between January 2007 and the date of IMI funding award.

- 'post-IMI' publications have been defined as all publications from the associated researchers, published between the date of IMI funding award and August 2013.

- Where a particular IMI-supported researcher is associated with more than one IMI project, 'pre-IMI' and 'post-IMI' have been defined by the start date of the earliest project.

6.7.2   GEOGRAPHICAL MAPS OF COLLABORATION

Figures 6.7.1 and 6.7.2 provide geographical maps of collaboration among IMI-supported researchers in Calls 1 to 6. These maps show:

- The number of publications by country (shaded from white to dark blue by total output). Countries with no contributing output are shaded in grey).

- The frequency of collaboration between collaborating partners. Lines connecting partners indicate co-authorship with thicker lines indicating a higher number of co-authored papers.

- The nature of collaboration between collaborating partners. If both collaborating partners are from the same sector, the line is black; if they are from different sectors, the line is orange. At the request of IMI, academia and research (other) are treated as a single combined sector (academic) for this particular analysis. This allows collaboration between private and public bodies to be highlighted more clearly as otherwise collaboration between academia and research (other) would dominate the maps.

- The location of the collaboration (red dot). Institutions were mapped at city-level using author address data from *Web of Science*.

Together, Figures 6.7.1 and 6.7.2 show that collaboration has increased as successive IMI projects have begun and that there has been a small increase in cross-sector collaborative activity.

Most collaboration has occurred within the combined academic sector (defined as above). Some projects, however, have driven extensive collaboration between research institutes and universities e.g. for SUMMIT, between National Institute for Health and Welfare in Helsinki and both the University of Oxford and Imperial College London.

There are also some strong cross-sector links, predominantly associated with GlaxoSmithKline and Pfizer in the UK and Janssen in Belgium, who are participants in a number of different projects.

There has been an increase in both types of collaborative activity (within-sector and cross-sector) between the 'pre-IMI' to 'post-IMI' periods. This means that there is less of a change in the overall cross-sector collaboration rate than may be inferred from comparing the maps; while there are some strong orange lines on the second map, there are correspondingly more black lines. In the 'pre-IMI' dataset, cross-sector collaboration was 16.5% (389 of 2 362 co-authorship pairs); in the 'post-IMI' dataset, this was 23.9% (1 381 of 5 772 co-authorship pairs).[23]

Overall, these analyses point towards increased collaboration among IMI-supported researchers, though it should be noted that researcher mobility during the time period or the recently observed global trend towards increased collaboration is not accounted for.[24]

---

[23] These figures are lower than those given earlier in section 6.6 because collaboration between research institutes and academic institutions are treated as within-sector for this analysis.

[24] Adams J (2013). Collaborations: the fourth age of research. Nature, 497, 557-560.

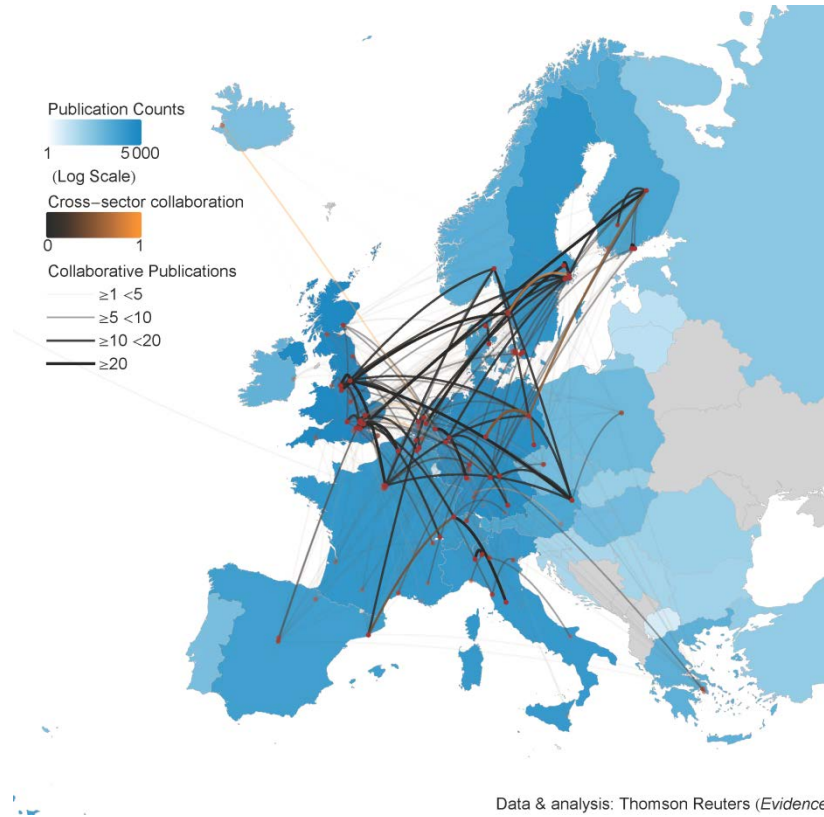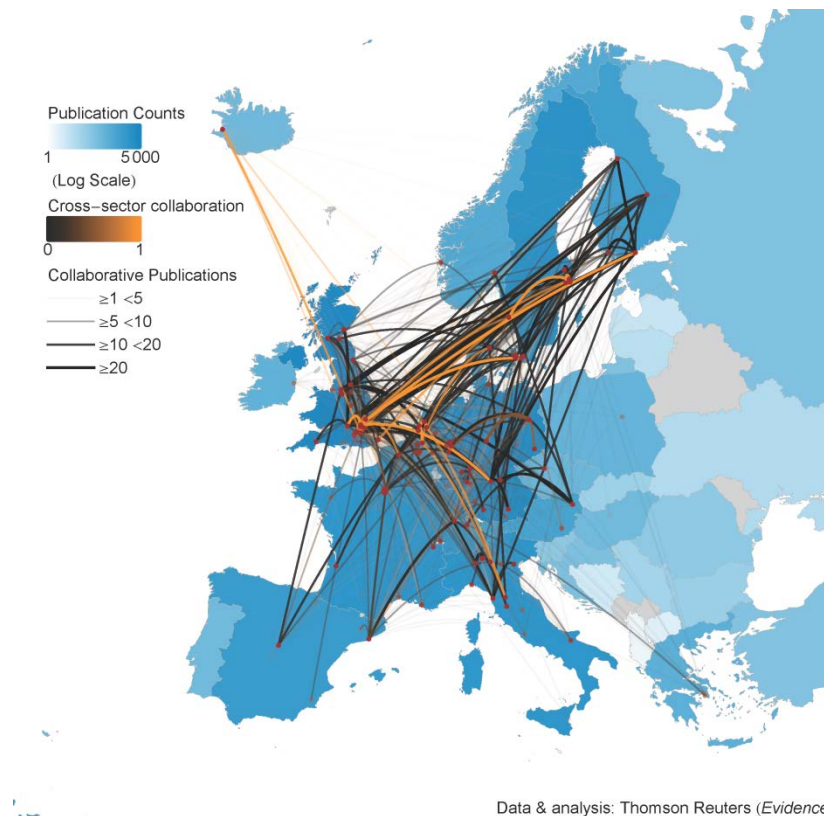FIGURE 6.7.1 COLLABORATIVE RESEARCH LINKS, IMI-SUPPORTED RESEARCHERS, PRE-IMI FUNDING AWARD

Data & analysis: Thomson Reuters (*Evidence*)

FIGURE 6.7.2 COLLABORATIVE RESEARCH LINKS, IMI-SUPPORTED RESEARCHERS, POST-IMI FUNDING AWARD

Data & analysis: Thomson Reuters (*Evidence*)

ANNEX 1: SUMMARY OF NEW IMI-SUPPORTED PUBLICATIONS

This Annex presents summary analyses of IMI publications identified since the second report to IMI in February 2013. These summary analyses should be borne in mind when considering IMI project research analyses presented in Sections 4 and 5 of the report.

A1.1 SUMMARY OF NEW IMI-SUPPORTED PUBLICATIONS – OUTPUT

A total of 164 new IMI-supported publications were identified for this report (as outlined in Section 4.1, 155 of which (94.6%) were abstracted in *Web of Science*. In addition to these 155, 12 new matches were identified in *Web of Science* for previously unmatched records, making 167 new publications linked to citation data.

- Most of this research was published in 2013 (Figure A1.1.1A).

- With the exception of ORBITO, no publications were found for projects in Calls 4 to 6.

- Two pre-2012 publications not found previously are new to the dataset. These are both associated with the Call 1 project SafeSciMet.

     o One publication is an editorial and has not yet been cited. The other publication is an article and has received 2 citations.

- The categorisation of these new publications by document type is very similar to that of the overall dataset (Section 4.2) with over 95% of new publications classified as articles and reviews (Figure A1.1.1B).

FIGURE A1.1.1 SUMMARY OF NEW IMI-SUPPORTED PUBLICATION OUTPUT

(A) Number of new *Web of Science* publications by year and call
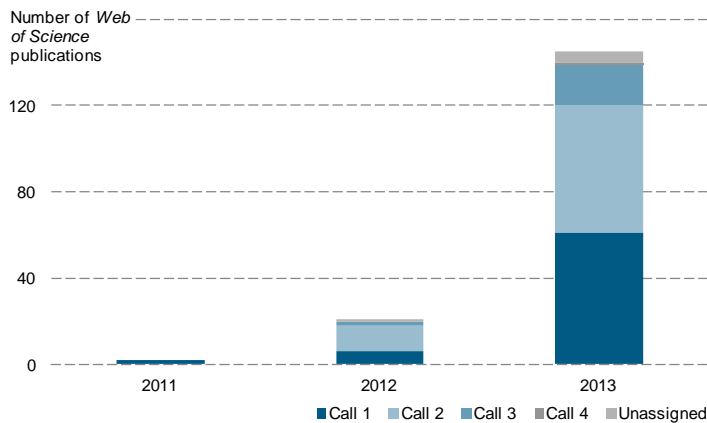
(B) Categorisation of new *Web of Science* publications by document type
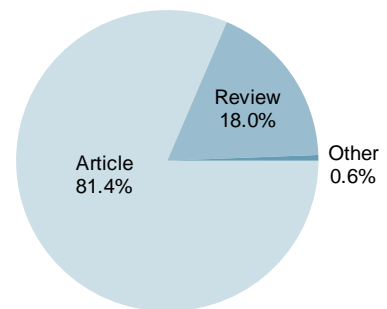
The additional publications have extended the range of journals in which IMI project research is published. Table A1.1.1 shows all new journals used more than once while Table A1.1.2 presents the top ten new journals with highest Journal Impact Factor.

TABLE A1.1.1 NEW JOURNALS IN WHICH NEW IMI PROJECT PUBLICATIONS HAVE BEEN PUBLISHED MOST FREQUENTLY (2009-2013)

Journal	Number of *Web of Science* publications	Number of papers	Journal Impact Factor (2012)	*Web of Science* journal categories
Revista de Psiquiatria Clinica	3	3	0.633	Psychiatry
Critical Care	2	2	4.718	Critical Care Medicine
European Journal of Immunology	2	2	4.97	Immunology
European Journal of Pharmaceutical Sciences	2	1	2.987	Pharmacology & Pharmacy
Gastroenterology	2	2	12.821	Gastroenterology & Hepatology
Genome Medicine	2	2	3.397	Genetics & Heredity
Journal of Bone and Mineral Research	2	2	6.128	Endocrinology & Metabolism
Nature Medicine	2	2	22.864	Biochemistry & Molecular Biology; Cell Biology; Research & Experimental Medicine
Neuroscience	2	2	3.122	Neurosciences

TABLE A1.1.2 TOP TEN NEW JOURNALS IN WHICH NEW IMI PROJECT PUBLICATIONS HAVE BEEN PUBLISHED, RANKED BY JOURNAL IMPACT FACTOR

Journal	Number of *Web of Science* publications	Number of papers	Journal Impact Factor (2012)	*Web of Science* journal categories
Nature Medicine	2	2	22.864	Biochemistry & Molecular Biology; Cell Biology; Research & Experimental Medicine
Pharmacological Reviews	1	1	22.345	Pharmacology & Pharmacy
Gastroenterology	2	2	12.821	Gastroenterology & Hepatology
Hepatology	1	1	12.003	Gastroenterology & Hepatology
Genome Biology	1	1	10.288	Biotechnology & Applied Microbiology; Genetics & Heredity
Nature Communications	1	1	10.015	Multidisciplinary Sciences
Trends In Genetics	1	1	9.772	Genetics & Heredity
Blood	1	1	9.06	Haematology
Autoimmunity Reviews	1	1	7.975	Immunology
Pharmacology & Therapeutics	1	1	7.793	Pharmacology & Pharmacy

These journal titles encompass a broad range of research disciplines and confirm the cross-disciplinary nature of IMI project research. Furthermore, in the last 6 months, IMI projects have produced publications in elite journals such as Nature Medicine (BTCure, EU-AIMS) and high impact field-specific journals such as Gastroenterology (BTCure).

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

Figure A1.2.1 shows the top *Web of Science* journal categories associated with new IMI project publications. Only categories with at least 6 publications are included.

FIGURE A1.2.1 TOP *WEB OF SCIENCE* JOURNAL CATEGORIES IN WHICH NEW IMI PROJECT PUBLICATIONS ARE PUBLISHED

Number of *Web of Science* publications
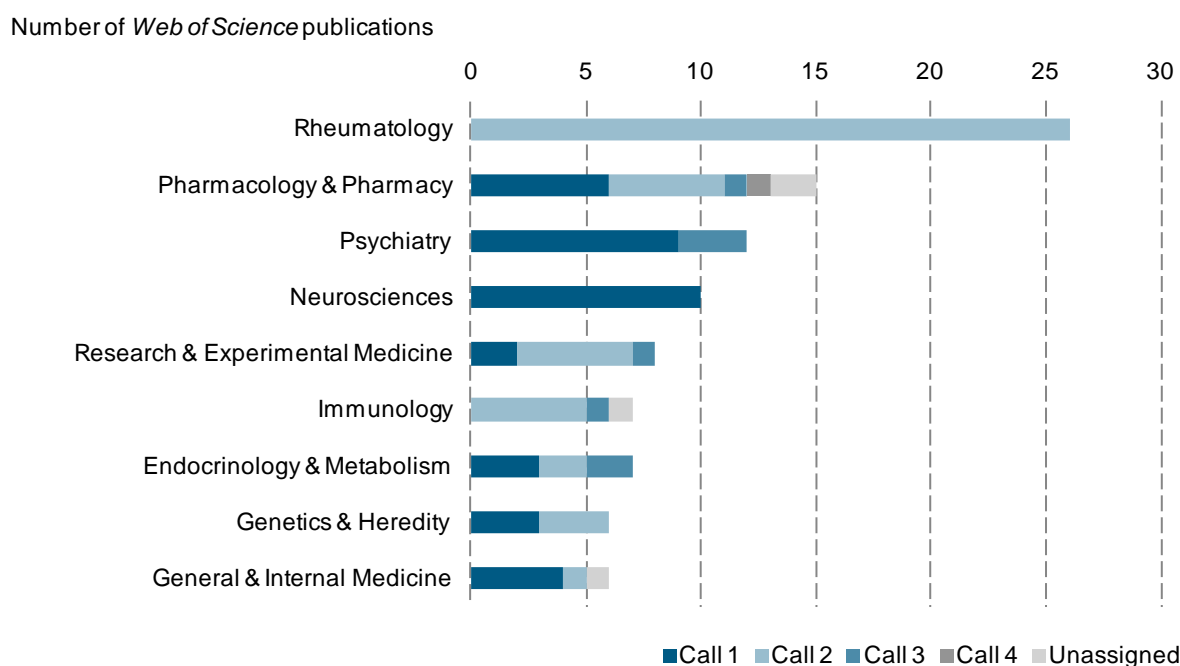
One category, General & Internal Medicine is new to this report. On the whole, these categories overlap with the top ten most frequently-used journal categories for IMI project research overall (Section 4.5), however, there are some differences in category usage between new IMI project publications and the overall IMI projects portfolio.

Psychiatry is the third most frequent category for new publications (7.2% share) but is seventh overall. It is associated primarily with the Call 1 project NEWMEDS with a smaller number of publications from EU-AIMS. Clinical Neurology (primarily associated with EUROPAIN and Pharma-Cog) is fifth in the overall IMI portfolio but no new publications are associated with this category

A1.2 SUMMARY OF NEW IMI-SUPPORTED PUBLICATIONS – CITATIONS

A summary of new IMI-supported publications is shown in Table A1.2.1. Although these publications are relatively recent, over one-quarter (28.7%) have already been cited.

TABLE A1.2.1 SUMMARY INDICATORS FOR NEW IMI PROJECT PUBLICATIONS

Call	Project	Number of IMI publications	Number of *Web of Science* publications	Number of cited *Web of Science* publications	Total citations (current)
n/a	(unassigned)	6	6	2	2
1	eTOX	7	7	2	4
1	EUROPAIN	13	13	5	9
1	IMIDIA	5	4	1	1

Call	Project	Number of IMI publications	Number of *Web of Science* publications	Number of cited *Web of Science* publications	Total citations (current)
1	MARCAR	10	9	3	11
1	NEWMEDS	25	20	4	11
1	Pharma-Cog	4	4	1	1
1	PROTECT	4	4	0	0
1	SafeSciMET	3	2	1	2
1	SUMMIT	5	5	2	2
1	U-BIOPRED	1	1	1	1
2	BTCure	44	44	15	20
2	DDMoRe	5	5	0	0
2	Onco Track	5	5	1	1
2	Open PHACTS	6	5	2	7
2	Predect	2	2	0	0
2	Quic-Concept	3	3	0	0
2	RAPP-ID	7	7	2	4
3	ABIRISK	6	6	1	1
3	BioVacSafe	2	2	0	0
3	DIRECT	2	2	0	0
3	EU-AIMS	7	7	4	8
3	MIP-DILI	1	1	1	1
3	PreDiCT-TB	3	3	0	0
4	ORBITO	1	1	0	0
	Overall	**176**	**167**	**48**	**86**

Of the 30 IMI projects in Calls 1 to 3, 16 had no new *Web of Science* publications (EMTRAIN, EU2P, Pharmatrain, PRO-active, SAFE-T, EHR4CR, EUPATI) and are not listed in Table A1.2.1. These projects had no *Web of Science* publications in the second report.

Calls 4 to 6 are newly included in the IMI projects dataset which has previously focused on projects funded through Calls 1 to 3. With the exception of ORBITO, no publications were found for the ten projects in Calls 4 to 6.

ANNEX 2: DEFINITION AND SCOPE OF *WEB OF SCIENCE* JOURNAL CATEGORIES

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

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.

General & Internal Medicine includes journals on medical specialties such as general medicine, internal medicine, clinical physiology, pain management, military and hospital medicine. Journals focusing on family medicine and primary health care services are placed in the Primary Health Care category.

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

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

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.

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

This Annex considers the cumulative dataset of IMI project publications that have been linked to records in Thomson Reuters citation databases.

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 A3.1 lists the 52 papers in the IMI project publications dataset that have been identified as highly-cited. This is a large increase over the second report to IMI where ten papers were classified as highly-cited.

Papers are listed in ascending alphabetical order (project, first author) and those papers also identified as 'hot papers' (Section A3.2) are highlighted (bold text).

A3.1 HIGHLY-CITED PAPERS ASSOCIATED WITH IMI PROJECTS

- BTCure: Cope, A *et al.* (2011) The Th1 life cycle: molecular control of IFN-gamma to IL-10 switching, Trends in Immunology, 32: 278-286, doi: 10.1016/j.it.2011.03.010.

- BTCure: Gerlag, DM *et al.* (2012) EULAR recommendations for terminology and research in individuals at risk of rheumatoid arthritis: report from the Study Group for Risk Factors for Rheumatoid Arthritis, Annals of the Rheumatic Diseases, 71: 638-641, doi: 10.1136/annrheumdis-2011-200990.

- **BTCure: Harre, U *et al.* (2012) Induction of osteoclastogenesis and bone loss by human autoantibodies against citrullinated vimentin, Journal of Clinical Investigation, 122: 1791-1802, doi: 10.1172/JCI60975.**

- BTCure: Nikitopoulou, I *et al.* (2012) Autotaxin expression from synovial fibroblasts is essential for the pathogenesis of modeled arthritis, Journal of Experimental Medicine, 209: 923-931, doi: 10.1084/jem.20112012.

- BTCure: Willemze, A *et al.* (2012) The ACPA recognition profile and subgrouping of ACPA-positive RA patients, Annals of the Rheumatic Diseases, 71: 268-274, doi: 10.1136/annrheumdis-2011-200421.

- **eTOX: Arighi, CN *et al.* (2011) Overview of the BioCreative III Workshop, BMC Bioinformatics, 12: , doi: 10.1186/1471-2105-12-S8-S1.**

- eTOX: Enoch, SJ *et al.* (2011) A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity, Critical Reviews in Toxicology, 41: 783-802, doi: 10.3109/10408444.2011.598141.

- 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: 10.1176/appi.ajp.2010.09111661.

- eTOX: Przybylak, KR and Cronin MTD (2012) In silico models for drug-induced liver injury - current status, Expert Opinion on Drug Metabolism & Toxicology, 8: 201-217, doi: 10.1517/17425255.2012.648613.

- eTOX: Taboureau, O *et al.* (2011) ChemProt: a disease chemical biology database, Nucleic Acids Research, 39: D367-D372, doi: 10.1176/appi.ajp.2010.09111660.

- eTOX: Vidal, D and Mestres, J (2010) In Silico Receptorome Screening of Antipsychotic Drugs, Molecular Informatics, 29: 543-551, doi: 10.1002/minf.201000055.

- **EU-AIMS: Kong, A *et al.* (2012) Rate of de novo mutations and the importance of fathers age to disease risk, Nature, 488: 471-475, doi: 10.1038/nature11396.**

- EU-AIMS: Stein, JL *et al.* (2012) Identification of common variants associated with human hippocampal and intracranial volumes, Nature Genetics, 44: 552-+, doi: 10.1038/ng.2250.

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

- EUROPAIN: Andersen, KG and Kehlet H (2011) Persistent Pain After Breast Cancer Treatment: A Critical Review of Risk Factors and Strategies for Prevention, Journal of Pain, 12: 725-746.

- EUROPAIN: Andersen, KG *et al.* (2012) Persistent pain after targeted intraoperative radiotherapy (TARGIT) or external breast radiotherapy for breast cancer: A randomized trial, Breast, 21: 46-49, doi: 10.1016/j.breast.2011.07.011.

- EUROPAIN: Baastrup, C *et al.* (2010) Spinal-, brainstem- and cerebrally mediated responses at- and below-level of a spinal cord contusion in rats: Evaluation of pain-like behavior, Pain, 151: 670-679, doi: 10.1016/j.pain.2010.08.024.

- **EUROPAIN: Finnerup, NB *et al.* (2010) The evidence for pharmacological treatment of neuropathic pain, Pain, 150: 573-581, doi: 10.1016/j.pain.2010.06.019.**

- EUROPAIN: Hauser, W *et al.* (2012) The Role of Antidepressants in the Management of Fibromyalgia Syndrome A Systematic Review and Meta-Analysis, CNS Drugs, 26: 297-307.

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

- EUROPAIN: 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.

- EUROPAIN: Serra, J *et al.* (2012) Microneurographic identification of spontaneous activity in C-nociceptors in neuropathic pain states in humans and rats, Pain, 153: 42-55, doi: 10.1016/j.pain.2011.08.015.

- EUROPAIN: Westermann, A *et al.* (2012) Different underlying pain mechanisms despite identical pain characteristics: A case report of a patient with spinal cord injury, Pain, 153: 1537-1540, doi: 10.1016/j.pain.2012.02.031.

- EUROPAIN: Wildgaard, K *et al.* (2012) Quantitative sensory testing of persistent pain after video-assisted thoracic surgery lobectomy, British Journal of Anaesthesia, 108: 126-133, doi: 10.1093/bja/aer325.

- IMIDIA: Ravassard, P *et al.* (2011) A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion, Journal of Clinical Investigation, 121: 3589-3597, doi: 10.1172/JCI58447.

- IMIDIA: Roggli, E *et al.* (2010) Involvement of MicroRNAs in the Cytotoxic Effects Exerted by Proinflammatory Cytokines on Pancreatic beta-Cells, Diabetes, 59: 978-986, doi: 10.2337/db09-0881.

- NEWMEDS: Bussey, TJ *et al.* (2012) New translational assays for preclinical modelling of cognition in schizophrenia: The touchscreen testing method for mice and rats, Neuropharmacology, 62: 1191-1203, doi: 10.1016/j.neuropharm.2011.04.011.

- NEWMEDS: Gastambide, F *et al.* (2012) Selective Remediation of Reversal Learning Deficits in the Neurodevelopmental MAM Model of Schizophrenia by a Novel mGlu5 Positive Allosteric Modulator, Neuropsychopharmacology, 37: 1057-1066, doi: 10.1038/npp.2011.298.

- NEWMEDS: Gilmour, G *et al*. (2012) NMDA receptors, cognition and schizophrenia - Testing the validity of the NMDA receptor hypofunction hypothesis, Neuropharmacology, 62: 1401-1412, doi: 10.1016/j.neuropharm.2011.03.015.

- 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.

- **NEWMEDS: Jacquemont, S *et al*. (2011) Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus, Nature, 478: 97-U111, doi: 10.1038/nature10406.**

- NEWMEDS: Kargieman, L *et al*. (2012) Clozapine Reverses Phencyclidine-Induced Desynchronization of Prefrontal Cortex through a 5-HT1A Receptor-Dependent Mechanism, Neuropsychopharmacology, 37: 723-733.

- NEWMEDS: Keeler, JF and Robbins TW (2011) Translating cognition from animals to humans, Biochemical Pharmacology, 81: 1356-1366, doi: 10.1016/j.bcp.2010.12.028.

- **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.**

- NEWMEDS: Llado-Pelfort, L *et al*. (2012) 5-HT1A Receptor Agonists Enhance Pyramidal Cell Firing in Prefrontal Cortex Through a Preferential Action on GABA Interneurons, Cerebral Cortex, 22: 1487-1497, doi: 10.1093/cercor/bhr220.

- NEWMEDS: Uher, R *et al*. (2012) Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms, Psychological Medicine, 42: 967-980, doi: 10.1017/S0033291711001905.

- Onco Track: Algar, WR *et al*. (2012) Quantum Dots as Simultaneous Acceptors and Donors in Time-Gated Forster Resonance Energy Transfer Relays: Characterization and Biosensing, Journal of the American Chemical Society, 134: 1876-1891, doi: 10.1021/ja210162f.

- Onco Track: Hildebrandt, N (2011) Biofunctional Quantum Dots: Controlled Conjugation for Multiplexed Biosensors, ACS Nano, 5: 5286-5290, doi: 10.1021/nn2023123.

- **Open PHACTS: Williams, AJ *et al*. (2012) Towards a gold standard: regarding quality in public domain chemistry databases and approaches to improving the situation, Drug Discovery Today, 17: 685-701.**

- **Pharma-Cog: Frisoni, GB *et al*. (2010) The clinical use of structural MRI in Alzheimer disease, Nature Reviews Neurology, 6: 67-77, doi: 10.1038/nrneurol.2009.215.**

- Pharma-Cog: Languille, S *et al*. (2012) The grey mouse lemur: A non-human primate model for ageing studies, Ageing Research Reviews, 11: 150-162, doi: 10.1016/j.arr.2011.07.001.

- PROTECT: Bhaskaran, K *et al*. (2012) Angiotensin receptor blockers and risk of cancer: cohort study among people receiving antihypertensive drugs in UK General Practice Research Database, British Medical Journal, 344, doi: 10.1136/bmj.e2697.

- PROTECT: Gallagher, AM *et al*. (2011) Risk of Death and Cardiovascular Outcomes with Thiazolidinediones: A Study with the General Practice Research Database and Secondary Care Data, PLoS One, 6, doi: 10.1371/journal.pone.0028157.

- PROTECT: van Staa, TP *et al*. (2012) Glucose-lowering agents and the patterns of risk for cancer: a study with the General Practice Research Database and secondary care data, Diabetologia, 55: 654-665, doi: 10.1007/s00125-011-2390-3.

- Quic-Concept : Lambin, P *et al.* (2012) Radiomics: Extracting more information from medical images using advanced feature analysis, European Journal of Cancer, 48: 441-446, doi: 10.1016/j.ejca.2011.11.036.

- Quic-Concept: Nguyen, QD *et al.* (2012) Imaging apoptosis with positron emission tomography: Bench to bedside development of the caspase-3/7 specific radiotracer [F-18]ICMT-11, European Journal of Cancer, 48: 432-440, doi: 10.1016/j.ejca.2011.11.033.

- Quic-Concept: Soloviev, D *et al.* (2012) [F-18]FLT: An imaging biomarker of tumour proliferation for assessment of tumour response to treatment, European Journal of Cancer, 48: 416-424, doi: 10.1016/j.ejca.2011.11.035.

- RAPP-ID: Barnkob, R *et al.* (2012) Measuring acoustic energy density in microchannel acoustophoresis using a simple and rapid light-intensity method, Lab on a Chip, 12: 2337-2344, doi: 10.1039/c2lc40120g.

- SUMMIT: Boekholdt, SM *et al.* (2012) Association of LDL Cholesterol, Non-HDL Cholesterol, and Apolipoprotein B Levels With Risk of Cardiovascular Events Among Patients Treated With Statins A Meta-analysis, JAMA-Journal of the American Medical Association, 307: 1302-1309.

- SUMMIT: Rocca, B *et al.* (2012) The recovery of platelet cyclooxygenase activity explains interindividual variability in responsiveness to low-dose aspirin in patients with and without diabetes, Journal of Thrombosis and Haemostasis, 10: 1220-1230, doi: 10.1111/j.1538-7836.2012.04723.x.

- 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.

- U-BIOPRED: Bel, EH *et al.* (2011) Diagnosis and definition of severe refractory asthma: an international consensus statement from the Innovative Medicine Initiative (IMI), Thorax, 66: 910-917, doi: 10.1136/thx.2010.153643.

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

Section A3.2 lists the nine papers from IMI projects that have been identified as 'hot papers'. Papers are listed in ascending alphabetical order (project, first author) and those papers also identified as highly cited (Section A3.1) are highlighted (bold text).

Papers from the BTCure and EU-AIMS projects are new to this list while papers associated with eTOX, EUROPAIN and Open PHACTS were identified as 'hot papers' in the second report to IMI. NEWMEDS has accumulated a third hot paper relative to the second report.

Eight of these nine papers are classified as highly-cited. The remaining paper had not accumulated citations at end-2012 when the percentile ranking used to define highly-cited papers is calculated.

A3.2 'HOT PAPERS' ASSOCIATED WITH IMI PROJECTS

- **BTCure: Harre, U *et al.* (2012) Induction of osteoclastogenesis and bone loss by human autoantibodies against citrullinated vimentin, Journal of Clinical Investigation, 122: 1791-1802, doi: 4.52333540759179.**

- **eTOX: Arighi, CN *et al.* (2011) Overview of the BioCreative III Workshop, BMC Bioinformatics, 12: , doi: 9.18008572053678.**

- **EU-AIMS: Kong, A *et al.* (2012) Rate of de novo mutations and the importance of fathers age to disease risk, Nature, 488: 471-475, doi: 13.1726368159204.**

- **EUROPAIN: Finnerup, NB *et al.* (2010) The evidence for pharmacological treatment of neuropathic pain, Pain, 150: 573-581, doi: 18.8641598142005.**

- **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: 5.05164330712507.**

- NEWMEDS: Kapur, S *et al.* (2012) Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Molecular Psychiatry, 17: 1174-1179.

- **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: 39.2973625443629.**

- **Open PHACTS: Williams, AJ *et al.* (2012) Towards a gold standard: regarding quality in public domain chemistry databases and approaches to improving the situation, Drug Discovery Today, 17: 685-701, doi: 7.31316779533484.**

- **Pharma-Cog: Frisoni, GB *et al.* (2010) The clinical use of structural MRI in Alzheimer disease, Nature Reviews Neurology, 6: 67-77, doi: 10.1038/nrneurol.2009.215.**

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

Section A3.3 lists the five papers from IMI projects scoring highest on interdisciplinarity. These papers were selected using similar criteria as in the second report to IMI. It should be noted that as the total volume of papers from IMI projects increases, a threshold in diffusion score is less easy to define.

Papers are listed in ascending alphabetical order (project, first author). Three of these papers were also identified as highly-cited (Section A3.1) and are highlighted (bold text). None of the five papers were also identified as 'hot papers'.

A3.3 TOP FIVE PAPERS WITH HIGHEST DIFFUSION SCORE THAT ARE ASSOCIATED WITH IMI PROJECTS

- **eTOX: Obiol-Pardo, C *et al*. (2011) A Multiscale Simulation System for the Prediction of Drug-Induced Cardiotoxicity, Journal of Chemical Information and Modeling, 51: 483-492, doi: 10.1021/ci100423z.**

- PROTECT: Brauer, R *et al.* (2011) The association between antipsychotic agents and the risk of myocardial infarction: a systematic review, British Journal of Clinical Pharmacology, 72: 871-878, doi: 10.1111/j.1365-2125.2011.04043.x

- **SUMMIT: Boekholdt, SM *et al*. (2012) Association of LDL Cholesterol, Non-HDL Cholesterol, and Apolipoprotein B Levels With Risk of Cardiovascular Events Among Patients Treated With Statins A Meta-analysis, Jama-Journal Of The American Medical Association, 307: 1302-1309.:**

- Quic-Concept : Waterton, JC *et al.* (2012) Qualification of imaging biomarkers for oncology drug development, European Journal of Cancer, 48: 409-415, doi: 10.1016/j.ejca.2011.11.037.

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

ANNEX 4: BIBLIOMETRICS AND CITATION ANALYSIS

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

Publications cite other publications. These citation links grow into networks, and their numbers are likely to be related to the significance or impact of the publication. The meaning of the publication is determined from keywords and content. Citation analysis and content analysis have therefore become a common part of bibliometric methodology. Historically, bibliometric methods were used to trace relationships amongst academic journal citations. Now, bibliometrics are important in indexing research performance.

Bibliometric data have particular characteristics of which the user should be aware, and these are considered here.

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

The origins of citation analysis as a tool that could be applied to research performance can be traced to the mid-1950s, when Eugene Garfield proposed the concept of citation indexing and introduced the Science Citation Index, the Social Sciences Citation Index and the Arts & Humanities Citation Index, produced by the Institute of Scientific Information (currently the IP & Science business of Thomson Reuters).[25]

We can count citations, but they are only 'indicators' of impact or quality – not metrics. Most impact indicators use average citation counts from groups of papers, because some individual papers may have unusual or misleading citation profiles. These outliers are diluted in larger samples.

A4.1 DATA SOURCE

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

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

Unlike other databases, the *Web of Science* and underlying databases are selective, that is: the journals abstracted are selected using rigorous editorial and quality criteria. The authoritative, multidisciplinary content covers over 12,000 of the highest impact journals worldwide, including Open Access journals, and over 150,000 conference proceedings. The abstracted journals encompass the majority of significant, frequently cited scientific reports and, more importantly, an even greater proportion of the scientific research output which is cited. This selective process ensures that the citation counts remain relatively stable in given research fields and do not fluctuate unduly from year to year, which increases the usability of such data for performance evaluation.

---

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

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

A4.2 DATABASE CATEGORIES

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

Thomson Reuters frequently 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.
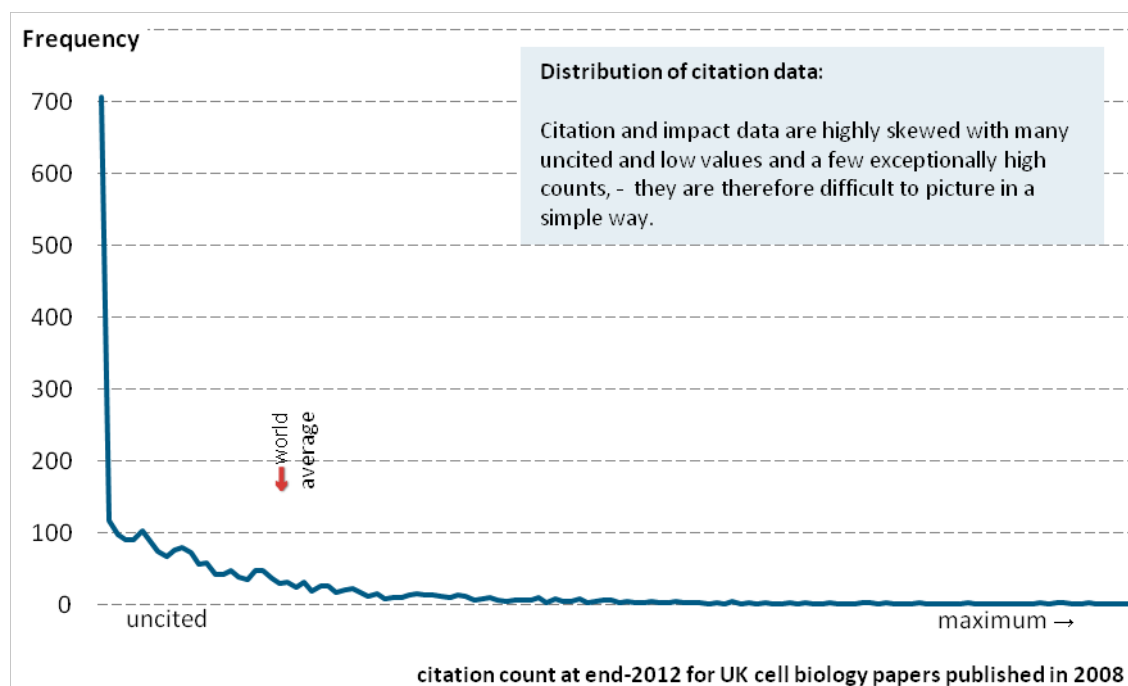
A4.4 CITATION COUNTS

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.

Frequency

Distribution of citation data:

Citation and impact data are highly skewed with many uncited and low values and a few exceptionally high counts, - they are therefore difficult to picture in a simple way.

citation count at end-2012 for UK cell biology papers published in 2008

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.
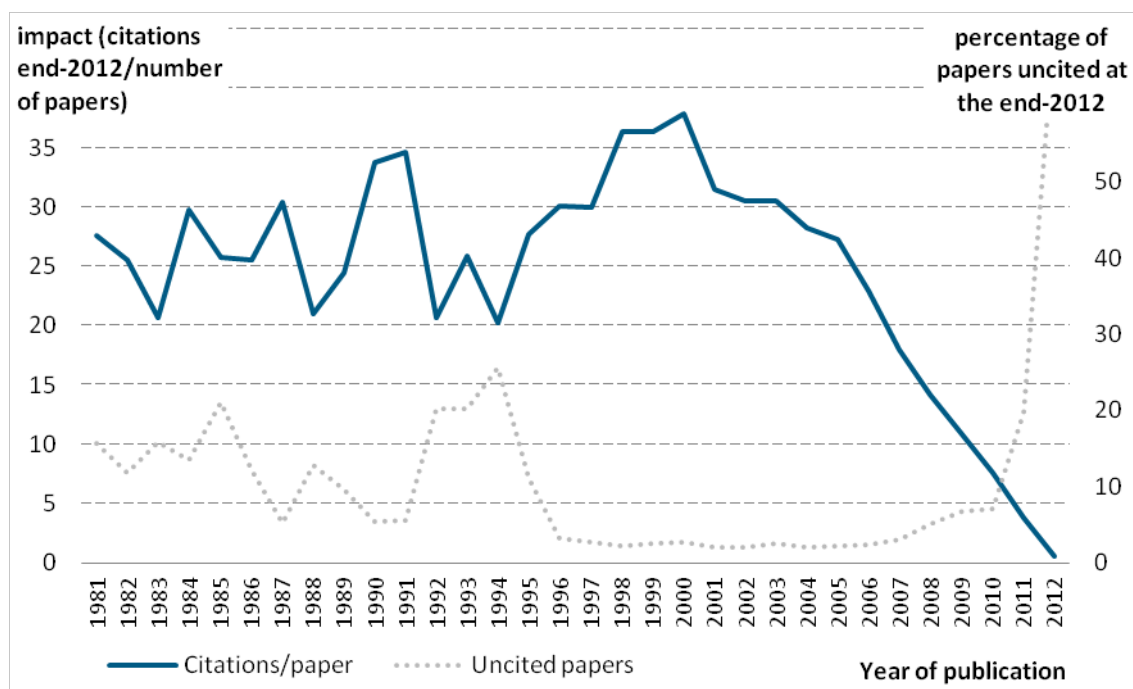
A4.6 DISCIPLINE FACTORS

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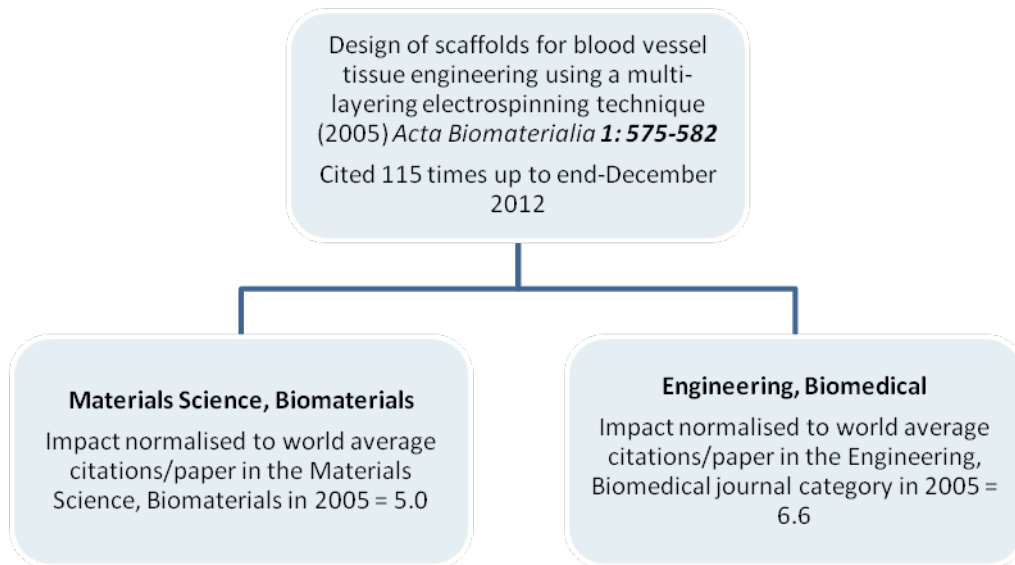
Some journals with a very diverse content, including the prestigious journals *Nature* and *Science* were classified as **Multidisciplinary** in databases created prior to 2007. The papers from these **Multidisciplinary** journals are now re-assigned to more specific research fields using an algorithm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

Design of scaffolds for blood vessel tissue engineering using a multi-layering electrospinning technique (2005) *Acta Biomaterialia* **1: 575-582**

Cited 115 times up to end-December 2012

Materials Science, Biomaterials

Impact normalised to world average citations/paper in the Materials Science, Biomaterials in 2005 = 5.0

Engineering, Biomedical

Impact normalised to world average citations/paper in the Engineering, Biomedical journal category in 2005 = 6.6

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 5.0 and the category-specific NCI_F for **Engineering, Biomedical** is higher at 6.6). 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 2012.

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 ((5.0 + 6.6)/2) = 5.8.

A4.9 IMPACT PROFILES®

We have developed a bibliometric methodology[26] that shows the proportion of papers that are uncited and the proportion that lie in each of eight categories of relative citation rates, normalised (rebased) to world average. An Impact Profile® enables an examination and analysis of the strengths and weaknesses of published outputs relative to world average and relative to a reference profile. This provides much more information about the basis and structure of research performance than conventionally reported averages in citation indices.

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

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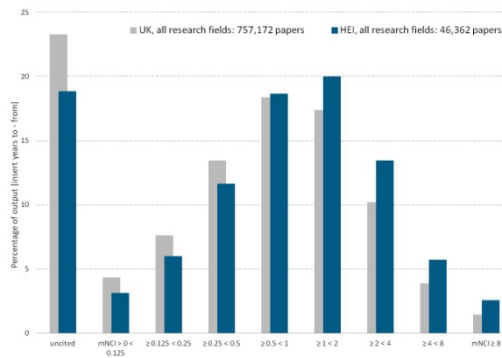
---

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

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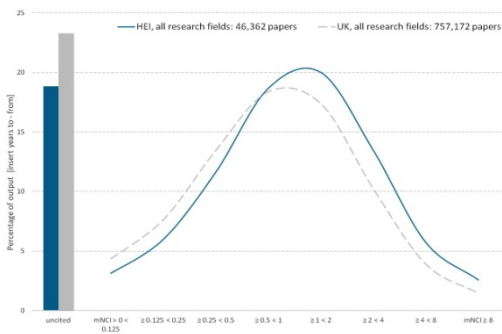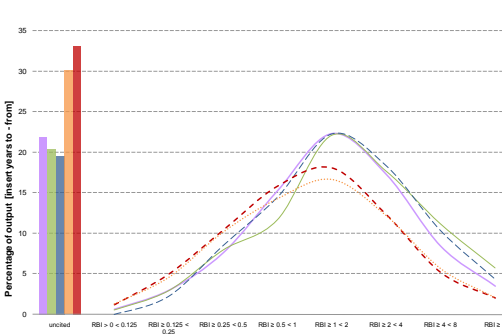
A
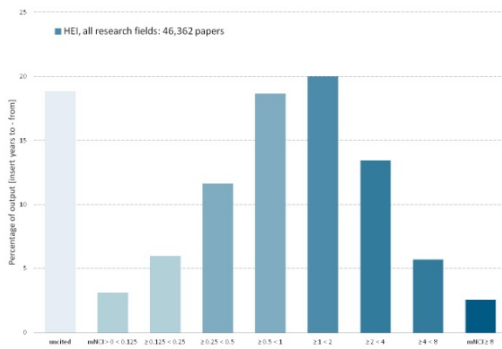
B

C

D

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

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

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

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

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

Attention should be paid to:

- The proportion of uncited papers on the left of the chart

- The proportion of cited papers either side of world average (1.0)

74

- 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 72 which has been cited 115 times to the end-December 2012, the expected citation rate for a paper in *Acta Biomaterialia* published in 2005 would be 28.7. Therefore, this paper has been cited more than expected for the journal. For a set of papers, we calculate the quality index as the percentage of papers which are cited more than expected for the relevant journals.

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

A4.11 *WEB OF SCIENCE* JOURNAL CATEGORIES

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

Language & linguistics

Language & linguistics theory

Law

Limnology

Linguistics

Literary reviews

Literary theory & criticism

Literature

Literature, African, Australian, Canadian

Literature, American

Literature, British Isles

Literature, German, Dutch, Scandinavian

Literature, romance

Literature, Slavic

Management

Marine & freshwater biology

Materials science, biomaterials

Materials science, ceramics

Materials science, characterization & testing

Materials science, coatings & films

Materials science, composites

Materials science, multidisciplinary

Materials science, paper & wood

Materials science, textiles

Math & computational biology

Mathematics

Mathematics, applied

Mathematics, interdisciplinary applications

Mechanics

Medical ethics

Medical informatics

Medical laboratory technology

Medicine, general & internal

Medicine, legal

Medicine, research & experimental

Medieval & renaissance studies

Metallurgy & metallurgical engineering

Meteorology & atmospheric sci

Microbiology

Microscopy

Mineralogy

Multidisciplinary sciences

Music

Mycology

Nanoscience & nanotechnology

Neuroimaging

Neurosciences

Nuclear science & technology

Nursing

Nutrition & dietetics

Obstetrics & gynaecology

Oceanography

Oncology

Operations research & management science

Ophthalmology

Optics

Ornithology

Orthopaedics

Otorhinolaryngology

Palaeontology

Parasitology

Pathology

Paediatrics

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 sciences, interdisciplinary

Social sciences, mathematical methods

Social work

Sociology

Soil science

Spectroscopy

Sport sciences

Statistics & probability

Substance abuse

Surgery

Telecommunications

Theatre

Thermodynamics

Toxicology

Transplantation

Transportation

Transportation science & technology

Tropical medicine

Urban studies

Urology & nephrology

Veterinary

Veterinary sciences

Virology

Water resources

Women's studies

Zoology