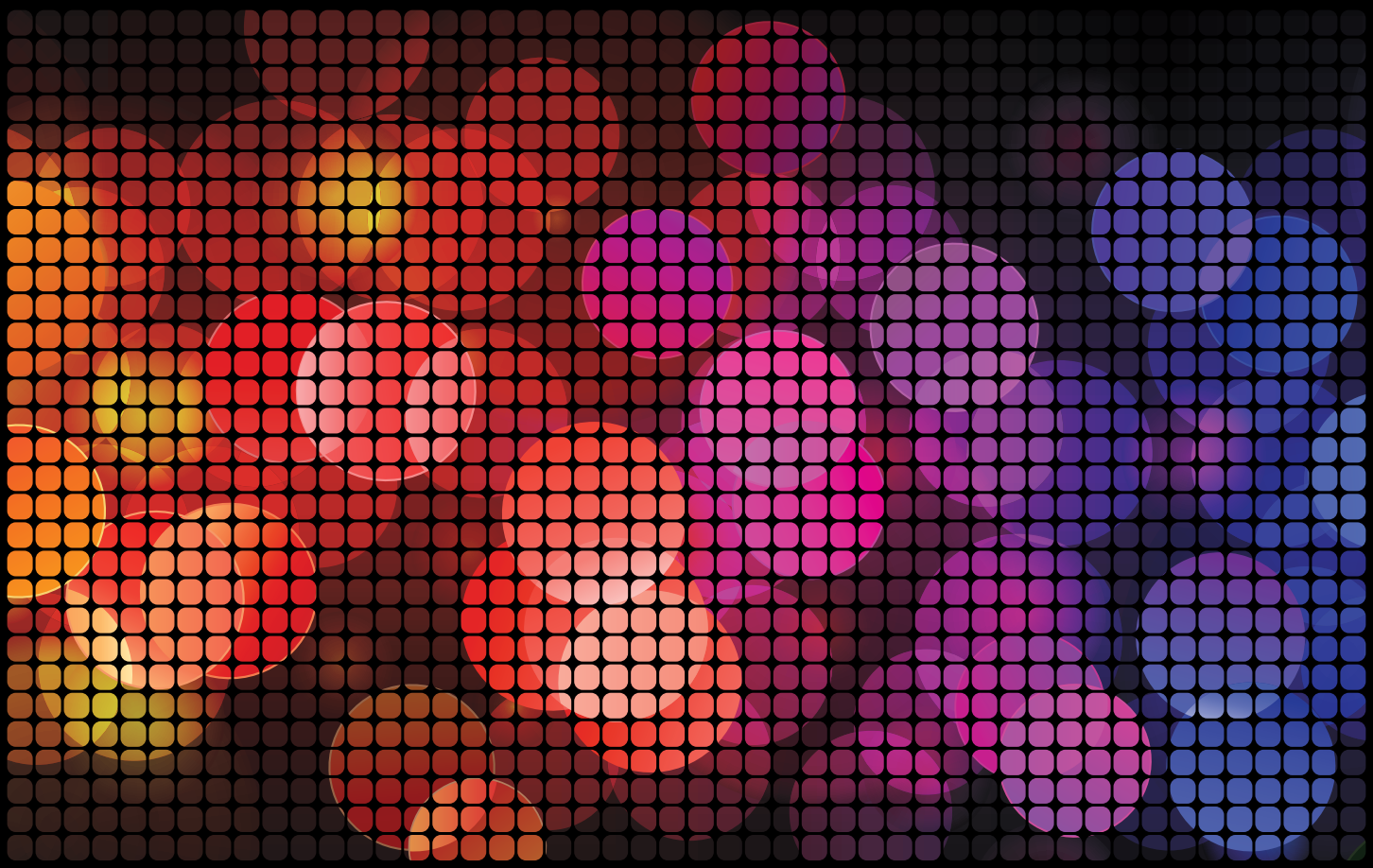

Digital Research Reports

Altmetric mentions and the communication of medical research

Disseminating research outcomes outside academia

Jonathan Adams & Tamar Loach

FEBRUARY 2015

DIGITAL science

About the Authors

Jonathan Adams joined Digital Science as Chief Scientist in October 2013. Previously he was the lead founder of Evidence Ltd (2000-2009) and Director of Research Evaluation for Thomson Reuters (2009-2013). Jonathan led the 2008 review of research evaluation in New Zealand and was a member of the Australian Research Council (ARC) indicators development group for its research excellence assessment (ERA). In 2004 he chaired the EC Evaluation Monitoring Committee for Framework Programme 6. In 2006 he chaired the Monitoring Group of the European Research Fund for Coal & Steel. In 2010 he was an Expert Advisor to the interim evaluation of the EU's 7th Framework Programme for Research (FP7).

Tamar Loach is Research Metrics Analyst at Digital Science. Her background is in Theoretical Physics, and she has undertaken research projects both in this area and on alternative research metrics with the Complexity and Networks group, all at Imperial College London. In 2014 she was an invited subject matter expert at the European Commission Institute for Prospective Technological Studies (JRC-IPTS) center for a workshop on alternative reputation mechanisms for scholars as a researcher of alternative metrics and has recently reported for NESTA on such metrics in medical research.

About Digital Science

Digital Science is a technology company serving the needs of scientific research. We offer a range of scientific technology and content solutions that help make scientific research more efficient. Whether at the bench or in a research setting, our products help to simplify workflows and change the way science is done. We believe passionately that tomorrow's research will be different - and better - than today's.

Visit www.digital-science.com

This report has been published by Digital Science, a business division of Macmillan Science and Education.

The Macmillan Building, 4 Crinan Street, London, N1 9XW
info@digital-science.com

ISBN: 978-0-9929477-1-2

Project Funders

Nesta...

Altmetric mentions and the communication of medical research

Historically, academic citations to journal articles and patents have been key to tracking important innovations in science and technology. Now, alternative metrics for research impact are available. Can these indicators spotlight medical research papers with professional and social impact?

High-impact research and technology emerges in many different social, economic and physical contexts and globalization means that it often extends beyond national borders. Policy makers and business leaders need a reliable horizon scanning capability. Stakeholders in networks beyond academia want to know about significant innovation affecting their interests. But conventional research metrics - based on journal publications and citations - only point to 'academic impact' and long after research is published.

Social and mainstream media mentions of research publications appear much more rapidly than conventional academic citations and are generated by a wider range of users. They therefore offer the potential for early and complementary indicators of research impact. Such indicators could also identify new kinds of economic and social impact.

In this report we explore the relevance of such new indicators to research in medical and health sciences.

Are altmetrics a complement to bibliometrics?

Social communication provides data from which informative signals can be extracted. For example, Boston-based Recorded Future has used the analysis of Twitter feeds to predict the timing and location of protest events and cyber-attacks around major Government-level meetings.[1]

The term 'altmetrics' (as 'alternative metrics' to traditional citation metrics for research papers) has been promoted by Jason Priem, the co-author in 2010 of the altmetrics manifesto.[2] He argued that altmetrics can expand our view of research impact by analyzing 'impact through sharing' via blogging, microblogging, and comments. In the USA, NISO (the National Information Standards Organisation) is leading an Alternative Assessment Metrics Project to collate international ideas about key requirements and potential action on altmetric applications.[3]

A number of studies on social and mainstream media that 'mention' research use data from altmetric.com, a key commercial source of systematically indexed and collated research mentions.

WHAT IS A MENTION?

Mentions are the altmetric equivalent of the citation in conventional bibliometrics.

A mention is the identification of a publication (such as a journal article) with or without additional commentary in a non-academic medium.

Mentions can be collated from sources among both mainstream (e.g. newspapers, broadcast) and social (e.g. Twitter, Facebook, blog) posts.

A score or index can be created from collated mentions by applying a weighted formula.

Here, 'Altmetric' refers to the company and 'altmetrics' refers to the concept.

Altmetric data on media mentions

Source	Mention count
Twitter	4,405,086
Facebook	253,992
News media	240,684
Scientific blogs	132,875
Googleplus	86,964
Reddit	15,053
Faculty of 1000	11,304
Sina Weibo	9,847
Video	8,126
Peer review	8,063
Policy documents	3,067

Table 1. Mentions to research documents collected and indexed by Altmetric, Aug 2013 - July 2014

Up to August 2014, Altmetric had acquired and collated mentions pointing towards at least 2.5 million scholarly documents. The majority of the mentions are tweets. In the last 12 months, Altmetric have collated over 4 million tweets, pointing towards almost 1 million articles, of which the most frequently mentioned received tweets from over 9,000 unique accounts.

Many academic papers receive no mentions but a comparison of mentioned documents to total publication counts shows that Altmetric attention is growing. The share of papers that are mentioned rose from fewer than one in twenty in 2009 to almost one quarter of output in 2013.

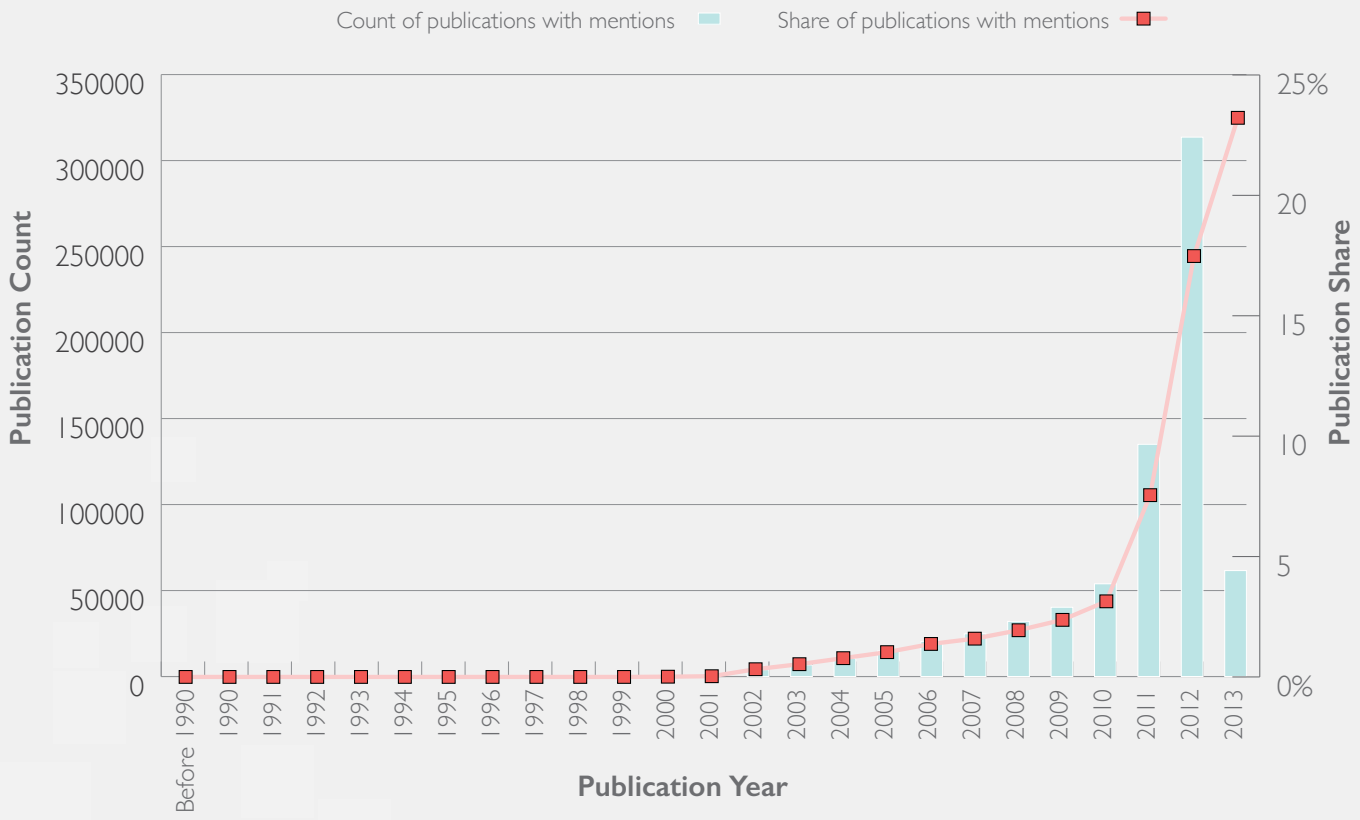

Figure 1. Research publications indexed in Thomson Reuters Web of ScienceTM that are mentioned at least once in Altmetric data. The database expanded over the period but the share of mentioned documents has grown faster (data and analysis: CWTS, University of Leiden).

In the case of Twitter, demographic data are available on individual users. This provides a way of mapping the source of mentions. An analysis by country of Twitter users' locations confirms that while users tweeting information about research publications are widely distributed, the USA and the UK stand out as key contributors.

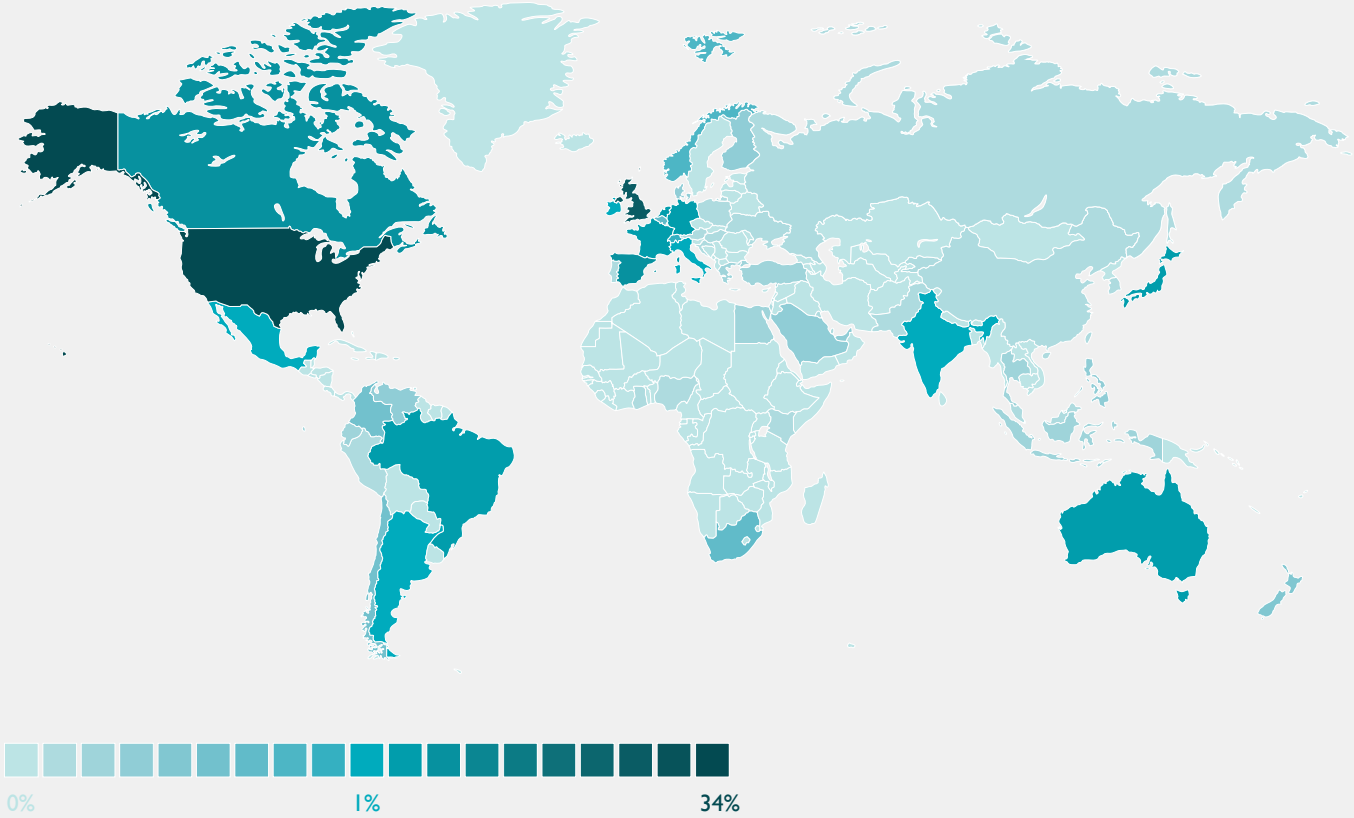

Figure 2. Twitter mentions of research publications, August 2013 through July 2014. Darker colour indicates more users (log scale of colour intensity); actual numbers given for named countries. Fig 2.a Global distribution: the USA, Australia and Brazil are very active.

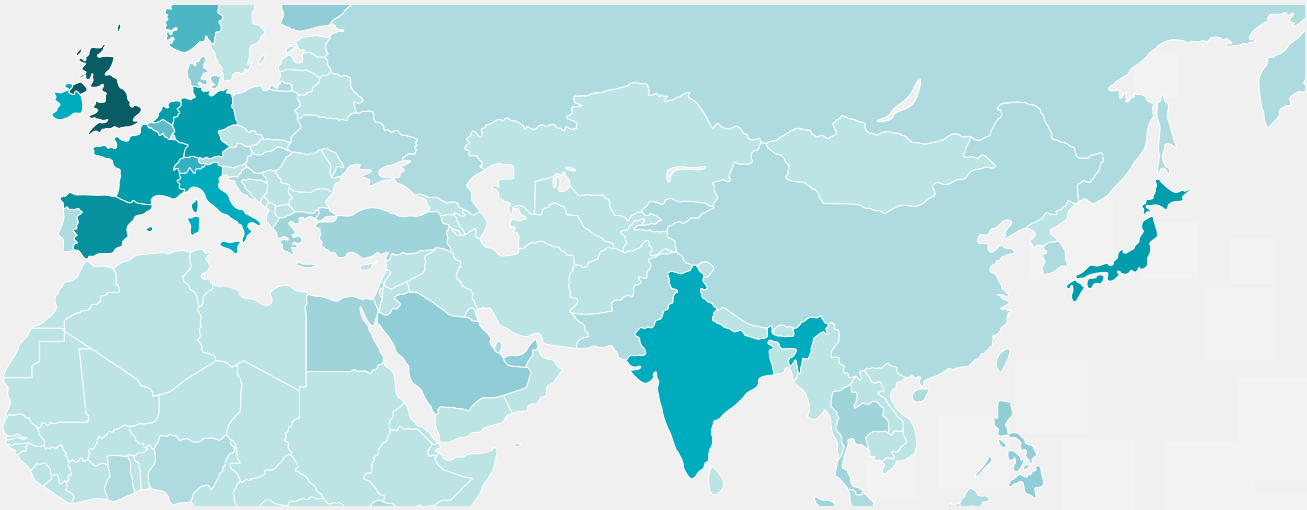

Figure 2.b. European-African distribution: Western Europe is a concentration of activity but, outside Egypt, there is a low level of media referencing of research within Africa.

The distribution of mentions across research papers is highly skewed so the 'average' is much greater than the median count. Just under half of mentioned documents have just one mention and 80% have received less than five mentions. Why are some papers mentioned frequently and does that pattern correlate with citation counts? There are both contradictory and confirmatory reports.

- The average number of mentions to papers in a specific journal correlates with the journal's Thomson Reuters 'impact factor'. Not all papers receive mentions but, where papers are mentioned, then those from journals with a higher impact factor had a higher average number of mentions. This suggests a coincidence between citation and mention behaviour.

- If the specific mention count for a research paper is compared with Thomson Reuters normalised citation index for that paper, then those with a higher number of mentions did not necessarily have more citations. This suggests that citations and mentions reference different kinds of significance.

These results present an apparent conundrum. Frequently-mentioned papers come from higher impact journals where there is strong competition among contributors. Yet these papers are not necessarily the most frequently cited in those journals. If a paper in a well-recognised journal receives many mentions, yet relatively few citations, the merit recognised by editors - and later by readers - could well come from content that goes beyond academic impact..

Mentions are clustered around medical research

Research papers that attract mentions are clustered in biomedical and health research journals (Figure 3). Research communities vary in size: there are more researchers publishing in biomedical sciences than in chemistry. Biomedicine produces shorter, more frequent publications. Engineers publish more in conference proceedings. So, clinical and biomedical research dominates publication databases. Even so, media point more towards biomedical and clinical sciences than to other sciences. They point towards social sciences that intersect with health research. And, within biomedicine they point more to clinical research than genetics and molecular biology.

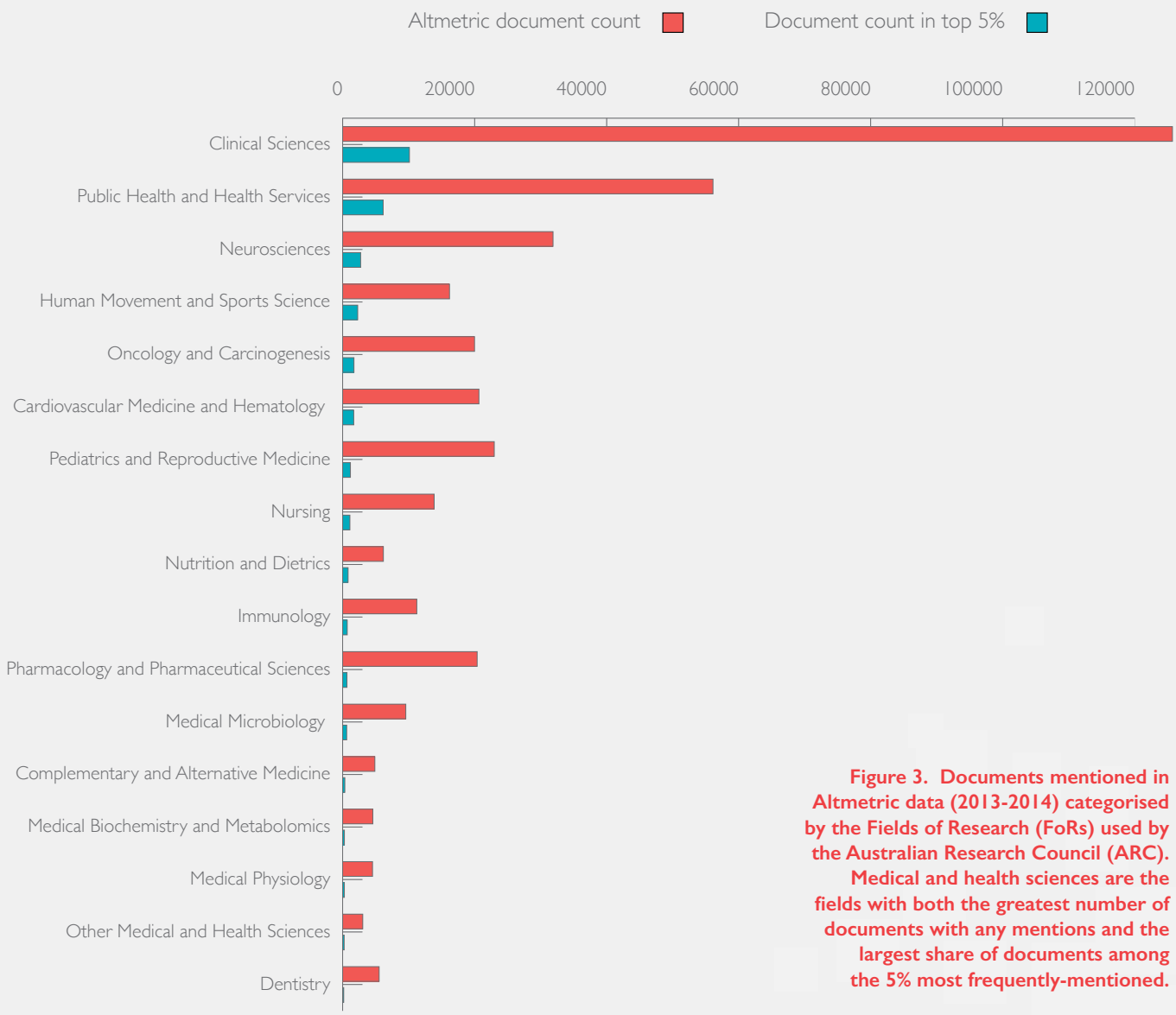

Figure 3. Documents mentioned in Altmetric data (2013-2014) categorised by the Fields of Research (FoRs) used by the Australian Research Council (ARC). Medical and health sciences are the fields with both the greatest number of documents with any mentions and the largest share of documents among the 5% most frequently-mentioned.

Are mentions motivated by new knowledge communities?

Why do mentions, but not always citations, focus on just some papers and why might they be clustered around medical research?

There are multiple motivations for 'mentions'.[6] They are a different referencing structure to the academic citation that links journal articles: no gatekeepers replace editors; no limit exists on the numbers of mentions that can be made; there are no filters and few constraints. Nonetheless, social structures mean that wholly uninformative or perverse referencing is likely to wither.

Two hypotheses about new routes to broadcasting research outcomes emerge from preliminary analyses around medical and health papers. Both point to new 'communities' of knowledge dissemination.

1. **Practical and professional impact**. Rapid and accessible communication of innovative research outcomes relevant to a '*community of practice*' such as research users and health managers, practitioners who do not always scan current literature.

2. **Stakeholder impact**. Patients, carers and charity supporters of disease groups represent a '*community of interest*' that is more likely to be looking at, able to understand and then to wish to communicate research about new treatments for critical conditions.

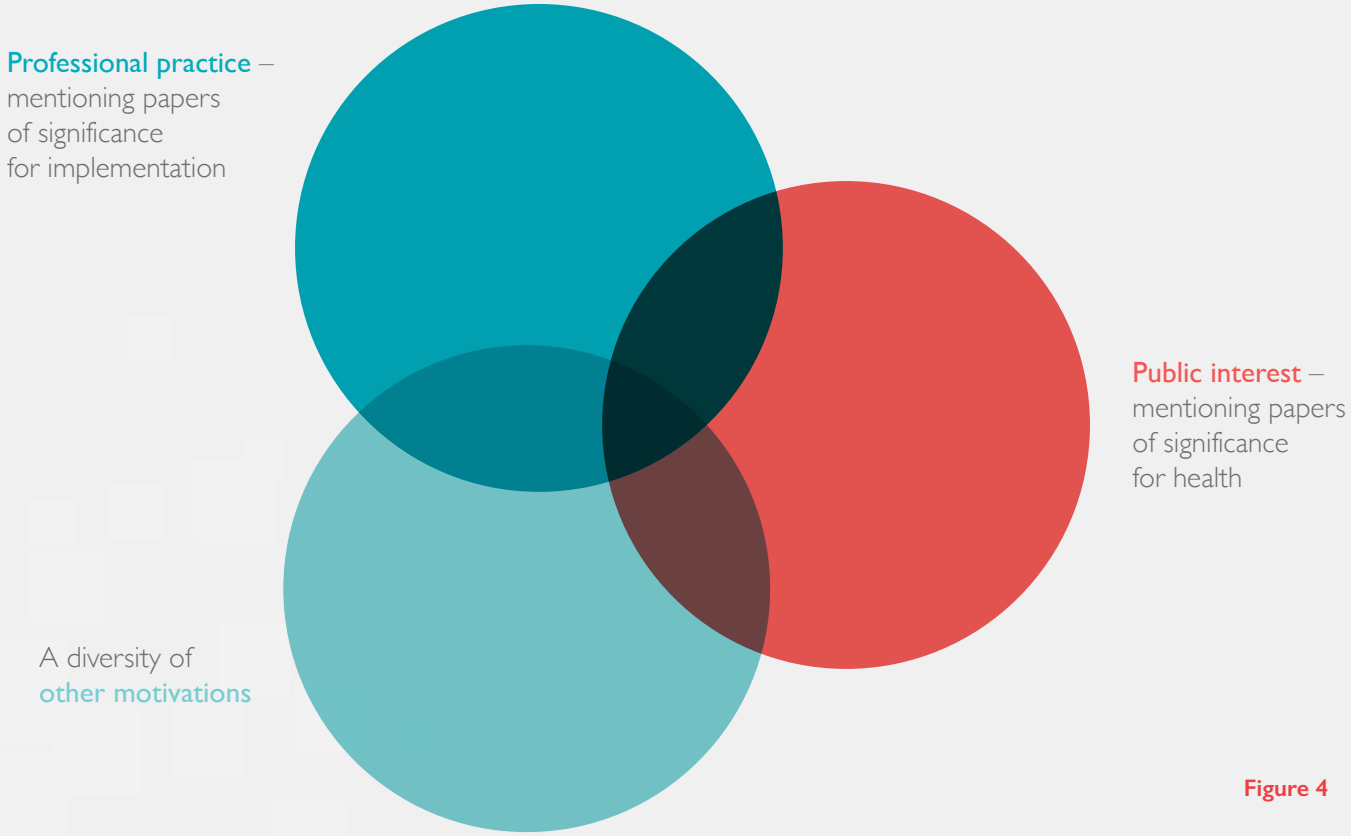

Figure 4

Linking mentioned papers to disease categories

Papers can be linked to diseases via library cataloguing. Altmetric had indexed 2,123,729 research documents at April 2014. More than half of these (1,234,681 documents) could be linked to bibliographic references in the US National Library of Medicine's Medline database.[7] The database records have subject tags (Medical Subject Headings, or MeSH terms) that describe content. There are 798,099 distinct bibliographic references in the Altmetric-linked data that have associated MeSH terms and from these 10,146,428 MeSH terms (an average of 12.7 MeSH terms per document) can be extracted.

For the set of 1,234,681 documents indexed by Altmetric and matched to Medline, it is apparent that a small proportion of documents receive a large proportion of mentions (Figure 5)

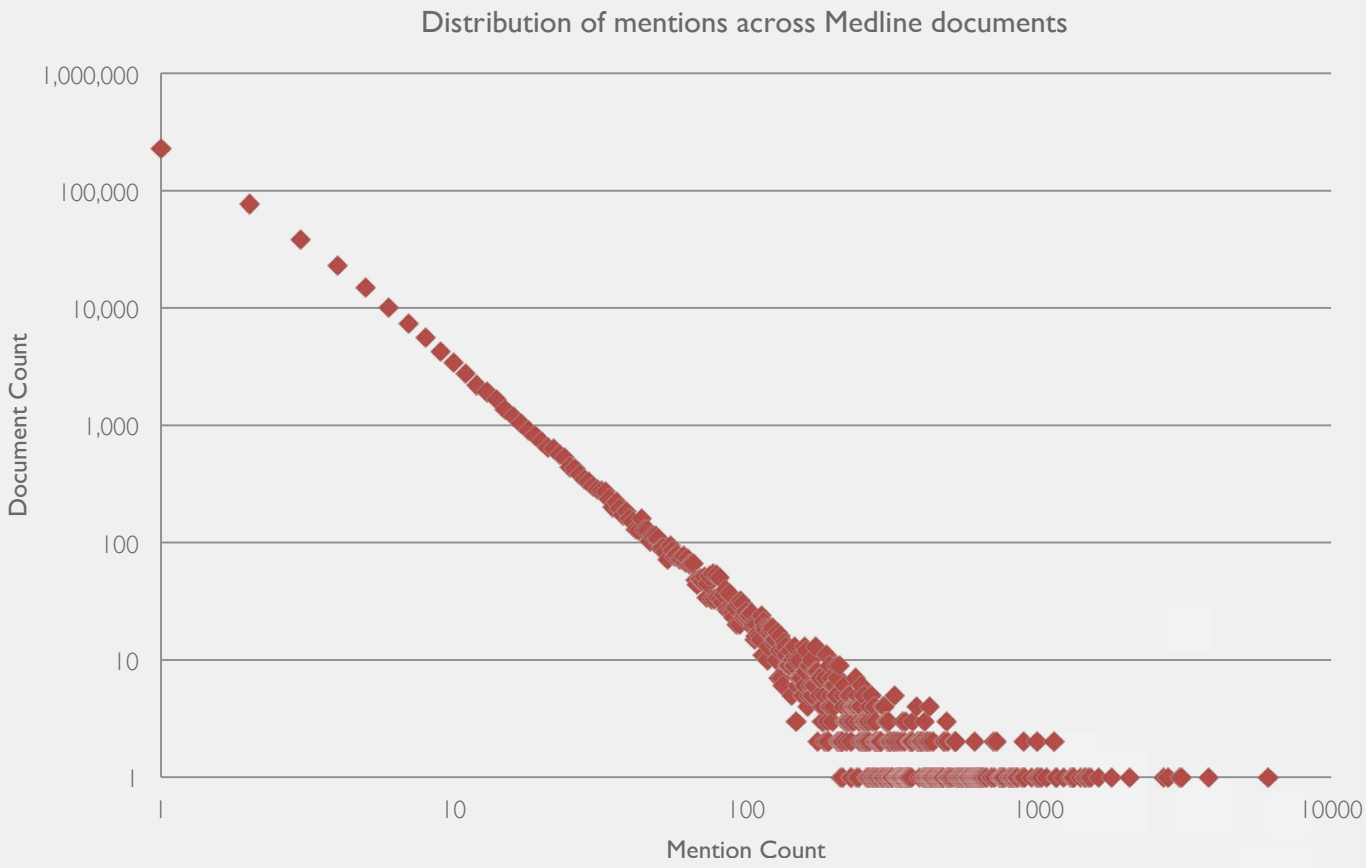

Figure 5. The numbers of Medline documents with one or more Altmetric mentions. We cannot say how many research documents have not been mentioned.

Mention counts are skewed and this affects any choice of analysis. For over one million Medline-indexed documents with at least one mention, the average count is 5.07. Statistically, the threshold for lower quartile of the distribution is 1, the median is 2, and the upper quartile is 3. This means that the vast majority of mentioned documents have a mention count below average. The maximum count of Altmetric mentions for a single research document (in this dataset at July 2014) is a massive 13,902.

Hypothesis 1 – 'mentions' indicate impact for research users

Mentions may be complementing academic citation by indicating papers that contain information of particular significance to a community of practice or for other innovative applications (as well as to further research, which will lead to their later citation).

The basis for this hypothesis is that frequently-mentioned papers:
- tend to have been published in journals of high citation impact;
- are not necessarily frequently cited, though they may be;
- are more likely, in biomedical research, to be in clinical and heath research categories than in basic biomedical research categories.

Why is this useful?

The problem of translation from research into practice is significant. Practitioners do not always have the opportunity to scan current literature. A mechanism that enables rapid and accessible communication of innovative research outcomes relevant to their practice would therefore be valuable.

If social media provide that mechanism, and if the frequency of mentions is an early guide to research outcomes that are likely to have significant value to innovation in practice or in processes and products, then this also provides an informative external reference for research and development managers and policy-makers.

How can we test the hypothesis?

To test the hypothesis, papers need to be categorised in terms of the type of impact they might have. This categorisation needs to be independent of the information about the frequency of mentions.

A journal publication bears no specific statement from author or editor that refers to its likely relevance or impact, though practice is changing in this regard. We might assume, however, that experience in a field allows the expert to develop a mental picture of research that carries significance.

We interviewed individuals with expert backgrounds as medical, biomedical and health researchers, research managers, research funders and research evaluators. We also compared frequently-mentioned articles with the immediately preceding and following articles in the same journal issue to see if any observable characteristics stood out.

Conclusions

Expert opinion leads to no clear identification of the characteristics of a research publication that was more likely to carry economic, social or professional impact. On balance, the view was that such papers might appear in different journals, possibly with different target readership. Within journals there was more likely to be similarity as regards impact, so the differences seen in the CWTS and Imperial studies are not readily explicable in these terms.

Side-by-side comparison within journal context was no more illuminating. However, a document-type analysis confirmed the diversity of motivations for mentioning a research-related document. Lectures and lecture notes were scarce in the sample but if mentioned at all then they were mentioned a lot. Such documents are not usually seen as original research publications, but in fact they can be an important research output in some technical disciplines.

Our informants suggested a second hypothesis to explain the observed patterns of social media mentions and academic citations for research papers. To quote one Dean from a university Medical School: "*Not all stakeholders in research have a direct influence but they may have an acute interest. Social media enable them to signal that interest where they see research publications of significance*".

Hypothesis 2 – 'mentions' provide a referencing medium outside academia

Social media can complement the academic route that draws attention to a research paper, by enabling non-research stakeholders – such as patients and carers – to share information with others in their network of interest

Why is this useful?

Social media mentions offer a non-academic parallel to citations. Citations acknowledge intellectual debts and point to work of influence, but they do so primarily for those already involved in research. No parallel system has been available to stakeholders interested in research outcomes but not also publishing. Social media now enables rapid identification and dissemination of recognition for innovative research outcomes perceived as relevant to a particular disease or condition.

How can we test the hypothesis?

If mentions are driven by stakeholder networks then we should expect a correlation between the average number of mentions that a paper receives and the size of the associated social network. To test the hypothesis, the size of the stakeholder community associated with a disease needs to be estimated and compared with the mention counts for papers about that disease.

We cannot directly measure the size of these groups so we need a proxy, such as information linked to the medical research charities. We might expect the number of people supporting particular charities, and therefore actively interested in the disease, to correlate with media activity. In fact, even calculating these numbers is difficult, so a further proxy is required in the form of the total research support going to each disease area.

Categorising disease data

Publication and research funding data are brought together via medical research categories.

The FoR pattern (Figure 3) is not granular enough, so we also assigned biomedical publications to topics via MeSH terms. There are over 27,000 subject headings in the 2014 Medline version, which associates individual papers with narrowly defined topics, making the system granular and flexible for analytical purposes. Where multiple disease terms exist for one document, that document - and its mentions – are counted fully towards each disease, with no fractional counting.[8]

Figure 6 summarises the data in terms of the total numbers of papers associated with each disease area. Figure 7 filters the same papers and shows the aggregate mention count for the most frequently-mentioned 10% and then just the top 1% of overall papers. There is a strong correlation across Figures 6 and 7: disease categories with many mentioned papers tend to have a high aggregate mention count among the overall frequently-mentioned papers. However, this highlights categories with a disproportionate rank on one indicator compared to the other.

Oncology (as Neoplasms), neurosciences (nervous system diseases), cardiovascular diseases and nutrition are the most frequent categories by either mode of counting. By contrast, three other categories shift rank markedly compared to the rest (five places in each case).

The disease categories with a higher position on the frequently-mentioned indicator are Virus diseases (up from 11th by paper count to 6th by mention count) and Wounds and injuries (up from 16th to 11th). The disease category which has a lower aggregate mention count considering its paper count is Digestive systems (down from 7th to 12th). So, when Virus and Wound papers are mentioned frequently then, compared to other papers, they often get an exceptionally high mention count. Digestive papers, on the other hand, get lower peak mention counts.

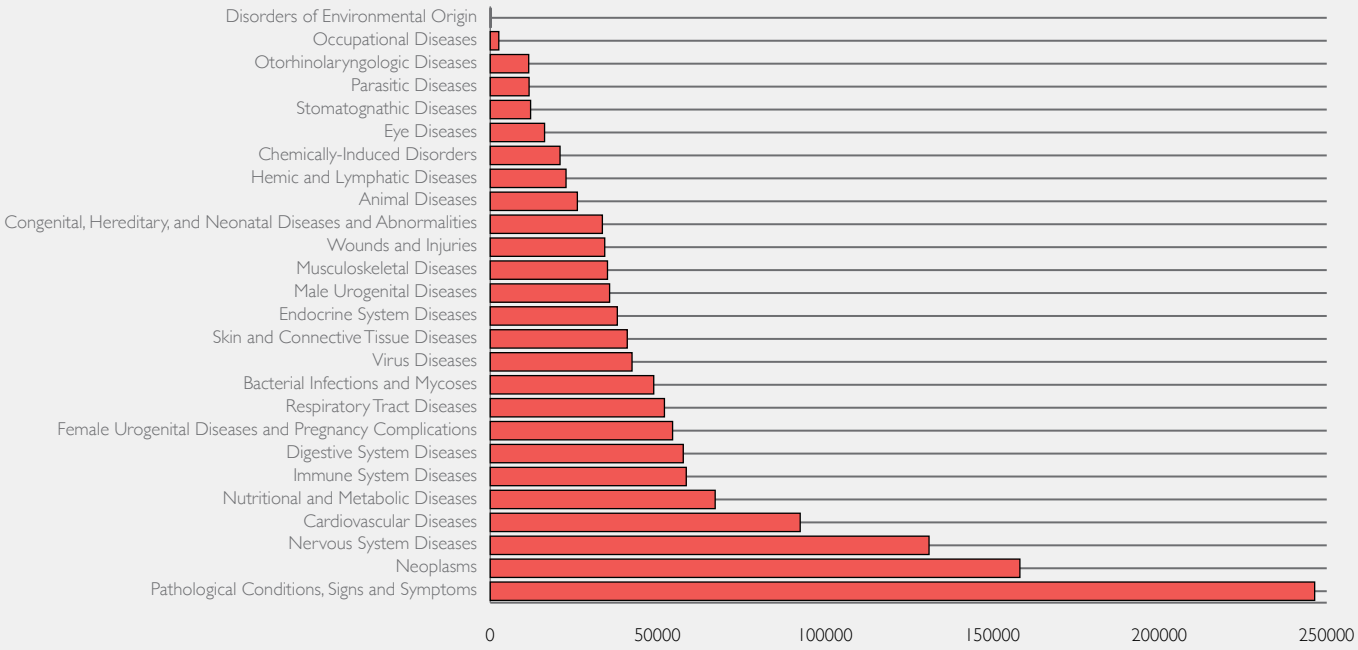

Figure 6. Count of Altmetric mentioned documents by MeSH disease term.

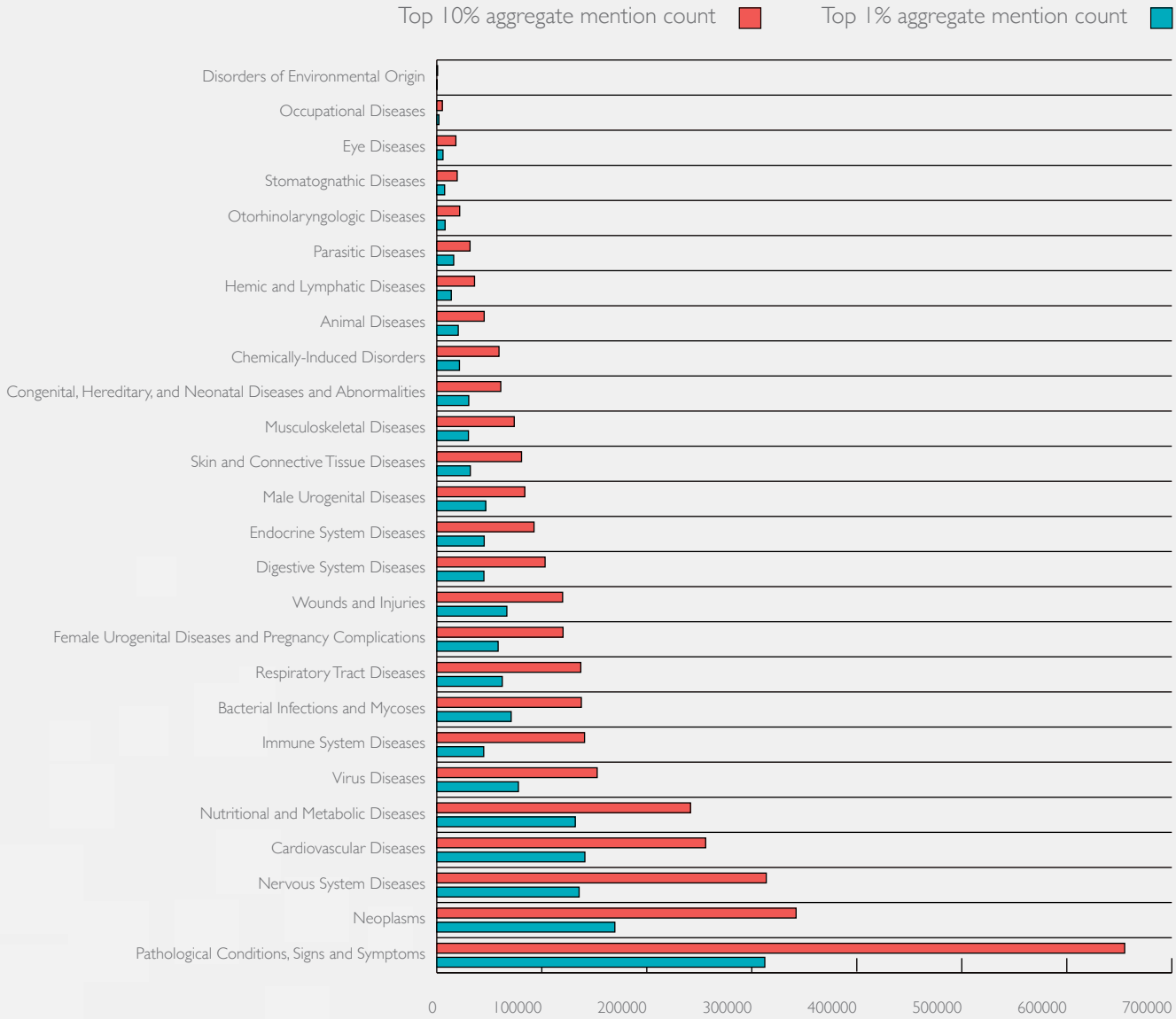

Figure 7. Aggregate mentions for most frequently-mentioned 10% (red) and 1% (blue) Altmetric indexed documents by MeSH disease term.

Categorising funding data

Funding for biomedical research in the UK comes from many sources, including the Medical Research Council (MRC), but the MRC supports much fundamental research addressing many possible disease targets. More specific research funding comes through the Association of Medical Research Charities (AMRC).

The data categorisation used by AMRC for its classification of medical research funding does not follow the Medline MeSH terms. Instead, it uses the Health Research Classification System (HRCS) developed by the UK Cancer Research Campaign partners in 2004.[9] However, we can map between MeSH terms and HRCS categories.

Collated AMRC data from 29 medium and smaller sized member charities confirm that cancer receives about 25% of UK charitable medical research funding and that inflammatory & immune, musculoskeletal, and neurological research each receive around 15%. The other 30% of funding is spread across a wide range of disease areas.[10]

We tested two expectations. One is that papers are more likely to be mentioned (i.e. attract media attention) where there is substantial charitable research support (a proxy for recent public/donor attention). The second expectation is that papers that are mentioned will receive a higher total count, on average, where there is a high pre-existing level of interest. Funding is related to the number and motivation of the fundraisers and serves as a proxy of the strength of interest.

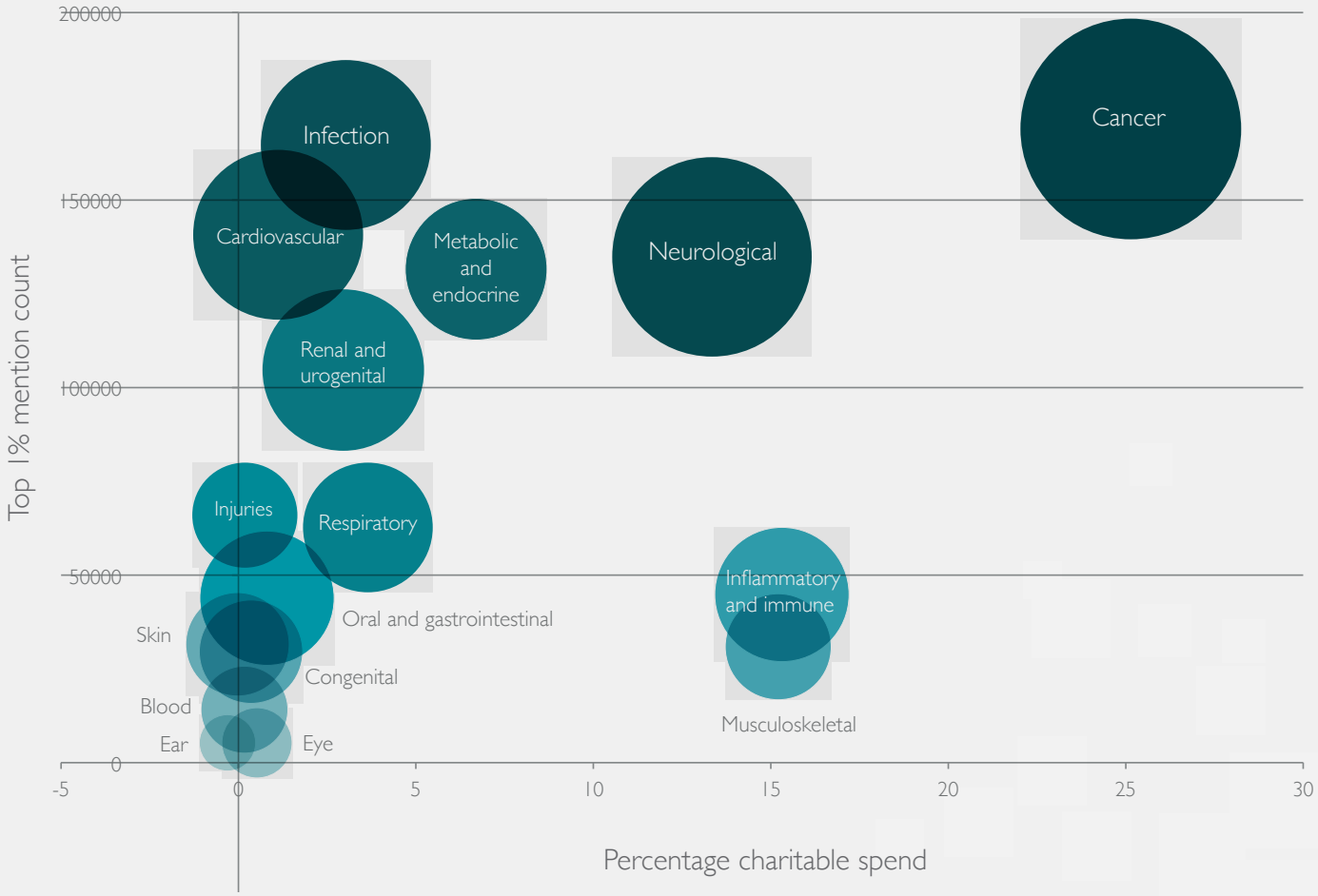

Figure 10. For HRCS disease areas, the relationship between that area's share of UK charitable research spend and the count of mentions for papers describing research on the disease that feature among the top 1% most mentioned in Altmetric data. Bubble size and colour depth reflect the number of Altmetric documents for that disease.

There are many barely-mentioned documents. To limit the analysis to papers receiving more than transient attention, we used only the 1% most frequently-mentioned disease-related papers. For each HRCS disease area, we compared share of UK charitable research spend with the count of mentions for papers referring to that disease and among that top 1%.

The position of Cancer and Neurological disease compared to Blood, Skin and Respiratory diseases shows a positive relationship between charitable research spend and the relative attention given to frequently-mentioned papers. Higher peaks of attention are achieved by research papers associated with diseases tackled by charities supported by larger research funds.

But the correlation is not quite significant ($0.06>P>0.05$) and other outcomes raise questions about the relative support and attention that different disease areas receive. Although there is a 'trend', Cardiovascular disease receives more attention than would be expected given the level of research support. By contrast, resources for Immune and for Musculo-skeletal diseases are high but the level of media attention to research papers is low compared to, for example, Neurological diseases.

Conclusions

Media mentions are a valuable new tool for professional and societal networks to draw rapid and informal attention to innovative research. They are complementary to publication and citation links, working in parallel but avoiding the hurdles and delays of conventional academic routes. However, we need wider recognition of engagement with the use of social and news media. We need to focus attention on research around — and innovation in — alternative communication of research outcomes, if such communication forms are to realise their potential benefits.

Media mentions can draw attention to innovation that is significant for both public and professional reasons. For example, a network of need and opportunity exists for stakeholder groups engaged with acute diseases such as communities of practise and interest.. Our data show that there is an emerging correlation between charitable research spend and altmetric mentions. This is already significant statistically – albeit at a low level – and it already raises challenging questions about diseases with high levels of attention but apparently lower levels of research support.

Methodologies for analysing mention data need to progress to assist non-academic groups in maximising beneficial use. For example, it would be valuable to develop profiles of individual communicators. It is likely that key facilitators communicate research results between the academic and the non-professional but informed networks. We anticipate that such key nodes in networks also exist in professional health/clinical contexts.

Rapidly growing data resources cannot yet be deconstructed sufficiently to confirm it statistically, but the relative frequency of mentions for review papers suggests that biomedical researchers and clinical and health professionals already use media mentions to draw attention to valuable and innovative outcomes, techniques and recommendations.

Both mentions and citations are used to communicate discoveries, but it may be misleading to draw too many comparisons. A supposed similarity can distract from the need to work towards individually optimal analytics and specific insights.

There is a relationship between the content of biomedical research papers and the frequency with which they are mentioned, but no analysis can yet provide any definitive outcomes. The plurality of potential stimuli for mentions is a source of cross-cutting signals without the custom filters that make the information content easy to discern. Such filters are needed.

DIGITAL science

1 https://www.recordedfuture.com/

2 http://altmetrics.org/manifesto/

3 http://www.niso.org/news/pr/view?item_key=0051dd6c2ee7962b1bd3cc35059326e1fafb2b00

4 Imperial College, London: Tim Evans and Tamar Loach (Complexity and Networks Group).

5 CWTS, Leiden: Costas, R., Zahedi, Z., & Wouters, P. (2014). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. arXiv preprint arXiv:1401.4321.

6 Haustein, S., Peters, I., Sugimoto, C., Thelwall, M. and Larivière, V. (2014). Tweeting biomedicine: an analysis of tweets and citations in the biomedical literature. Journal of the Association for Information Science and Technology, 65, 656-669.

7 http://www.nlm.nih.gov/bsd/pmresources.html

8 The most common term (Pathological Conditions etc.) is not analytically informative as this is a generic term that is applied across many disease areas.

9 http://www.hrcsonline.net/

10 Figure 4 in: UK CRC. (2007). From donation to innovation: an analysis of health research funded by medium and small sized medical research charities. AMRC/CRC, London.

11 A similar idea has been proposed by Stefanie Haustein in the context of astronomy, where there is again a significant non-professional network of interested individuals Haustein, S. and Larivière, V. (2014). Astrophysicists on Twitter and other social media metrics research. Harvard-Smithsonian Center for Astrophysics; February 7, 2014