SPRINGER NATURE

springernature.com

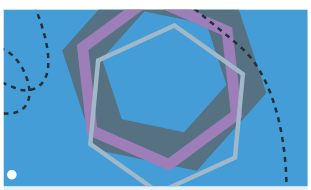
Illustration inspired by the work of Jean-Claude Bradley

Open Research

ASSESSING THE OPEN ACCESS EFFECT FOR HYBRID JOURNALS

White paper

ADVANCING DISCOVERY



Open Research

Journals Data
Books Tools

Partnering with

DIGITAL science

Contents

Authors
Hélène Draux, Mithu Lucraft, John Walker

June 2018

This white paper has been made openly available in the figshare repository.

Access case study:
https://doi.org/10.6084/m9.figshare.6396290

Foreword

At Springer Nature, we believe our role is to help researchers advance discovery. An important route to achieving this is by making their findings – research articles, books and datasets – as discoverable, accessible, understandable, usable, reusable, and shareable as possible. Open approaches benefit the whole scientific and research community, facilitating collaboration, reducing friction and inefficiency, speeding up discovery, aiding the application of research to solve real-world problems, fostering economic growth, and increasing the public's appreciation of research. As such we are committed to moving to an Open Science approach, in support of governments, research funding bodies, institutions, and researchers, wherever they are also committed to this goal and it is practical and sustainable to do so.

Steven Inchcoombe,
Chief Publishing Officer,
Springer Nature

For nearly twenty years, we have provided researchers with the ability to publish immediate 'gold' open access (OA), primarily through launching and growing new fully OA journals, and by offering OA options for our Springer and Palgrave subscription journals (i.e. hybrid OA). As a publisher, Springer Nature supports gold OA, believing that in most cases it provides the simplest, most open and most sustainable route to OA, as well as offering the greatest benefits to the research community and beyond.

We recognise though, that currently the subscription model remains the most viable route for the majority of research and, for the foreseeable future, for highly selective journals that require significant editorial investment, such as *Nature*. We therefore also offer some of the most liberal self-archiving policies to support 'green' OA, we encourage sharing of all articles via our free service, SharedIt, and we work with Scholarly Collaboration Networks, such as ResearchGate, to facilitate the sharing of articles on such platforms. Equally it is important that all articles are made available via machine readable interfaces and that their bibliographic reference lists and metadata are made openly accessible. At Springer Nature we achieve this via CrossRef and via SciGraph.

Returning to gold OA articles, last year we published over 75,000 OA articles in more than 600 fully OA journals – the most significant portfolio of OA journals in the world. We pioneered the hybrid approach, launching Springer Open Choice in 2004. Today, we offer authors OA options in more than 1,900 journals, representing 92% of our English-language subscription-based journals. Last year, we published over 15,000 OA articles via this mechanism.

Why is this breadth of OA publishing so important to the success of OA, open research and to advancing discovery more broadly? As we recently illustrated in our UK case study, much of the growth of OA has been facilitated through hybrid journals as well as through our fully OA journals. There are a number of reasons why hybrid journals, in our view, remain key:

1. Funding: A recent report from Research Consulting found gold OA uptake is largely driven by, and reliant upon, the availability of funding. With an incredibly mixed picture internationally for OA funding, hybrid journals – with their stable income via the subscription model – have enabled us as a publisher to support the take-up and growth of OA in this complex market in a sustainable way. As a global publisher we need to ensure that we serve the whole community, including researchers in disciplines which are unable to attract public funding, and those who are in countries where research funds are limited.

2. **Author choice**: Hybrid journals continue to present an attractive choice to authors. We know that authors are motivated first and foremost by their desire to be published in a relevant peer-reviewed journal with a strong reputation in their community. OA is rarely their first concern. Our regular author surveys (last year completed by over 70,000 authors from all disciplines and regions) have shown for many years that researchers' top four criteria when choosing where to submit their draft manuscript are a journal's reputation, its relevance, the quality of its peer-review, and its Impact Factor. OA has risen from #10 five years ago to #8 today. Many fully OA journals, especially those published by Springer Nature, now have high levels of citations and usage – and good reputations, readership and author communities – they are challenging well-established journals. However, it remains the case that the vast majority of journals today that can offer immediate OA publication are hybrid.
3. **Cost of transition**: Without this mixed model approach, the cost of facilitating OA options would be significantly greater: in order to support the global research community, we would need to create new OA journals to mirror our 1,900 subscription hybrid journals; we could not simply adapt all of these existing journals. The additional cost/time/risk/disruption for the whole research ecosystem as well as to publishers would be huge compared with the opportunity to progress an orderly evolution. And yet, in 2018, a number of research funders are considering excluding hybrid journals or capping their APCs, which we believe would risk a significant regression in OA uptake, based on the continued demand from authors for these established journals.

Recognising the importance of the hybrid option, we commissioned Digital Science to undertake the analysis summarised in this white paper, to show whether there is real benefit for authors, their institutions, and funders in choosing the gold OA publishing option in hybrid journals. This topic spawns much debate, particularly around the economic value of the hybrid model. In our view, in the complex international research ecosystem, hybrid journals are critical for facilitating the on-going growth of OA in a sustainable way, where underpinning support remains via subscriptions in most cases.

The results of this analysis clearly show that hybrid OA offers significant benefits for researchers, increasing usage, citations, and attention. On average, OA articles are downloaded four times as often as non-OA articles. Some of this is undoubtedly usage by interested people that do not have the benefit of an affiliation with a subscribing institution, but some is likely by researchers from subscribing institutions that are travelling or just not on their campus. Turning to citations, on average OA articles are cited 1.6 times more frequently than similar subscription articles. And looking at Altmetric, on average, OA articles attracted 2.4 times more attention than non-OA articles.

As noted in the discussion at the end of this white paper, we cannot control for all variables in this type of analysis, and in particular there is a selection bias risk – for example, that authors choose OA for their most significant work. Nonetheless, several of the specific findings from this study indicate that OA does of itself confer benefits, including the comparison of article cohorts from a single country (the UK). The most direct comparison is between recognised users where usage of OA articles is approximately 1.5 times higher, indicating likely enhanced discovery, resulting in part from greater sharing.

We hope that the findings in this white paper demonstrate the value hybrid journals are bringing, to complement fully OA journals, directly to researchers, and by extension to funders, institutions, and to society more broadly.

Foreword

To create an open research ecosystem, open access is just one part of a greater whole. To have open research we must also have open peer review, open data, open systems around provenance and reproducibility, and open frameworks for indicators and metrics. Open access (OA) is open communication and open dissemination of research results, which is a critical first piece in a longer journey.

Daniel Hook,
Digital Science

At Digital Science, we have sought to complement OA with technologies that support the core of OA but also the wider ecosystem of open research around it. The connections between objects that are needed to meet the needs of initiatives such as OpenAIRE are codified in the data held in articles, which are often not structured or formulated in a standardised infrastructure. At least three Digital Science products (Dimensions, Symplectic Elements, and figshare) attempt to solve that problem of linkage from different perspectives.

Research itself is changing fundamentally. The relationship that research has with data is driving a revolution across fields – almost all researchers now need data skills. Digital Humanities is an emergent and exciting field that uses these new skillsets in a context that could not have been imagined 30 years ago. Collaboration is increasingly global as the internet facilitates communication beyond political and geographical boundaries. All this is driven by a technology that enables us to dream big about the possibilities of opening up research to increase the pace of discovery.

With this in mind, it is a pleasure to work with our sister company, Springer Nature, to produce this analysis on a topic so close to both our hearts. Gaining insights into how to move OA forward in a sustainable way to power the open research ecosystem of the future is core to both our missions. We also want it to help others in academia and beyond, to understand the infrastructural challenges that we need to overcome in the next few years.

On a final note, the outlook seems positive. The case for OA is established. The results in this study make it clear that OA certainly benefits the scholarly community and we make a strong case that the signal from Altmetric for the route to impact of open research demonstrates that openness makes a huge difference. The direction of travel should be clear.

Executive summary

This white paper explores the impact advantage of open access (OA), looking specifically at Springer Nature hybrid journals. Previous studies have defined 'impact' in various ways; here we consider usage (downloads), research impact (citations), and broader impact (looking at Altmetric scores, news, and policy mentions).

The results present strong evidence that OA articles in hybrid journals attract significantly more downloads, citations, and attention compared with articles published non-OA in hybrid journals.

Methodology:
We performed two multidisciplinary studies: first, taking a global sample of 73,925 journal articles published in Springer Nature hybrid journals from January to June 2014; and second, focusing on articles in Springer Nature hybrid journals with corresponding authors affiliated to UK institutions, including 3,087 OA articles published in 2016, along with a comparison set of 6,027 non-OA articles published in 2014 and 2015. Informed by earlier research, we examined the relationship between OA and usage (measured in terms of downloads), citations, and broader impact (using Altmetric data). In a model, we corrected for the influence of variables at the author level (institutional reputation, based on the proxy of a university ranking, and geographic region) and the journal level (Impact Factor, as a proxy for perceived journal prestige, and subject field).

Key findings:
Across both studies (global and UK), we found OA articles in hybrid journals benefit from an advantage across all metrics considered, attracting significantly more downloads, citations, and attention compared to non-OA articles.

Downloads:
Global:

- OA articles are downloaded significantly more often than non-OA articles, even when controlling for Impact Factor and institution ranking.
- In the global study, OA articles were downloaded on average four times more often than non-OA articles. After controlling for several variables, our model predicted 269% more downloads.
- A usage advantage was found across all subject fields.

UK:

- The UK study found a similar usage benefit, with 3.2 times more downloads for OA articles on average.

OA articles in hybrid journals benefit from an advantage across all metrics

Citations:

Global:

- In the global study, we found that OA articles attract an average of 1.6 times more citations.
- The citation advantage was found across all subjects, with the most significant gain for articles in clinical medicine, where OA articles attracted almost twice as many citations.
- The model for the global study predicted that OA articles receive 36% more cumulative citations, after controlling for the influence of other variables.

UK:

- In the UK study, after two years, OA articles had gained an average of 1.6 times more citations than non-OA articles. The model predicted that OA articles were cited 30% more than non-OA articles.
- The most recent articles published in the UK study were only 15 months old at the time of analysis, which is relatively early in terms of assessing scholarly impact. These results should therefore be considered as directional only.

Attention:

Global:

- In the global study, OA articles attracted an average of 2.4 times more attention.
- OA articles received 1.9 times more news mentions on average, with the model predicting that OA articles have 219% more news mentions.
- On average, OA articles received 1.2 times as many mentions in policy documents. The model predicted OA articles have 166% more policy mentions.

UK:

- In the UK study, the average Altmetric score after one year for OA articles was 3.2 times higher than for non-OA articles.

In this white paper, we first quantified the OA advantage in terms of averages. However, owing to the non-normalized distribution of impact data, we also used statistical models to quantify the advantage while controlling for confounding factors. Whilst we found strong evidence of an OA advantage while controlling for some factors that also likely influence downloads, citation, and mentions (including Impact Factor, author institutional affiliation, and subject), we acknowledge that there are a number of other factors that may also play a role, which are not addressed here, such as the availability of articles through other routes such as green OA or sharing services. As one of the first major analyses of hybrid usage data, this white paper sets out a strong case for an hybrid OA impact advantage. We would encourage other publishers to conduct similar analyses and to continue to build on a shared understanding of the benefits of hybrid journals and the effects of choosing open access, both to provide further insights to authors on the benefits of OA, and ultimately to support a transition to OA that benefits funders, research, and the world at large.

Introduction

In the 17 years since Steve Lawrence wrote in *Nature* that free online availability of a research paper substantially increases its impact[1], measured by citation rates, there have been a number of studies that have considered the impact advantage of open access (OA)[2]. Studies have approached the OA effect from a wide number of angles, with a significant number identifying some advantage from OA: Scholarly Publishing and Academic Resources Coalition (SPARC) Europe compiled a list of 70 studies on OA citation advantage published between 2001 and 2015[3]: 46 studies found a positive advantage for publishing OA, 17 found no advantage, and seven were inconclusive or measured other effects.

Some studies have looked within a single discipline, such as astrophysics[4], condensed matter[5], and agricultural research[6]. Others have considered a single journal[7] or looked across journals[8]. Certain studies of the advantage of OA have controlled for confounding factors which potentially affect impact, including journal ranking[9] or quality[10]; temporal changes[11]; institution[12], and country or region[13]. In the Gargouri *et al*. comparative study[14] comparing self-selective self-archiving with mandatory self-archiving articles, the authors included article-level factors (article age, number of co-authors, references or pages, article type, country, and field) and one journal-level variable, the journal IF.

The measurement of impact has, in the context of these existing studies, looked primarily at usage (measured by downloads) and citation rates, both of which give an indication of academic impact and potential direct benefits of OA to the research community. Research impact studies have more recently also begun to explore the measurement of impact on society and the public[15]. One tool that tracks proxies for societal impact is Altmetric, which has been used to measure broader impact in the field of climate change[16] and the societal impact of researchers at the University of Sheffield[17].

Hybrid OA has grown significantly over the past decade, with more than 45,000 OA articles published in hybrid journals in 2016[18]. The availability of hybrid OA has been cited as a necessary part of the transition to a fully OA system[19]. Research by the Wellcome Trust in 2016 reported that the key deciding factors that matter to researchers are journal reputation, journal audience, high-quality peer review, and journal IF[20]. This is supported by author research conducted by Springer Nature[21], finding that authors prioritise journal reputation over OA; they will submit to the best journal for their research whether it offers OA or not. The cost of publishing under a hybrid model has led to discussions around the value of publishing OA[22]. If a primary motivation for authors publishing in high impact journals is to gain more citations, is there a return on investment – in terms of increased impact – from paying an APC to publish OA in a hybrid journal?

This study considers the value of OA in hybrid journals. We took two multi-disciplinary approaches, one at the global scale and another focused on articles with authors affiliated to UK institutions. Informed by earlier studies, we examined the relationship between OA and usage (measured in terms of downloads), citations, and broader impact (using Altmetric data). We corrected with variables at the author level (institutional reputation, based on the proxy of a university ranking, and their geographic region) and the journal level (IF, as a proxy for perceived journal prestige, and subject field).

This study considers the impact of OA in hybrid journals

Methodology

To examine the impact advantage of publishing OA, we compared OA and non-OA articles in terms of usage (downloads), research impact (citations), and broader impact (Altmetric attention). While alternative metrics cannot claim to quantify the impact of an article, they indicate early attention outside academia. While not a perfect impact measure, it provides a signal of societal attention.

Our study has two parts. The first includes articles published in Springer and Palgrave Macmillan hybrid journals, across all author affiliations. We corrected for common variables known to affect the performance of academic papers: the perceived prestige of the journal (using IF as a proxy), the first author's institutional affiliation, and the journal subject field. The second part then takes a single-country focus, looking only at articles published by corresponding authors at a specific set of UK institutions, in order to control for differences across countries. As recently published in a separate Springer Nature case study[23], the UK is a global leader for OA, publishing a significantly higher proportion of articles via the OA route than the global average. The time period selected for the UK study was chosen to enable a comparison of articles published before and after the introduction of Springer Nature's Compact agreement with Jisc[24] which covers both content access and publishing fees. The UK study also included journal IF as an explanatory variable.

Global study

This covered 73,925 journal articles, published from January to June 2014. In the dataset, 3,004 articles were OA (4%) while 70,921 (96%) articles were non-OA. Some of these non-OA articles may have been freely available to non-subscribing users for a period of time, for example for marketing purposes. We only included articles written in English, and classified as research articles, conference papers, reviews, or short surveys. The monthly distribution and proportion of OA articles was consistent over the period. We examined three commonly-used metrics: downloads, citations, and attention.

- **Downloads** are tracked by Springer Nature. We used the total number of downloads between publication and data retrieval (December 2017), distinguishing between:
    - "**recognised use**", where the user's IP address is recognised as being that of a registered institution (i.e. the institution has or has had some form of subscription to a Springer Nature product);
    - "**non-recognised use**" – the remainder.
- **Citations** were extracted in March 2018 from Dimensions, the scholarly database developed by Digital Science. We used cumulative citations from the date of online publication. A comparison between Scopus and Dimensions citation totals confirmed that both datasets were comparable (Pearson correlation 0.97).
- **Attention** was sourced from Altmetric, which tracks mentions of research articles in mainstream media, policy sources, blogs, social media sources (Twitter, Facebook, Google+), online references (Wikipedia), and videos. We used the overall score (a weighted sum of all mentions), and the separate scores for mentions in mainstream media and in policy documents. These three metrics allow a broad comparison of the two datasets.

We then considered multiple independent variables that, based on earlier studies in the

literature, we expected could influence the impact of articles:

- **Journal Impact Factor (IF) 2014** (source: Clarivate Analytics). The IF was used as a proxy for perceived journal prestige. We recognise that the IF is subject to a range of criticisms, and included it as a variable in the study not as an endorsement but rather because it is a metric well-recognised by researchers. The range of IF scores varies across subjects. We also ran the models using an in-house, subject-weighted version of the IF; this gave very similar results. In this white paper we report on the results with the official IF, for simplicity.
- **Subject field** (source: Springer Nature): Biomedicine, Clinical Medicine, Human Sciences, Life Sciences, Mathematics / Computer Sciences, Physical Sciences / Engineering, Social Sciences and Humanities, and Other (articles outside the other categories).
- **Research institution** affiliation of the first author (source: Dimensions). As a proxy for perceived institutional prestige, we used the 2018 Times Higher Education World University Rankings (THE[i]). THE ranks around 1,000 universities based on composite scores across five categories: Teaching, Research, Citations, International Outlook, and Industry Income. For this study, we used the score rather than the ranking since it is a continuous measure. We acknowledge that various criticisms have been made of the concept and practice of university rankings; it is used here merely as a directional proxy for perceived institutional prestige.
- **Geography**, based on the affiliations of all authors in the Global Research Institution Database (GRID). We considered five regions: Africa & Middle East, Asia Pacific, Central & South America, Europe, and North America. When the authors worked in different regions, we distinguished between bilateral collaboration (two regions) and multilateral collaboration (three regions or more).

Descriptive statistics are based on the full dataset, but in the models we used a smaller dataset, due to the inclusion of the journal IF and of the university ranking of the first author. 10% of articles were published in a journal without a 2014 IF, and 50% of the articles did not have an author affiliation to one of the 1,103 research institutions ranked by THE[i]. When keeping only the articles which had both an IF and an affiliation in the THE ranking, the OA sample is 47.6% of its original size (1,367 articles) and the non-OA sample is 45.5% (33,095 articles). The average number of cumulative citations in the smaller sample is quite similar to the full dataset: non-OA has an average of 7.5 (full data: 7.7) and the average for OA is 12.3 (full data: 12.3).

UK case study

The Jisc Compact agreement with Springer Nature enables researchers in selected UK institutions to publish OA, without payment, in the majority of Springer hybrid journals. The agreement started in October 2015; for simplicity we included only the 3,087 OA articles published in 2016. We compared them to 6,027 non-OA articles published in the two years prior to 2016. All articles had a corresponding author affiliated with an institution covered by the Jisc Compact agreement. Again, we examined three commonly-used metrics: downloads, citations, and attention.

- **Downloads** are tracked by Springer Nature. Since the articles were published in different years, and some of them relatively recently, we used the monthly downloads from the date of online publication.
- **Citations** were again extracted from Dimensions, and we used the yearly citation data to accommodate for the fact that articles were published in different years.
- **Attention** was provided by Altmetric, and we used the score one year after publication.

i. A rank could not be attributed to certain articles for two reasons: the affiliation reported in the article did not match any GRID ID, or the affiliated institution was not ranked by THE.

The regional variable was kept constant (the UK), so we only considered the variables that had shown most influence in the first dataset:

- **Journal Impact Factor 2016** (source: Clarivate Analytics). The IF had proven a useful correcting factor in the global study, so we used it again in the UK study.
- **Subject field**: Biomedicine, Clinical Medicine, Human Sciences, Life Sciences, Mathematics / Computer Sciences, Physical Sciences / Engineering, Social Sciences and Humanities, and Other.

Statistical models

Although many studies have used a linear regression to assess the advantage of OA[25], others have used more advanced models. Mueller-Langer & Watt used negative binomial, Poisson, and generalised method of moments, and instrumental variable methods regressions[26]. These models are robust when dealing with bibliometrics.

Downloads, citations, and Altmetric data all have a high prevalence of zero values (undownloaded, uncited, or zero-scored articles), and are therefore not easily modelled by linear approaches. We used the Negative Binomial Generalised Linear Models (NB GLM) as they are adapted to zero-inflated datasets[27]. To include time series as an independent variable, we used the Negative Binomial Generalised Linear Mixed Models (NB GLMM). We used the NB GLM and NB GLMM based on models run in R (package lme4). A graphical representation of the models and their description can be found in Appendix A: models.

Results

Global study: OA articles published in hybrid journals

OA in hybrid journals attracts significantly more downloads, citations, and Altmetric mentions

Summary
We considered two sets of articles published in hybrid journals: the first set contained 3,004 OA articles (4% of the total dataset); the second set contained 70,921 non-OA articles (96%), which formed our control set. All articles were published in the same six-month window: January to June 2014. The monthly distribution and ratio of OA articles was stable over the period.

We used three common metrics to measure the impact of journal articles on the research community and society more broadly. The count of downloads, tracked by Springer Nature, shows the usage of articles by recognised (whose IP is from a recognised institution) or non-recognised users. The count of citations, provided by Dimensions, shows the use and recognition by authors' peers. The scores from Altmetric, and its specific mainstream media and policy mention scores, are a signal of societal attention.

Overall, our results show that OA articles in hybrid journals attract significantly more downloads, citations, and attention/mentions than non-OA articles. Table 1 shows a summary of these findings.

	Measure	Averages	Model predictions
Downloads	All downloads	On average, OA articles were downloaded 4 times more.	The model predicts that OA has 269% more downloads.
Citations	Cumulative citations	On average, OA articles received 1.6 times more citations. The biggest gain was in Clinical Medicine, with twice as many citations.	OA articles receive 36% more cumulative citations.
Altmetric	Score	On average, OA articles attracted 2.4 times more attention.	The overall Altmetric score for OA articles is 251% higher than that of non-OA articles.
	News	On average, OA articles attracted 1.9 times more news mentions.	OA articles have 219% more news mentions.
	Policy	On average, OA articles attracted 1.2 times more policy mentions.	OA articles have 166% more policy mentions.

Table 1: Summary of findings for three metrics, global study

Usage
Springer Nature tracks the number of times articles are downloaded on its platform. The cumulative downloads for the OA and non-OA articles were significantly different (Student's t-test, $p<0.0001$). OA articles were downloaded on average four times more often. The average values are shown in Figure 1, which shows a gain for both users at recognised institutions and other users.

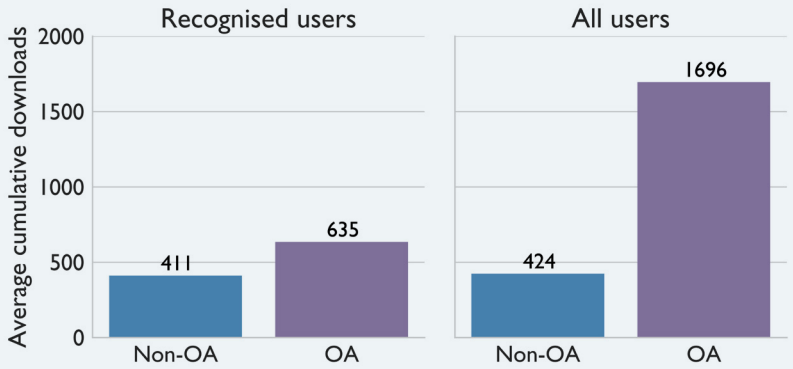


Figure 1: Average downloads of OA and non-OA articles; recognised and all users

Figure 2 shows the logarithmic value of the downloads for OA and non-OA articles[ii].

The average usage benefit of OA was significant for each subject field (see Table 2 for individual p-values, at least $p<0.01$), as shown in Figure 3:

- The biggest gains were in Social Sciences and Humanities (4.3 times more on average for all users) and Human Sciences (3.6 times).

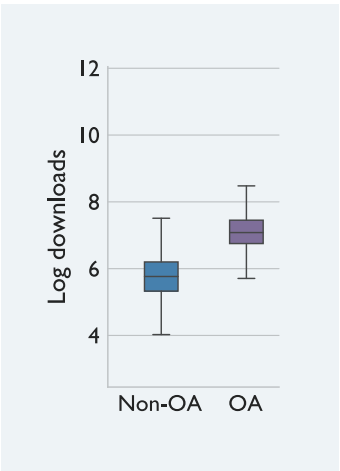


Figure 2: Distribution of (log) downloads, for all users

ii. The logarithmic value distinguishes more between low values and compensates for the effect of outliers. This does not change the relative values so allows for a clearer comparison.

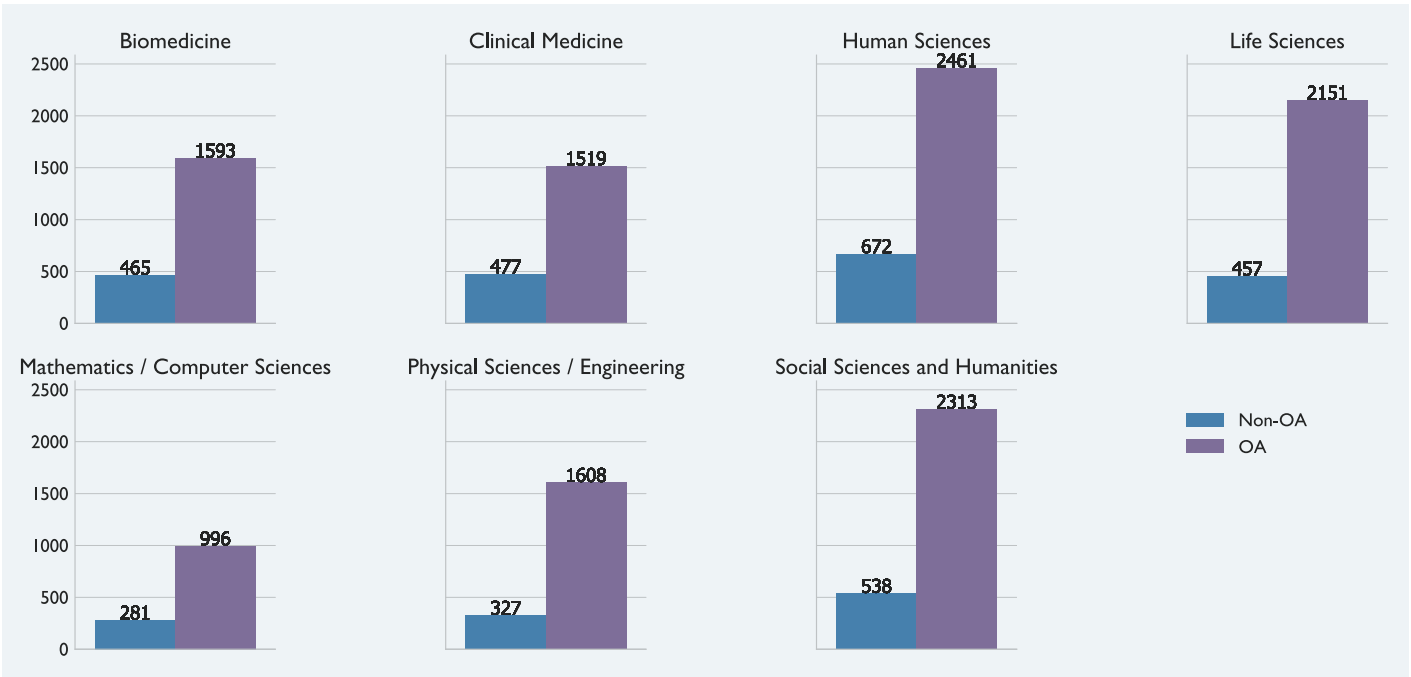


Figure 3: Average downloads by subject field, for all users

	All users		NB GLM p-values
	Non-OA	OA	
Biomedicine	465	1,593	p<0.01
Clinical Medicine	477	1,519	p<0.01
Human Sciences	672	2,461	p<0.0001
Life Sciences	457	2,151	p<0.0001
Mathematics / Computer Sciences	281	996	p<0.0001
Physical Sciences / Engineering	327	1,608	p<0.0001
Social Sciences and Humanities	538	2,313	p<0.0001
All	**424**	**1,696**	**p<0.0001**

Table 2: Average downloads, by subject

We modelled the difference in the number of cumulative downloads between OA and non-OA articles, and found a significant difference (NB GLM p<0.0001) influenced by the OA status, journal IF and university ranking (see Table 3 and Model 1 in appendix B, also available on figshare).

	Open access status
All users	269% more
Recognised users	49% more

Table 3: Modelled download gains attributable to OA

OA articles are significantly more downloaded (NB GLM p<0.0001), even by recognised users and when controlling for journal IF and institution ranking in the model.

Journal Tiers

We considered five tiers of journals, based on their IF:

- Tier 1: 0-1
- Tier 2: 1-3
- Tier 3: 3-5
- Tier 4: 5-8
- Tier 5: 8+

We looked at the averages in these tiers, in Table 4. In all tiers, OA articles received more downloads on average. The sample size for Tier 5 was small.

		Tier 1	Tier 2	Tier 3	Tier 4	Tier 5
Impact Factor		0-1	1-3	3-5	5-8	8+
Average	Non-OA	247 ±SD 3	400 ±SD 3	628 ±SD 9	887 ±SD 24	1,397 ±SD 109
	OA	1,046 ±SD 67	1,494 ±SD 40	2,407 ±SD 270	2,117 ±SD 157	2,828 ±SD 366
Median	Non-OA	176	313	478	672	1,003
	OA	826	1,101	1,409	1,683	2,322
Sample size	Non-OA	8,533	45,238	9,256	1,012	203
	OA	187	1,891	619	96	23

Table 4: Average downloads in five journal tiers for all users

Citations

The articles in this sample were all published in the first six months of 2014. With 3.75 to 4.25 years of cumulative citations possible, scholarly impact can be estimated by the number of times the articles are cited.

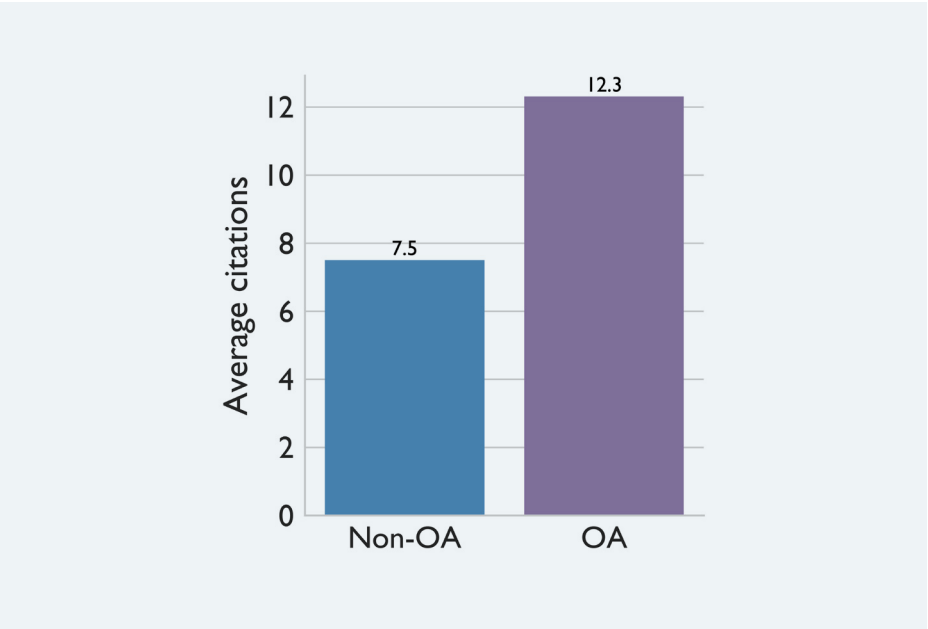


Figure 4: Average cumulative citations for OA and non-OA articles

OA articles received 1.6 times more citations on average (see Figure 4).

Figure 5 shows the different distribution of citations (using the logarithmic value) for OA and non-OA articles, with the median citation number for non-OA articles being lower than OA (4 vs. 6; average 7.5 vs 12.3).

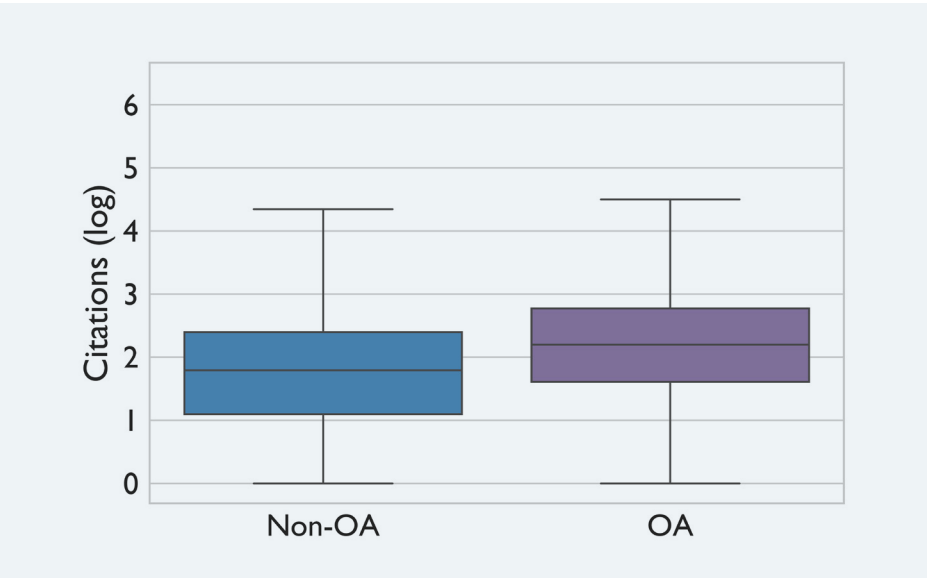


Figure 5: Distribution of (log) citations accrued over 3.75 to 4.25 years

Citations in different subject fields

The citation advantage was found to be significant for all subjects (NB GLM, at least p<0.05) except Social Sciences and Humanities, and Mathematics / Computer Sciences as shown in Table 5 and Figure 6 (showing the logarithmic value of citations).

- In all subjects, OA articles received more citations on average, compared to non-OA articles;
- The biggest gain was for Clinical Medicine with almost twice as many citations for OA articles.

	Sample size		Average		NB GLM p-values	Median	
	Non-OA	OA	Non-OA	OA		Non-OA	OA
Biomedicine	10,513	651	9.2 ±SD 0.1	14.0 ±SD 0.7	p<0.001	6	9
Clinical Medicine	11,807	532	9.3 ±SD 0.1	16.1 ±SD 0.8	p<0.001	6	10
Human Sciences	5,059	182	7.2 ±SD 0.1	10.2 ±SD 0.7	p<0.05	5	7
Life Sciences	6,832	353	7.7 ±SD 0.1	11.9 ±SD 0.8	p<0.001	6	8
Mathematics / Computer Sciences	4,712	114	4.6 ±SD 0.1	5.2 ±SD 0.6	-	2	3
Physical Sciences / Engineering	22,290	876	6.4 ±SD 0.1	10.6 ±SD 0.6	p<0.001	4	7
Social Sciences and Humanities	1,897	48	5.9 ±SD 0.2	6.3 ±SD 1.2	-	3	3
All	**63,968**	**2,812**	**7.5 ±SD 0.0**	**12.3 ±SD 0.3**	**p<0.001**	**5**	**8**

Table 5: Average and median citations, by subject

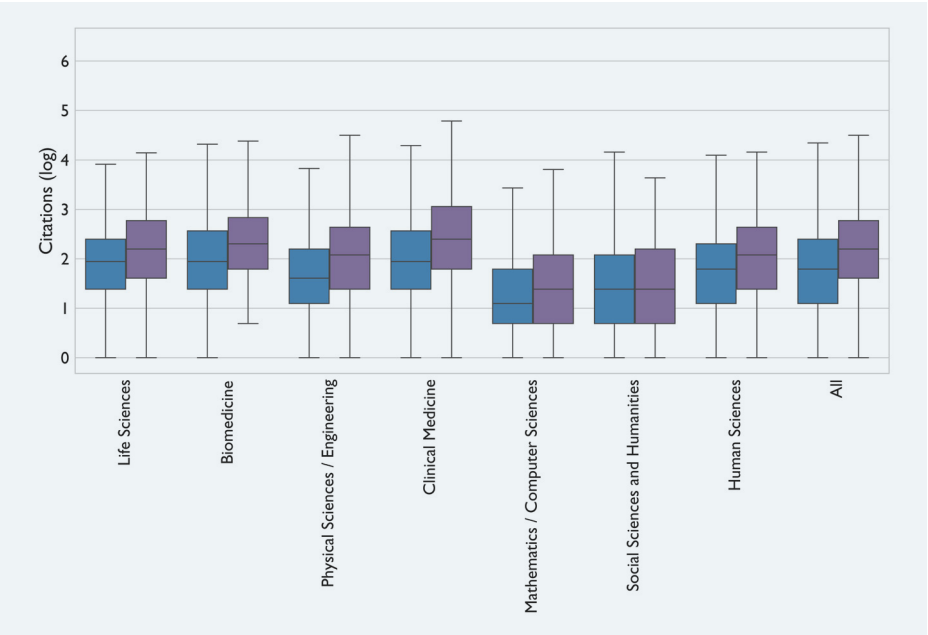


Figure 6: Distribution of (log) citations accrued over 3.75 to 4.25 years, by subject

Regional differences

The citation gain varied across regions, as shown in Table 6:

- Articles published by authors from Europe and North America, as well as those resulting from bilateral and multilateral collaborations, had a significant (at least p<0.001) average citation gain;

Figure 16 and Table 9 show the differences across subject fields.

- There was not enough data from the Africa & Middle East and Central & South America regions for comparison.

	Sample size		Average		NB GLM p-values	Median	
	Non-OA	OA	Non-OA	OA		Non-OA	OA
Africa & Middle East	2,902	39	5.9 ±SD 0.1	5.6 ±SD 0.8	-	4	4
Asia and Pacific	21,184	435	6.4 ±SD 0.1	9.9 ±SD 0.6	-	4	6
Bilateral	10,716	531	8.5 ±SD 0.1	14.2 ±SD 0.8	p<0.0001	5	9
Central & South America	1,846	22	5.6 ±SD 0.1	9.4 ±SD 1.5	-	4	7.5
Europe	12,563	1,193	7.8 ±SD 0.1	11.0 ±SD 0.4	p<0.0001	5	7
Multilateral	2,505	194	11.6 ±SD 0.3	20.2 ±SD 2.0	p<0.001	7	12
North America	10,045	301	8.6 ±SD 0.1	14.3 ±SD 0.9	p<0.0001	6	10
All	**63,968**	**2,812**	**7.5 ±SD 0.0**	**12.3 ±SD 0.3**	**p<0.0001**	**5**	**8**

Table 6: Descriptive statistics for citations, by region

Figure 7 shows the distribution of the number of citations in these regions, using logarithmic values. As expected, articles involving international collaboration performed better than the rest. In particular, articles with authors from three or more different countries received more citations (referred to here as "multilateral collaboration"), and articles with authors from two different countries ("bilateral collaboration") performed on a par with articles with authors from Europe and North America.

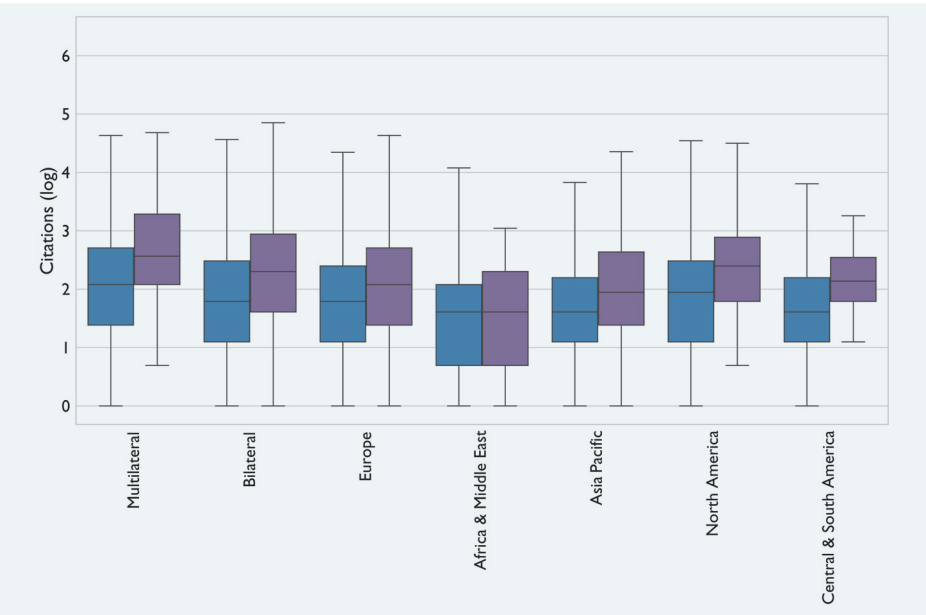


Figure 7: Distribution of (log) citations accrued since online publication, by region

Modelled cumulative citations

To model the citation impact, we used the cumulative citations from date of online publication to March 2018, and we included the IF and institution ranking as independent variables in a GLM model (see Model 2 in appendix). The difference was significant (NB GLM p<0.0001) for the three independent variables.

- This showed a significant advantage for OA compared to non-OA articles, with the model predicting there would 36% more citations over the 3.75-4.25-year period.

Altmetric attention score

OA articles would be expected to have a greater reach in mainstream media and with policy makers. Indeed, only 24.7% of non-OA articles had an Altmetric score, compared with 39.8% of OA articles. The average score for articles that received any attention was 2.4 times higher for OA articles (4.3 vs 10.3), as seen in Figure 8.

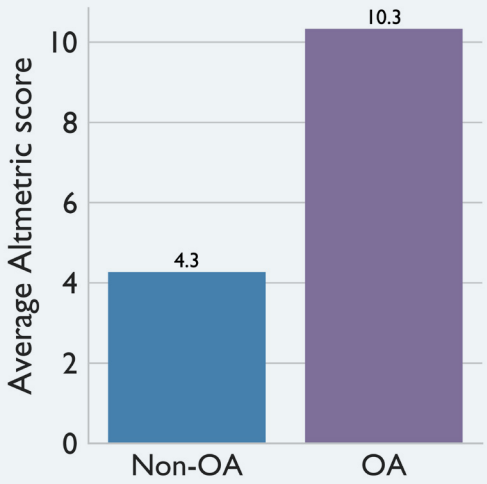


Figure 8: Average attention score for OA and non-OA articles

This picture was seen in all subject areas (see Figure 9 and Table 7):

- The largest significant (NB GLM $p<0.05$) gain was in Clinical Medicine, with 3.2 times more attention to OA articles;
- Life Sciences had the smallest gain (NB GLM $p<0.01$), but still had an average score 1.6 times higher;
- Social Sciences and Humanities had a 4.6 times higher score on average for OA articles. However, this difference was not significant, due to a small sample size and large outliers.

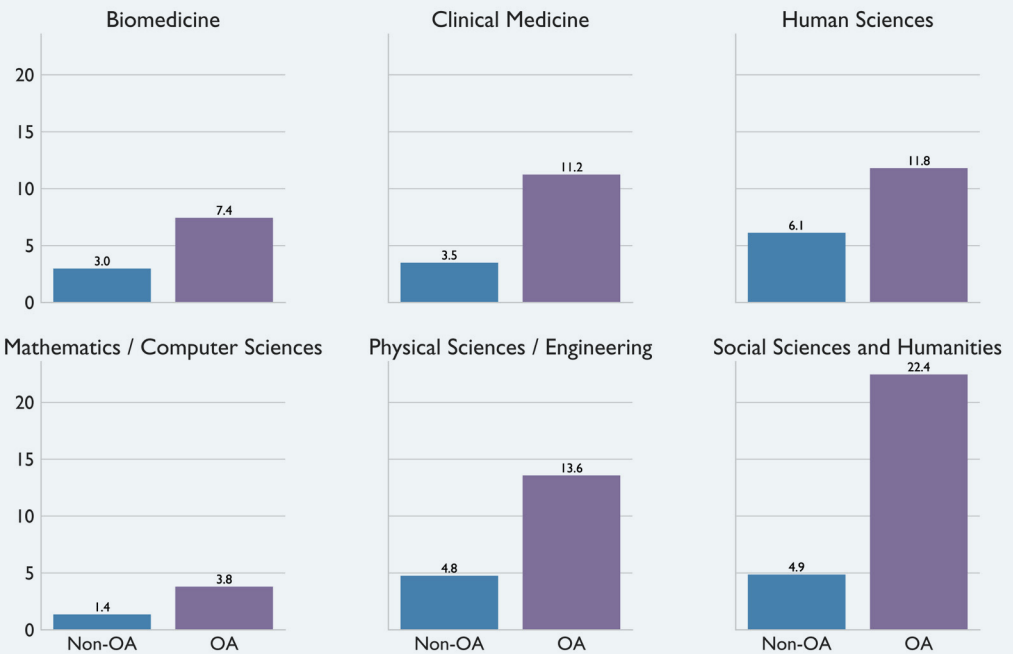


Figure 9: Average attention score for OA and non-OA articles, by subject

	Average score		Gain	p-values
	Non-OA	OA		
Biomedicine	3.0 ±SD 1.7	7.4 ±SD 0.2	2.5x	p<0.05
Clinical Medicine	3.5 ±SD 2.9	11.2 ±SD 0.2	3.2x	p<0.05
Human Sciences	6.1 ±SD 2.3	11.8 ±SD 0.4	1.9x	p<0.0001
Life Sciences	4.7 ±SD 1.8	7.5 ±SD 0.4	1.6x	p<0.01
Mathematics / Computer Sciences	1.4 ±SD 1.8	3.8 ±SD 0.1	2.7x	p<0.0001
Physical Sciences / Engineering	4.8 ±SD 6.1	13.6 ±SD 0.6	2.8x	p<0.0001
Social Sciences and Humanities	4.9 ±SD 16.7	22.4 ±SD 0.7	-	-
All	**4.1 ±SD 0.1**	**10.1 ±SD 1.6**	**2.5x**	**p<0.0001**

Table 7: Average attention score for OA and non-OA articles, by subject

The distributions of Altmetric scores were different for OA and non-OA articles, as shown in Figure 10. OA articles had a wider range of scores, with larger outliers.

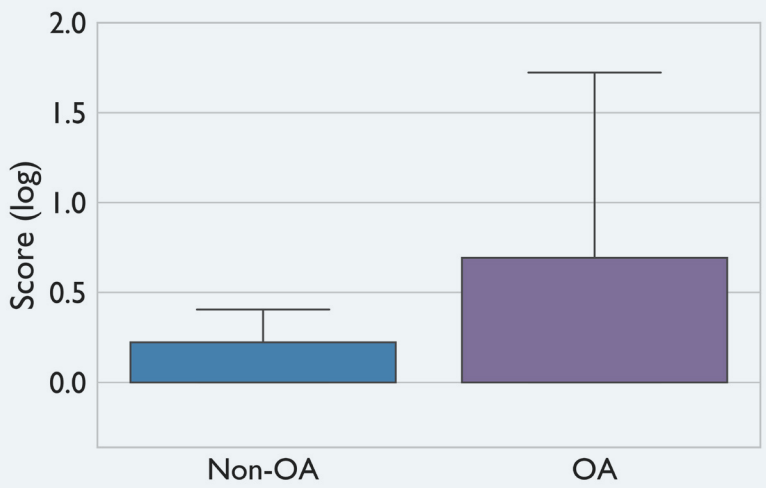


Figure 10: Distribution of (log) Altmetric scores for OA and non-OA articles

Modelled Altmetric Scores

To model the impact of OA on the overall score, we included the IF and the ranking of the institution (see Model 3 in Appendix B). This model showed a significant (NB GLM p<0.0001) advantage for OA:

- OA articles attracted significantly more attention than non-OA articles; with a 251% higher score.

Subject fields were affected in different ways, as shown in Figure 11. All subject fields have a higher score for OA articles, but we also have three clear groups:

- Group A: non-OA articles already attract some attention, but there is an OA benefit (Life Sciences, Biomedicine, Clinical Medicine, and Human Sciences);
- Group B: low attention for non-OA articles, but substantially higher for OA articles (Physical Sciences / Engineering and Social Sciences and Humanities);
- Group C: no substantial difference (Mathematics / Computer Sciences).

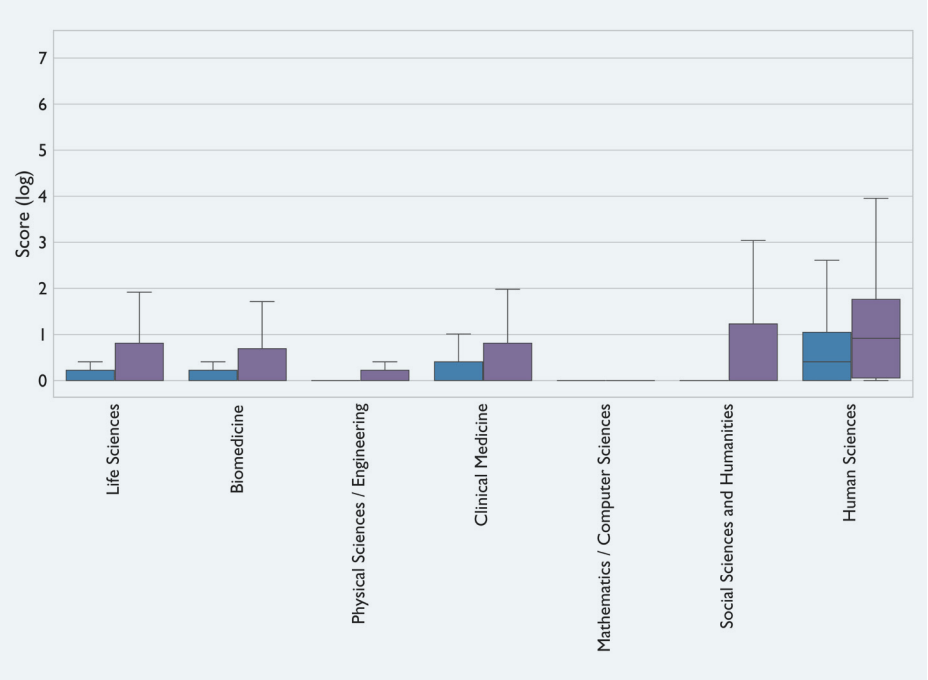


Figure 11: Distribution of (log) Altmetric scores for OA and non-OA articles, by subject

Regions were also affected in different ways. For each, articles received greater attention when published OA (see Figure 12). The difference, however, was not significant in the Asia Pacific, and Central & South America regions due to a lower number of articles.

We again observe three similar groups:

- Group A: non-OA articles attract some attention (Europe, North America, Bilateral and Multilateral collaborations[iii]);
- Group B: low attention for non-OA articles, but substantially better for OA articles (Asia Pacific);
- Group C: no substantial difference (Africa & Middle East, and Central & South America).

iii. Where Bilateral and Multilateral refer to articles with authors from two, and three or more, regions respectively.

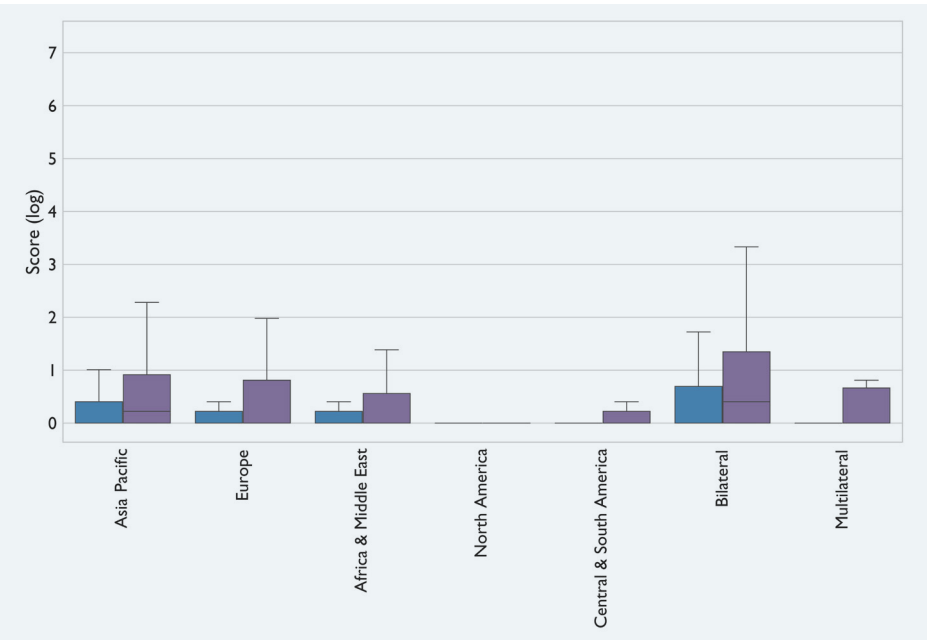


Figure 12: Distribution of (log) Altmetric scores for OA and non-OA articles, by region

News

Articles cited in news media outlets likely have a greater impact on the general public. News outlets tracked by Altmetric include general international and national news sources, and more specialist news sources (such as industry magazines and newsletters).

Figure 13 shows the average news scores for OA and non-OA articles (mentioned articles only). OA articles had gained 1.9 times more news attention than non-OA articles.

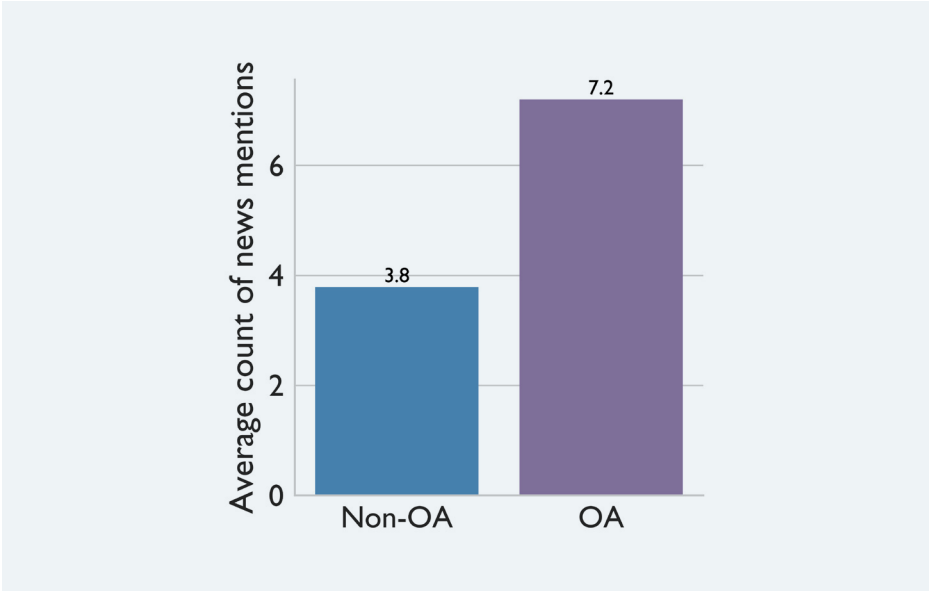


Figure 13: Average news mentions for OA and non-OA mentioned articles

The difference in average news mentions between OA and non-OA articles, shown in Figure 14 and Table 9, was only significant (NB GLM $p<0.05$) in Physical Sciences / Engineering, with 2.3 times more news mentions on average.

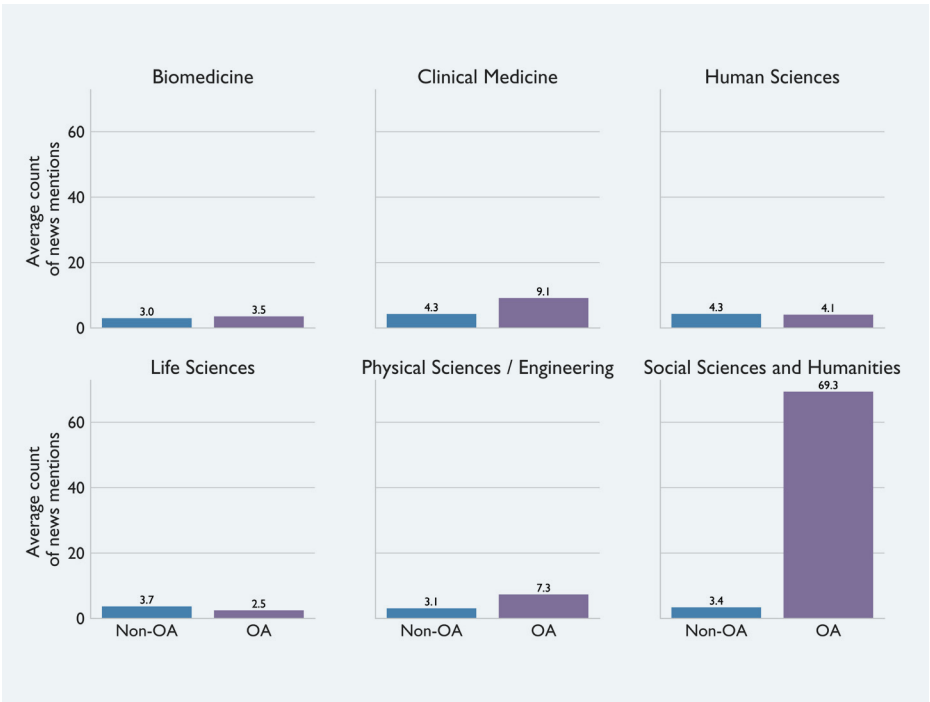


Figure 14: Average news mentions for OA and non-OA mentioned articles, by subject

	Average news mentions		p-values
	Non-OA	OA	
Biomedicine	3.0 ±SD 0.6	3.5 ±SD 1.4	-
Clinical Medicine	4.3 ±SD 0.7	9.1 ±SD 3.5	-
Human Sciences	4.3 ±SD 0.4	4.1 ±SD 0.8	-
Life Sciences	3.7 ±SD 0.4	2.5 ±SD 0.4	-
Physical Sciences / Engineering	3.1 ±SD 0.3	7.3 ±SD 4.4	p<0.05
Social Sciences and Humanities	3.4 ±SD 0.9	69.3 ±SD 68.3	-
All	**3.8 ±SD 0.2**	**7.2 ±SD 2.2**	**p<0.0001**

Table 8: Average news mentions for OA and non-OA mentioned articles, by subject

We modelled the number of mentions in the news with the IF and the institution ranking (see Model 4 in Appendix B), and found a significant (NB GLM p<0.0001) OA benefit:

- OA articles in the model received 219% more news mentions than non-OA articles;
- The difference within each subject field was not significant in most cases, due to small samples and skewed distributions.

Policy

Altmetric tracks policy documents from government bodies, policy institutes and non-governmental organisations, such as guidelines, reports, and white papers. Policy-making bodies covered include the World Health Organisation (WHO), UNICEF, and the UK's National Institute for Health and Care Excellence (NICE).

As with news mentions, OA articles received on average more mentions from policy-making bodies than non-OA articles: their average citation rate was 1.2 times higher (NB GLM p<0.001), as shown in Figure 15 (mentioned articles only).

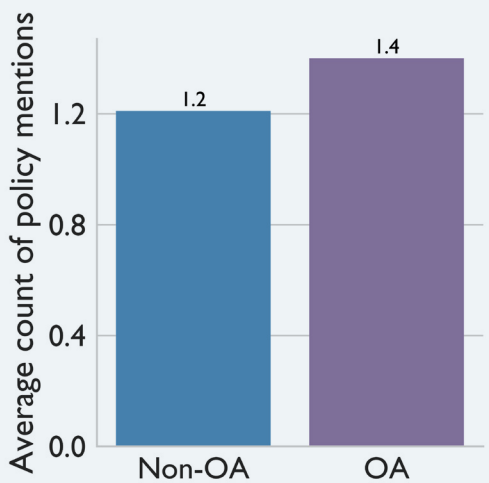


Figure 15: Average count of policy mentions for OA and non-OA mentioned articles

- OA did not give an advantage in all subject fields;
- OA articles in Physical Sciences / Engineering had on average more mentions (1.25 times more) in policy documents (NB GLM p<0.05);
- In Life Sciences, OA articles received fewer mentions (0.9 times) on average than non-OA articles (NB GLM p<0.05);
- In the other subject fields, the differences were not significant.

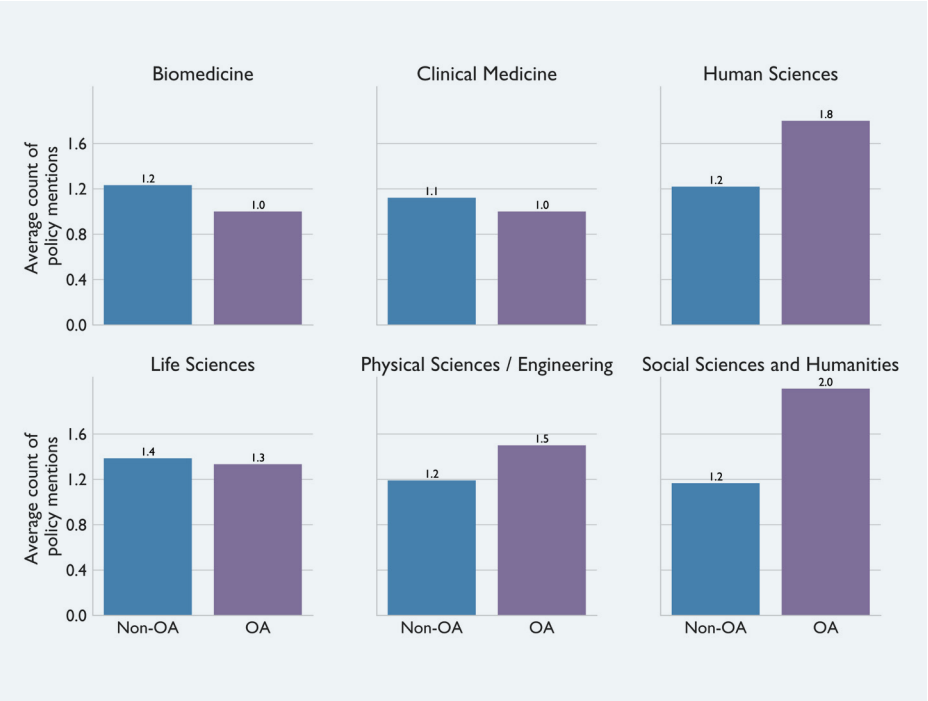


Figure 16: Average policy mentions for OA and non-OA mentioned articles, by subject

	Average policy mentions		NB GLM p-values
	Non-OA	OA	
Biomedicine	1.2 ±SD 0.0	1.0 ±SD 0.1	-
Clinical Medicine	1.1 ±SD 0.0	1.0 ±SD 0.0	-
Human Sciences	1.2 ±SD 0.4	1.8 ±SD 0.0	-
Life Sciences	1.4 ±SD 0.1	1.3 ±SD 0.1	p<0.05
Physical Sciences / Engineering	1.2 ±SD 0.4	1.5 ±SD 0.1	p<0.05
Social Sciences and Humanities	1.2 ±SD 1.0	2.0 ±SD 0.1	
All	**1.2 ±SD 0.0**	**1.4 ±SD 0.1**	**p<0.001**

Table 9: Average policy mentions for OA and non-OA mentioned articles, by subject

We included the number of mentions by policy documents in a similar model (see Model 5 in Appendix B) and found the three variables significant:

- OA articles had 166% more policy mentions than non-OA articles;
- The breakdown per category was not significant for most subjects, because the sample was too small and distribution skewed.

OA is associated with more mentions overall, as well as when focusing on platforms with a direct societal impact.

Comparing descriptive statistics for the modelled subsets of articles and the full dataset

The model we used was only applied to the set of articles in journals with an IF and with an institution ranking. To understand the impact this had on the samples, for each of these variables we compared their subsets with the full dataset. For example, we compared the average downloads for articles in a journal with an IF, with the average downloads for all articles, including those in journals without an IF. The descriptive statistics in Table 10 for institution ranking and Table 11 for the IF show broadly comparable values for the average and median downloads, as well as cumulative citations.

Table 10: Comparison of average and median downloads for articles with an institution ranking, and for all articles

		With ranking		All	
		Non-OA	OA	Non-OA	OA
	Sample size	35,978	1,445	70,641	3,000
Downloads	Average	441 ±SD 3	1,739 ±SD 92	416 ±SD 2	1,676 ±SD 62
	Median	327	1,204	310	1,175
Citations	Average	7.4 ±SD 0.1	12.0 ±SD 0.4	7.2 ±SD 0.0	12.1 ±SD 0.3
	Median	5	8	4	8

		With Impact Factor		All	
		Non-OA	OA	Non-OA	OA
	Sample size	35,978	1,445	70,641	3,000
Downloads	Average	424 ±SD 2	1,696 ±SD 66	416 ±SD 2	1,676 ±SD 62
	Median	318	1,187	310	1,175
Citations	Average	7.5 ±SD 0.0	12.3 ±SD 0.3	7.2 ±SD 0.0	12.1 ±SD 0.3
	Median	5	8	4	8

To further consider the impact of institution prestige, we compared the descriptive statistics for two sets of articles: those with a first author affiliated with an institution in the top 200 of the THE ranking, and those at other institutions. Articles published by researchers at the top 200 ranked institutions receive more downloads than those at other institutions, but there appears to be an OA benefit for each set of articles.

Table 11: Comparison of averages and medians for articles in journals with Impact Factors, and for all articles

			Non-Top 200 institutions	Top 200 institutions
Downloads	Sample size	Non-OA	58,040	12,601
		OA	2,373	627
	Average	Non-OA	395 ±SD 2	515 ±SD 6
		OA	1,595 ±SD 61	1,979 ±SD 187
	Median	Non-OA	298	374
		OA	1,142	1,316
Citations	Sample size	Non-OA	58,040	12,601
		OA	2,373	627
	Average	Non-OA	6.9 ±SD 0.0	8.5 ±SD 0.1
		OA	11.5 ±SD 0.3	14.0 ±SD 0.6
	Median	Non-OA	4	5
		OA	7	10

Table 12: Usage and citations of articles with a first author affiliated with a top 200 ranked institution

UK case study

Summary
For the UK study, we modelled the impact of OA on three standard metrics: the usage of articles (downloads), citations, and Altmetric scores. The OA articles had been published from January to December 2016, while the non-OA articles were published one or two years prior, from January 2014 to December 2015.

The most recent articles in the dataset were published in December 2016, so at the time of analysis were just 15 months old. It is relatively early therefore to assess scholarly impact, so our results here are only an early signal. The control group of non-OA articles was also up to two years older. To account for this, we used the citations at two years after publication of the earliest articles in each sample, and the Altmetric score at one year after publication.

Despite being a relatively recent sample, OA gave a significant advantage on all metrics, as shown in Table 13.

Impact	Measure	Averages	Model predictions
Usage	Downloads per month	During their lifetime (shorter for OA articles), OA articles were downloaded 3.2 times more than non-OA articles.	The model predicted that OA articles are downloaded 607% more per month than non-OA articles.
Citations	2-year citations	After two years, OA articles had received 1.6 times more citations than non-OA articles.	OA articles are cited 30% more than non-OA articles.
Altmetric	Score at one year after publication	The score for OA articles was 3.2 higher than that of non-OA articles.	OA articles have a score 444% higher than non-OA articles.

Table 13: Summary of impact of three metrics for the UK case study

Downloads
Springer Nature tracks the number of downloads of each article on their platforms. We used the monthly values for each article. Using a generalised additive model, we found that the downloading rates for OA and non-OA articles were significantly different (NB GLM $p<0.0001$), as shown on Figure 17, where the shading represents the 95% confidence interval.

Even though the OA articles were published later than the non-OA articles, the average cumulative downloads since publication was 3.2 times higher for OA articles (1,772 vs. 555). To establish if this difference was significant, we modelled the monthly downloads as the dependent variable, using the IF and the number of days since publication as independent variables (see Model 6 in the appendix).

The model shows that:

- IF and OA are both significant (NB GLM $p<0.0001$);
- As time passes after publication, the relative impact of OA starts to wane, but there remains a significant difference in the cumulative downloads. This could be due to the initial 'burst' of interest as a new article is highlighted in journal newsletters, social media, etc.

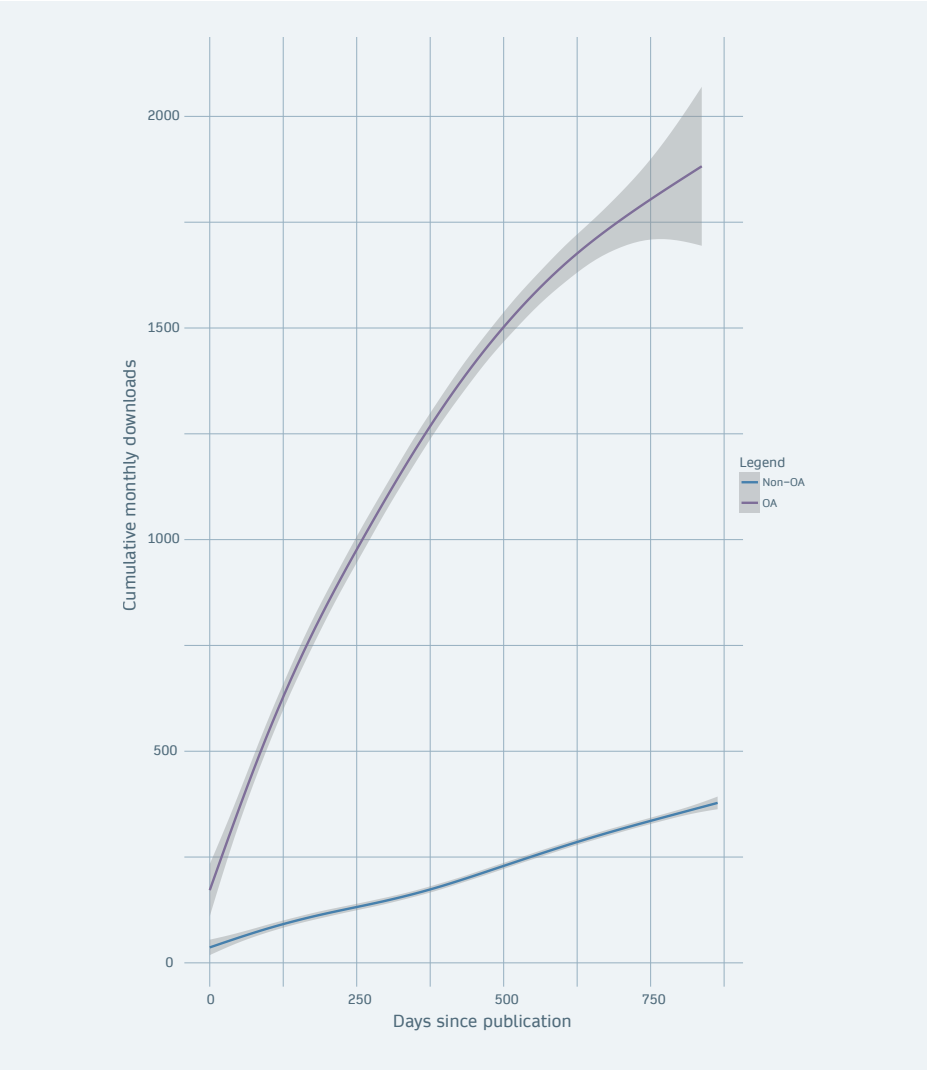


Figure 17: Cumulative monthly download trends (using a generalised additive model)

The order of magnitude of the effect of the IF was much lower than that of the OA status, suggesting that the availability of the research was more important than the journal in which it was published.

Citations

The non-OA articles in the UK study were published between January 2014 and December 2015, and the OA articles in our sample were published in 2016. For a fair comparison, we looked at citations occurring over the same length of time: the longest that all articles had been published for. Online publication date was distributed evenly throughout the year for both sub-samples, evening out possible seasonal effects.

Citations within two years of publication

The most recent sample, the OA articles, were published from January to December 2016, so could potentially be cited in 2016 and 2017. The control sample was published in 2015 and 2016, so we included citations respectively from 2015 and 2016, and 2016 and 2017. On average, after two years, OA articles had gained 1.6 times more citations than non-OA articles (NB GLM $p<0.001$).

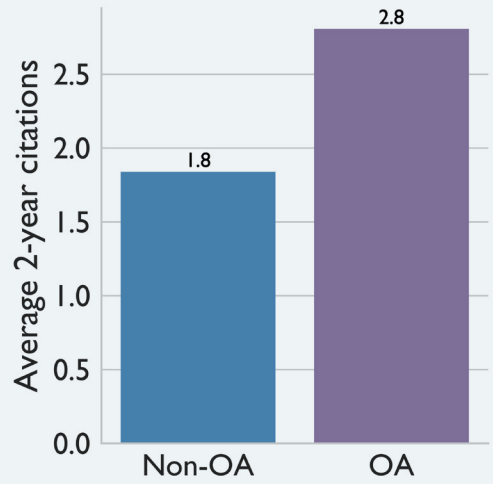


Figure 18: Average cumulative citations two years after publication

We looked at the difference across journal subjects: the median was higher for OA in every subject, apart from Clinical Medicine for which the median was similar in both samples; the distribution showed higher quartile values for the OA articles. As shown in Table 14, this was significant (at least NB GLM $p<0.05$) for all subjects except Mathematics / Computer Sciences.

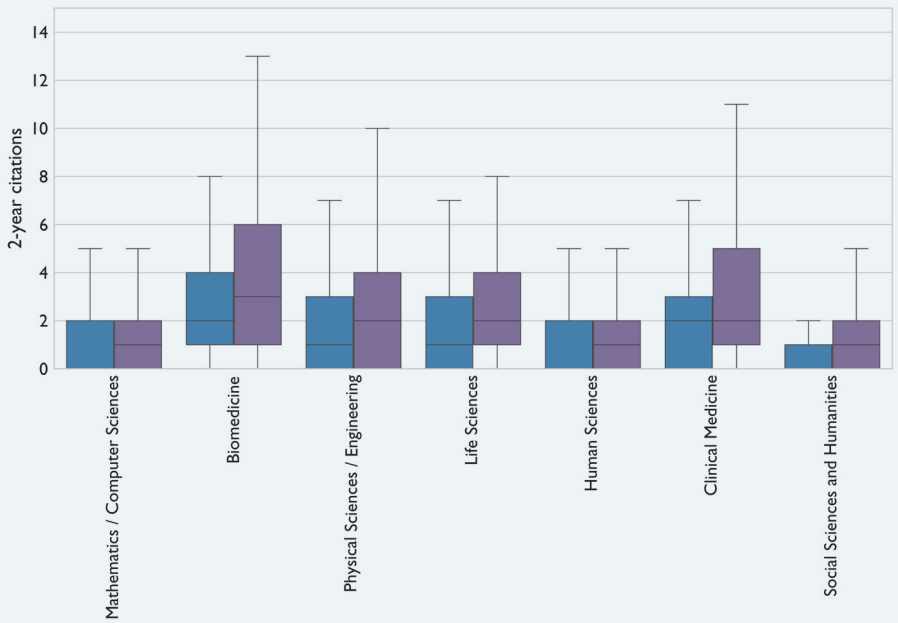


Figure 19: Distribution of citations at two years, by journal subject

OA articles in all subjects had more citations on average, two years after publication, than non-OA articles:

- The biggest gain for OA articles was found in Clinical Medicine, with 1.6 times more citations on average, and Biomedicine with 1.5 times more citations;
- All other subjects had at least a gain of 1.4 times more citations on average for OA articles.

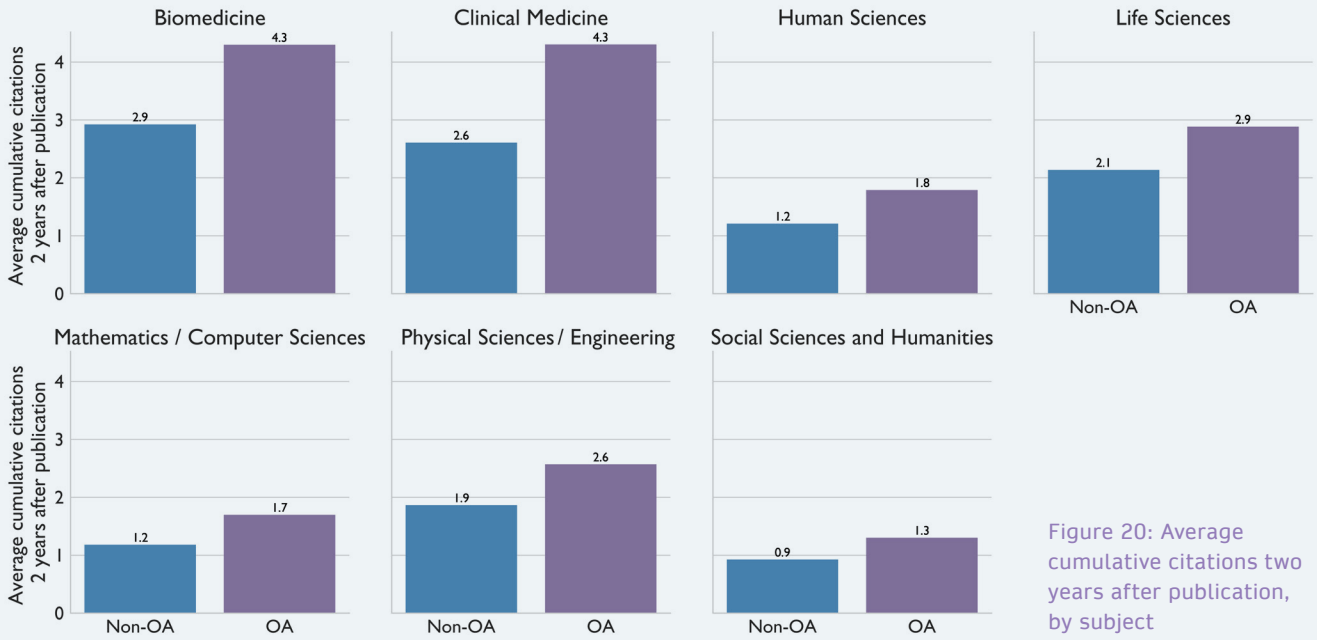


Figure 20: Average cumulative citations two years after publication, by subject

	Average		NB GLM p-values	Median	
	Non-OA	OA		Non-OA	OA
Biomedicine	2.9 ±SD 0.1	4.3 ±SD 0.3	p<0.0001	2	3
Clinical Medicine	2.6 ±SD 0.1	4.3 ±SD 0.3	p<0.0001	2	2
Human Sciences	1.2 ±SD 0.1	1.8 ±SD 0.1	p<0.0001	0	1
Life Sciences	2.1 ±SD 0.1	2.9 ±SD 0.3	p<0.001	1	2
Mathematics / Computer Sciences	1.2 ±SD 0.1	1.7 ±SD 0.2	-	0	1
Physical Sciences / Engineering	1.9 ±SD 0.1	2.6 ±SD 0.1	p<0.0001	1	2
Social Sciences and Humanities	0.9 ±SD 0.1	1.3 ±SD 0.1	p<0.05	0	1
All	**1.8 ±SD 0.0**	**2.8 ±SD 0.1**	**p<0.001**	**1**	**2**

Table 14: Average cumulative citations two years after publication, by subject

We applied a model (see Model 7 in Appendix B) taking into account the IF and OA status. Both are found to be significant (NB GLM p<0.0001):

- OA articles had 30% more citations within a two-year window than non-OA articles.

Although this represents a short-term view of citations, there appears to be an OA citation benefit in this sample.

Altmetric

For every article tracked, Altmetric provides the historical score at one year after publication. This is a comparable metric for all articles in both samples. In their first year, OA articles were mentioned more often on average than non-OA articles (22.4% vs. 5.4%). When considering only the articles that had attracted any attention, the average score for OA after one year was 3.2 times higher than non-OA. Table 15 shows the differences between OA and non-OA in terms of mentioned articles (as a percentage). The average historical score for OA articles at one year was higher in all subject fields.

	Percentage of articles mentioned within one year of publication		
	Non-OA	OA	p-values
Mathematics / Computer Sciences	2.7%	10.8%	p<0.001
Biomedicine	5.2%	27.8%	p<0.0001
Physical Sciences / Engineering	2.3%	8.7%	p<0.0001
Life Sciences	5.2%	20.8%	p<0.0001
Human Sciences	8.8%	37.4%	p<0.0001
Clinical Medicine	6.5%	23.8%	p<0.0001
Social Sciences and Humanities	4.8%	15.5%	p<0.0001
All	**5.4%**	**22.4%**	**P<0.0001**

Table 15: Mentioned OA and non-OA articles at one year after publication, by subject

Table 16 shows the median and average Altmetric scores across subject fields.

- There was an OA advantage, with a significant (NB GLM p<0.0001) 3.2 times more mentions on average after a year;
- The biggest relative advantage was seen in Social Sciences and Humanities (6.7 times higher on average) and Clinical Medicine (5.0 times higher on average).

	Sample size		Average		NB GLM p-values	Median	
	Non-OA	OA	Non-OA	OA		Non-OA	OA
Biomedicine	195	178	1.1 ±SD 0.4	3.8 ±SD 0.6	p<0.0001	0	0.75
Clinical Medicine	235	164	1.2 ±SD 0.4	5.7 ±SD 1.9	p<0.0001	0	0.5
Human Sciences	443	296	3.3 ±SD 1.3	5.7 ±SD 1.4	p<0.0001	0	1.3
Life Sciences	110	67	1.7 ±SD 0.7	3.8 ±SD 0.8	p<0.0001	0	0.85
Mathematics / Computer Sciences	392	163	0.1 ±SD 0.1	0.3 ±SD 0.1	p<0.0001	0	0
Physical Sciences / Engineering	794	323	0.2 ±SD 0.1	1.0 ±SD 0.3	p<0.0001	0	0
Social Sciences and Humanities	214	108	0.2 ±SD 0.1	1.5 ±SD 0.4	p<0.0001	0	0
All	**2,455**	**1,360**	**1.1 ±SD 0.2**	**3.5 ±SD 0.4**	**p<0.0001**	**0**	**0.25**

Table 16: Median and average Altmetric scores at one year for OA and non-OA mentioned articles, by subject

Using a similar model (see Model 8 in Appendix B), we find the OA and IF significantly (NB GLM p<0.0001) affect the Altmetric score after one year.

- OA articles had been mentioned 444% more than non-OA articles after the first year of publication.

Altmetric tracks multiple attention sources. As in the global study, we looked at mainstream media and policy mentions. Mainstream media attention often happens shortly after publication. Although the UK study is smaller and includes more recently published articles, we found significant differences in the level of mentions in news articles. However, mentions in policy documents are slower and the frequency relatively low, so there was not enough policy document mentions to be able to compare OA and non-OA articles.

News

Although OA articles were published more recently than non-OA articles, during their lifetime they attracted twice as many media mentions as non-OA articles, as shown in Table 17.

	Average number of days since publication	Average mentions in mainstream media
OA	629 days	9.2
Non-OA	1,189 days	4.7

Table 17: Average mainstream media attention in the UK case study

The increased mainstream media attention to OA articles affected all subjects, as can be seen in Figure 21 showing the distribution of news articles mentioning journal articles. The largest difference can be seen in Physical Sciences / Engineering, and Social Sciences and Humanities, where the OA median is almost as high as the non-OA third quartile (the top of the box plot). There is not enough data for Mathematics / Computer Sciences, so the box plot just shows outliers.

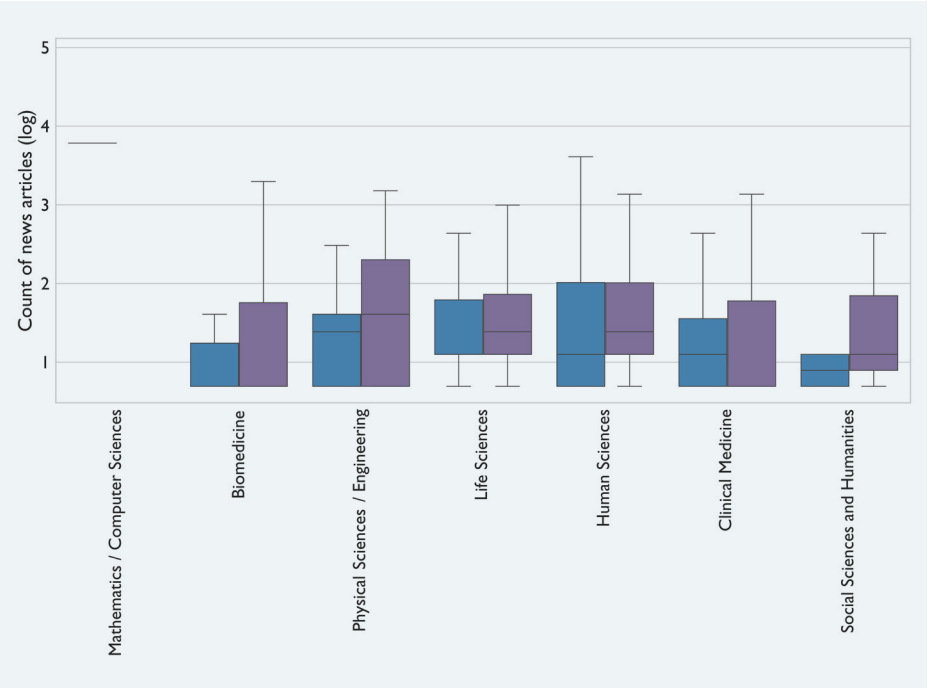


Figure 21: Distribution of news article counts across all subjects, for mentioned articles

Discussion and conclusions

Our findings indicate a strong OA advantage for articles in hybrid journals, considering usage, citations, and Altmetric data (including mentions in news and policy documents). In both the global and UK studies, OA articles were downloaded with much greater frequency than non-OA articles: on average four times more in the global study and 3.2 times more for the UK study. This effect is seen even when accounting for IF and institution ranking in the model.

In the global dataset, we differentiated downloads by users from recognised institutions and non-recognised users, finding that OA articles were downloaded more often by both groups (1.5 times more for recognised users). The fact that there is increased OA usage from recognised users may reflect OA articles' greater discoverability via search engines, or through sharing sites. At the same time, it may reflect a level of selection bias, where authors are choosing the OA model for their better research, which is more likely to be read. Neither of these possibilities can be substantiated by the research done here. However, the wider increase in downloads from other (i.e. non-registered) users shows a clear benefit of OA for those accessing articles from outside of academic institutions. A comparison of usage for the whole dataset and for the subset of articles with associated journal IFs and institutional rankings confirmed higher usage of OA articles. A usage benefit was also found across all subjects.

For the UK data, the usage benefit of OA started soon after publication. Using monthly data, we showed that not only were OA articles consistently downloaded more than non-OA articles, but also that the rate of increase in usage was steeper.

Citations of OA articles were also significantly higher in both studies. Cumulative citations in the UK case study showed that within only two years, OA articles had been cited on average six times more than non-OA articles. The fact that higher citations were found for the OA articles from a consistent set of UK institutions is supportive of a discrete benefit of OA, rather than the advantage being due to a selection bias. The articles from the global study (which had been published for longer: 3.75 to 4.25 years) showed a similar result, with OA articles attracting on average 1.6 times more cumulative citations. These results were obtained by including the IF (as a proxy for perceived journal prestige; in both studies), journal subject field (in both studies), and ranking of the institution (in the global study) to correct for possible confounding factors. The citation advantage was found across all subjects, except Mathematics / Computer Sciences (both UK and global studies), and Social Sciences and Humanities (global study only).

Both studies show increased attention for OA articles compared to non-OA, as tracked by Altmetric data. We first considered the overall Altmetric weighted score, which was significantly higher for OA articles in both studies. In the global study, OA articles achieved a score 251% higher than non-OA articles. The effect is also seen when looking at attention via mentions in mainstream media and policy documents. In the global study, mainstream media mentions were higher for OA articles, by 219%. Policy documents included 166% more mentions of OA articles in the global study.

As one of the first large-scale analyses of hybrid article data, this white paper sets out a strong case for an OA impact advantage

In both these cross-disciplinary studies of OA in hybrid journals, globally and UK-focused, we found a consistent benefit for publishing OA in terms of usage (downloads), citations, and Altmetric score (overall, and in mainstream media and policy documents). These metrics are related: the availability of articles contributes to their chances of being cited by other researchers and mentioned by news outlets and policymakers. Our global study showed that OA articles were more often downloaded both inside and outside universities, and the UK study showed that the benefit started soon after publication, with a higher downloading rate. This usage almost certainly contributed to increased citations and mentions. This is also consistent with previous studies that have looked at the academic and societal impact of OA in one subject[28] or one journal[29] or across subjects and journals[30].

Whilst our findings in this report show strong evidence of an OA advantage, we acknowledge that there are a number of other factors that may contribute to increased downloads, citations, and mentions that are not addressed here. We have controlled for a number of variables: perceived journal prestige (IF), perceived institutional prestige (ranking), geography, and subject. However the geographic distribution of OA in 2014 means that some regions (Asia Pacific, and Central & South America) are not sufficiently represented here. Further, the reduction of the sample size (for modelling) to include only articles that had available data on all variables, is a limitation.

A further limitation is the focus in this study on relatively recently published articles, which reduces the cumulative usage, citations, and mentions available for analysis. The global dataset was most appropriate for reviewing citation data, as 3.75 to 4.25 years had passed since their publication. As the UK dataset was selected to examine the impact of the Jisc Compact agreement, and therefore takes two different time periods, these results should only be seen as directional. Compared to the global study, the UK study provided a more homogeneous example, with articles published non-OA (before 2016) and others OA (in 2016). However, to include more explanatory variables would have further reduced the sample size. We chose the variables from the global study that showed the most impact (namely, the IF). Additional explanatory variables may have resulted in a smaller dataset to analyse.

Other factors that we have not controlled for in these studies include whether authors select their 'best' work when choosing whether or not to publish OA; selection biases relating to funding, or prior publications; the impact of sharing articles (green OA, pre- and post-prints, repositories, or other sharing services such as ResearchGate or Springer Nature's SharedIt); and journal promotional activities. It is also not possible from these results to quantify the wider impact of OA on society, in terms of societal or economic impact.

As one of the first large-scale analyses of hybrid article data, this white paper sets out a strong case for an OA impact advantage, subject to the potential limitations acknowledged here. We would encourage other publishers to conduct similar analyses and to continue to build on a shared understanding of the benefits of hybrid journals and the effects of choosing OA, both to provide further insights to authors on the benefits of OA, and ultimately in supporting a transition to OA that benefits research, institutions, funders, and the world at large.

Acknowledgements

We would like to thank the following people for their contributions to this case study:

Organisation	Name	Job title
Digital Science	Martin Szomszor	Chief Data Scientist
Springer Nature	Harald Wirsching	Vice President, Strategy & Market Intelligence
	Emma Goldsmith	Research & Development Coordinator
	Veronika Spinka	Senior Manager Open Access, Global Hybrid OA Initiatives
	Mirko Bitsch	Specialist Business Support, Process & Content Management
	Kelly Duoos	Marketing Manager

Figures & tables

Appendices

Appendix A: models

Global study
The diagram Figure 22 shows the variables used in the model for the global study

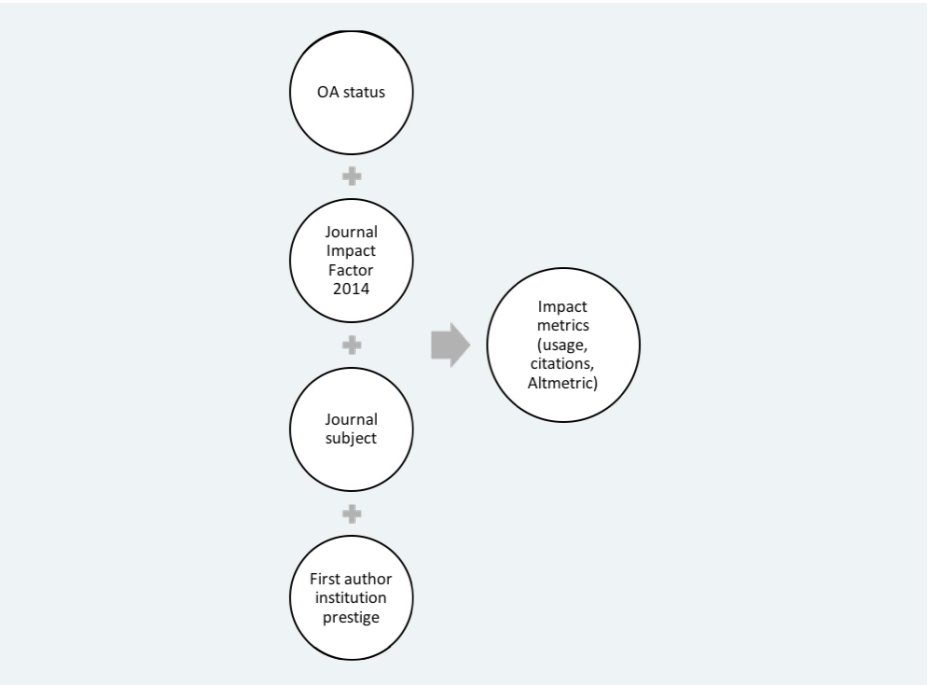


Figure 22: Graphical representation of the model used for the global study

UK case study
The diagram Figure 23 depicts the variables used in the model for the UK study.

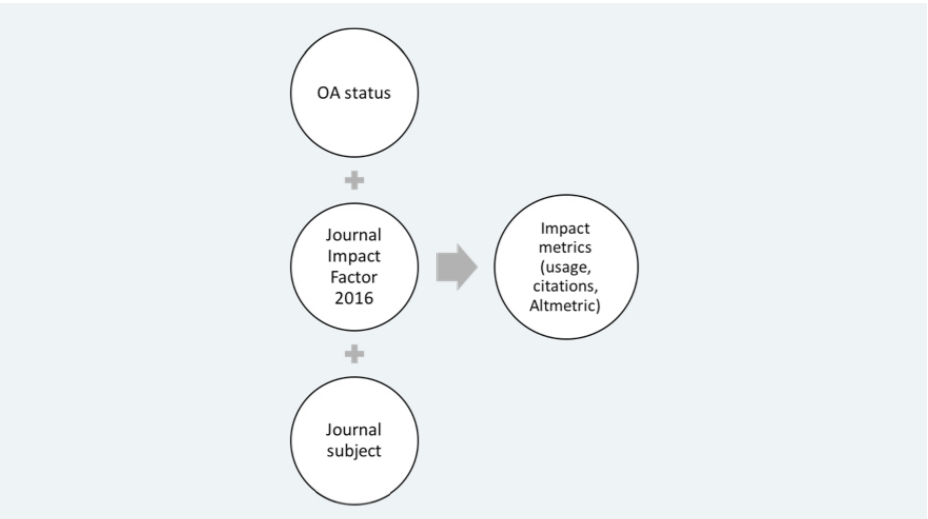


Figure 23: Graphical representation of the model used for the UK study

Appendix B: results from models

Global study

In the following models, df refers to the dataframe with all articles (OA=0 for non-OA, OA=1 for OA). The models were programmed in R (glm.nb of NB GLM or glmer.nb for NB GLMM). With the glmer.nb models, the dataframe df_gather was used. This is the same data but with a value for each year.

The following variables were used

- usage_total: all downloads
- oa: 1 if oa, 0 if non-oa
- if_2016: impact factor 2016
- the_score_first: overall score of the first institution
- category: journal subject
- usage_registered: users connecting from institutions recognised by sn
- times_cited: dimensions data for citations
- cites_in_year: number of citation for a specific year
- citation_year: the year corresponding to cites_in_year
- score: overall altmetric score (0 when no score)
- count_news: altemtric news (0 when no news)
- count_policy: altmetric policy (0 when no policy)

MODEL 1: Usage (Downloads)

This shows the raw results from the GLM in R. A summary of the coefficient and how it affects values is after the three models.

Downloads (all users)

glm.nb(usage_total ~ oa + category + if_2016 + the_score_first, data = df)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	5.3384452	0.0134438	397.093	< 0.0000000000000002	***
oaOA	1.3061621	0.0172892	75.548	< 0.0000000000000002	***
categoryClinical Medicine	-0.0725176	0.0122762	-5.907	3.48E-09	***
categoryHuman Sciences	0.4943884	0.0140537	35.178	< 0.0000000000000002	***
categoryLife Sciences	0.0772432	0.0133782	5.774	7.75E-09	***
categoryMathematics / Computer Sciences	-0.2735687	0.0147372	-18.563	< 0.0000000000000002	***
categoryOther	-0.0321374	0.0314345	-1.022	0.307	
categoryPhysical Sciences / Engineering	-0.1577603	0.0104481	-15.099	< 0.0000000000000002	***
categorySocial sciences and humanities	0.3543982	0.019965	17.751	< 0.0000000000000002	***
if_2016	0.2528368	0.0031893	79.276	< 0.0000000000000002	***
the_score_first	0.0039022	0.0001752	22.278	< 0.0000000000000002	***

Recognised Downloads

glm.nb(formula = usage_registered ~ oa + category + if_2016 + the_score_first, data = df)

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	5.3483055	0.0132105	404.853	< 0.0000000000000002	***
oaOA	0.3987098	0.0170074	23.443	< 0.0000000000000002	***
categoryClinical Medicine	-0.0902322	0.0120635	-7.48	7.44E-14	***
categoryHuman Sciences	0.4938304	0.0138092	35.761	< 0.0000000000000002	***
categoryLife Sciences	0.0608241	0.0131463	4.627	3.7153E-06	***
categoryMathematics / Computer Sciences	-0.2834365	0.0144825	-19.571	< 0.0000000000000002	***
categoryOther	-0.0357293	0.0308896	-1.157	0.247	
categoryPhysical Sciences / Engineering	-0.1602966	0.0102669	-15.613	< 0.0000000000000002	***
categorySocial sciences and humanities	0.3415654	0.0196181	17.411	< 0.0000000000000002	***
if_2016	0.2460744	0.0031339	78.521	< 0.0000000000000002	***
the_score_first	0.0035932	0.0001721	20.876	< 0.0000000000000002	***

SUMMARY EFFECTS

To translate the coefficient of a GLM, we take the exponential:

		OA			IF			THE		
		COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE
1	Usage total	1.3061621	3.69197705	269%	0.2528368	1.28767311	29%	0.0039022	1.00390982	0.4%
	Recognised users	0.3987098	1.48990119	49%	0.2460744	1.27899473	28%	0.0035932	1.00359966	0.4%

Model 2: Cumulative citations

glm.nb(times_cited ~ oa + category + if_2016 + the_score_first, data = df)

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.1127157	0.0204305	54.463	< 0.0000000000000002	***
oaOA	0.3102008	0.0258811	11.986	< 0.0000000000000002	***
categoryClinical Medicine	-0.0806615	0.0185096	-4.358	0.0000131	***
categoryHuman Sciences	-0.0741604	0.0214338	-3.46	0.00054	***
categoryLife Sciences	-0.0233162	0.0202759	-1.15	0.250166	
categoryMathematics / Computer Sciences	-0.3302827	0.023056	-14.325	< 0.0000000000000002	***
categoryOther	-0.1401558	0.0472589	-2.966	0.00302	**
categoryPhysical Sciences / Engineering	-0.1328756	0.0158601	-8.378	< 0.0000000000000002	***
categorySocial sciences and humanities	-0.1185448	0.0309961	-3.825	0.000131	***
if_2016	0.3632203	0.0047368	76.68	< 0.0000000000000002	***
the_score_first	0.0037723	0.0002681	14.071	< 0.0000000000000002	***

Effect

		OA			IF			THE		
		COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE
2	Cumulative citations	0.3102008	1.36369892	36%	0.3632203	1.43795261	44%	0.0037723	1.00377942	0.4%

Model 3: Score

glm.nb(score ~ oa + category + if_2016 + the_score_first, data = df)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.0837344	0.0642937	-32.41	< 0.0000000000000002	***
oaOA	1.2564226	0.0776833	16.174	< 0.0000000000000002	***
categoryClinical Medicine	0.2123026	0.0569979	3.725	0.000196	***
categoryHuman Sciences	1.4862259	0.0638709	23.269	< 0.0000000000000002	***
categoryLife Sciences	0.8865875	0.0614582	14.426	< 0.0000000000000002	***
categoryMathematics / Computer Sciences	-1.5909543	0.0861107	-18.476	< 0.0000000000000002	***
categoryOther	1.1454529	0.1395731	8.207	2.27E-16	***
categoryPhysical Sciences / Engineering	-0.2927449	0.0498716	-5.87	4.3586E-09	***
categorySocial sciences and humanities	0.3640148	0.0940007	3.872	0.000108	***
if_2016	0.4592631	0.0146307	31.39	< 0.0000000000000002	***
the_score_first	0.0167535	0.0008359	20.043	< 0.0000000000000002	***

Effect

		OA			IF			THE		
		COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE
3	Score	1.2564226	3.51283218	251%	0.4592631	1.58290711	58%	0.0167535	1.01689463	1.7%

Model 4: Overall news

glm.nb(count_news ~ oa + category + if_2016 + the_score_first, data = df)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-6.07322	0.24358	-24.933	< 0.0000000000000002	***
oaOA	1.1614	0.27393	4.24	2.2372E-05	***
categoryClinical Medicine	0.26154	0.20662	1.266	0.2056	
categoryHuman Sciences	1.80266	0.22483	8.018	1.08E-15	***
categoryLife Sciences	1.26523	0.21897	5.778	7.5577E-09	***
categoryMathematics / Computer Sciences	-3.72674	0.76404	-4.878	1.0736E-06	***
categoryOther	1.12118	0.48114	2.33	0.0198	*
categoryPhysical Sciences / Engineering	-0.11788	0.18573	-0.635	0.5256	
categorySocial sciences and humanities	0.08476	0.35941	0.236	0.8136	
if_2016	0.49356	0.05187	9.516	< 0.0000000000000002	***
the_score_first	0.02977	0.00308	9.665	< 0.0000000000000002	***

Effect

		OA			IF			THE		
		COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE
4	Overall news	1.1614	3.19440231	219%	0.49356	1.63813762	64%	0.02977	1.03021756	3.0%

Model 5: Count_policy

glm.nb(count_policy ~ oa + category + if_2016 + the_score_first, data = df)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-7.269914	0.310371	-23.423	< 0.0000000000000002	***
oaOA	0.977942	0.260956	3.748	0.000179	***
categoryClinical Medicine	0.587612	0.261599	2.246	0.024689	*
categoryHuman Sciences	1.710936	0.255489	6.697	2.1318E-11	***
categoryLife Sciences	0.672615	0.292269	2.301	0.021372	*
categoryMathematics / Computer Sciences	-0.424023	0.433289	-0.979	0.32777	
categoryOther	1.784854	0.422283	4.227	2.3717E-05	***
categoryPhysical Sciences / Engineering	-0.010467	0.262559	-0.04	0.9682	
categorySocial sciences and humanities	1.392104	0.351913	3.956	7.6273E-05	***
if_2016	0.213717	0.049177	4.346	1.3871E-05	***
the_score_first	0.026507	0.003527	7.516	5.65E-14	***

Effect

		OA			IF			THE		
		COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE
5	Count policy	0.977942	2.65897843	166%	0.213717	1.23827217	24%	0.026507	1.02686144	2.7%

UK study

In the following models, df refers to the dataframe with all articles (OA=0 for non-OA, OA=1 for OA). The models were programmed in R (glm.nb of NB GLM or glmer.nb for NB GLMM). With the glmer.nb models, the dataframe df_gather was used. This is the same data but with a value for each year.

The following variables were used:

- oa: 1 if oa, 0 if non-oa
- impact_factor_2014: impact factor 2014
- category: journal subject
- downloads_by_month: monthly downloads
- times_cited_year1_2: cumulative citation for the 2 years following publication
- score_history_1y: Altmetric score overall, recorded at 1 year after online publication (provided by Altmetric)

Model 6: Monthly downloads

glmer.nb(downloads_by_month ~ oa*days_from_pub + category + impact_factor_2014 + (1 | art_no), data = df_gather, control = glmerControl(calc.derivs = FALSE), nAGQ=0)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	2.242760397	0.031487166	71.228	< 0.0000000000000002	***
oaOA	1.955958753	0.017881381	109.385	< 0.0000000000000002	***
days_from_pub	-0.000658652	0.000005309	-124.053	< 0.0000000000000002	
categoryClinical Medicine	0.009191393	0.031423877	0.292	0.77	***
categoryHuman Sciences	0.472397088	0.030647511	15.414	< 0.0000000000000002	***
categoryLife Sciences	0.054397819	0.035896119	1.515	0.13	***
categoryMathematics / Computer Sciences	-0.314047945	0.035209935	-8.919	< 0.0000000000000002	*
categoryOther	0.26880598	0.044849853	5.993	2.05418E-09	
categoryPhysical Sciences / Engineering	-0.246508739	0.027916717	-8.83	< 0.0000000000000002	
categorySocial Sciences and Humanities	0.339340244	0.041966578	8.086	6.17E-16	***
impact_factor_2014	0.187925384	0.007192177	26.129	< 0.0000000000000002	***
oaOA:days_from_pub	-0.001187376	0.00001582	-75.056	< 0.0000000000000002	

Effect

		OA			IF		
		COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE
6	Monthly downloads	1.95595875	7.07069482	607%	0.18792538	1.20674347	21%

Model 7: 2-year citation

glm.nb(times_cited_year1_2 ~ oa + impact_factor_2014 + category, data = df)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.35902	0.04503	7.973	1.55E-15	***
oaOA	0.09553	0.0258	3.703	0.000213	***
impact_factor_2014	0.2505	0.00973	25.745	< 0.0000000000000002	***
categoryClinical Medicine	-0.04122	0.04435	-0.93	0.352612	
categoryHuman Sciences	-0.19265	0.04664	-4.131	3.61567E-05	***
categoryLife Sciences	-0.14607	0.05346	-2.732	0.006292	**
categoryMathematics / Computer Sciences	-0.32484	0.05689	-5.71	1.13192E-08	***
categoryOther	0.03446	0.06232	0.553	0.580302	
categoryPhysical Sciences / Engineering	-0.13228	0.04096	-3.23	0.00124	**
categorySocial Sciences and Humanities	-0.22946	0.07188	-3.192	0.001411	**

Effect

		OA			IF		
		COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE
7	2-year citation	0.26448	1.30275337	30%	0.2937	1.34138143	34%

Model 8: Altmetric: score history 1 year

glm.nb(score_history_1y ~ oa + impact_factor_2014 + category, data = df)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.31617	0.21968	-10.543	< 0.0000000000000002	***
oaOA	1.69384	0.11784	14.374	< 0.0000000000000002	***
impact_factor_2014	0.38263	0.04949	7.732	1.06E-14	***
categoryClinical Medicine	-0.04181	0.21856	-0.191	0.84828	
categoryHuman Sciences	1.77486	0.21057	8.429	< 0.0000000000000002	***
categoryLife Sciences	0.40944	0.24884	1.645	0.09988	.
categoryMathematics / Computer Sciences	0.693	0.24505	2.828	0.00468	**
categoryOther	0.90117	0.30497	2.955	0.00313	**
categoryPhysical Sciences / Engineering	-0.86797	0.20002	-4.339	1.4289E-05	***
categorySocial Sciences and Humanities	-0.06867	0.30258	-0.227	0.82047	
impact_factor_2014	0.187925384	0.007192177	26.129	< 0.0000000000000002	***
oaOA:days_from_pub	-0.001187376	0.00001582	-75.056	< 0.0000000000000002	

Effect

		OA			IF		
		COEFF	EXP(COEFF)	INCREASE	COEFF	EXP(COEFF)	INCREASE
8	Altmetric: score history 1 year	1.69384	5.44033152	444%	0.38263	1.46613546	47%

References

1. Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, 31 May 2001.
2. Hitchcock, S. (2004). The effect of open access and downloads ('hits') on citation impact: a bibliography of studies. http://opcit.eprints.org/oacitation-biblio.html
3. https://sparceurope.org/what-we-do/open-access/sparc-europe-open-access-resources/open-access-citation-advantage-service-oaca/oaca-list/
4. Schwarz, G. J., & Kennicutt Jr, R. C. (2004). Demographic and Citation Trends in Astrophysical Journal papers and Preprints. ArXiv:Astro-Ph/0411275. Retrieved from http://arxiv.org/abs/astro-ph/0411275
5. Moed, H. F & Henk, F. (2007). The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section. *Journal of the American Society for Information Science and Technology*, 58(13), 2047–2054. https://doi.org/10.1002/asi.20663
6. Kousha, K., & Abdoli, M. (2010). The citation impact of Open Access agricultural research: A comparison between OA and non-OA publications. *Online Information Review*, 34(5), 772–785. https://doi.org/10.1108/14684521011084618
7. https://www.nature.com/press_releases/ncomms-report2014.pdf
8. Hajjem, C., Harnad, S., & Gingras, Y. (2006). Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. ArXiv:Cs/0606079. Retrieved from http://arxiv.org/abs/cs/0606079

See also Davis, P. M. (2011), Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *The FASEB Journal* 25(7), 2129-2134.

9. Kamat, P. (2018). Impact of Open Access Papers in Hybrid Journals. *ACS Energy Letters*. 3 (2), 410-411. DOI: 10.1021/acsenergylett.8b00061
10. Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *PLoS ONE*, 5(10). https://doi.org/10.1371/journal.pone.0013636
11. Sotudeh, H. & Estakhr, Z. (2018). Sustainability of open access citation advantage: the case of Elsevier's author-pays hybrid open access journals. *Scientometrics.* 115(1), 563-576. https://doi.org/10.1007/s11192-018-2663-4
12. Mueller-Langer, F. & Watt, R. (2018). How many more cites is a $3,000 open access fee buying you? empirical evidence from a natural experiment. *Economic Inquiry*, 56(2), 931–954. https://doi.org/10.1111/ecin.12545
13. Science-Metrix (2018), Analytical Support for Bibliometrics Indicators Open access availability of scientific publications. Science-Metrix. http://science-metrix.com/sites/default/files/science-metrix/publications/science-metrix_open_access_availability_scientific_publications_report.pdf
14. Gargouri, Y. *et al*. (2010). See #10
15. Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics*, 103(2), 555–564. https://doi.org/10.1007/s11192-015-1547-0
16. Bornmann, L., Haunschild, R., & Marx, W. (2016). Policy documents as sources for measuring societal impact: how often is climate change research mentioned in policy-related documents? *Scientometrics*, 109(3), 1477–1495. https://doi.org/10.1007/s11192-016-2115-y
17. Tattersall, A., & Carroll, C. (2018). What Can Altmetric.com Tell Us About Policy Citations of Research? An Analysis of Altmetric.com Data for Research Articles from the University of Sheffield. *Frontiers in Research Metrics and Analytics*, 2. https://doi.org/10.3389/frma.2017.00009
18. Björk. B. (2017). Growth of hybrid open access, 2009–2016. *PeerJ* 5, e3878. https://doi.org/10.7717/peerj.3878
19. Calder *et al*. (2018). Gold Open Access in the UK: Springer Nature's transition. Springer Nature. https://10.6084/m9.figshare.6230813
20. Van Den Eynden, V., Knight, G., Vlad, A., Radler, B., Tenopir, C., Leon, D., Manista, F., Whitworth, J., Corti, L., (2016). Survey of Wellcome researchers and their attitudes to open research. Technical Report. UNSPECIFIED, figshare. https://doi.org/10.6084/m9.figshare.4055448.v1
21. https://figshare.com/articles/Author_Insights_2015_survey/1425362
22. Van Noorden, R. (2013). Open access: The true cost of science publishing. *Nature News*, 495(7442). 426-429. https://doi.org/10.1038/495426a

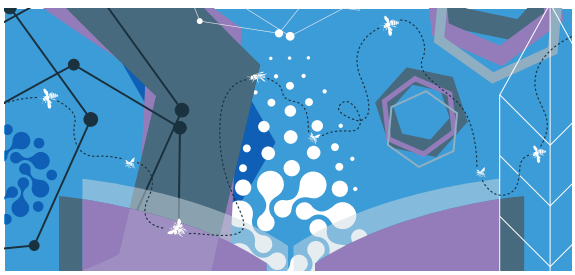
See also https://scholarlykitchen.sspnet.org/2008/08/19/oa-advantage-deminishin/

23. Calder et. al (2018). See #19
24. https://www.springer.com/gb/open-access/springer-open-choice/springer-compact/for-uk-authors-intro/731990
25. Davis, P. M., Lewenstein, B. V., Simon, D. H., Booth, J. G., & Connolly, M. J. L. (2008). Open access publishing, article downloads, and citations: randomised controlled trial. *BMJ*, 337, a568. https://doi.org/10.1136/bmj.a568.

See also Eysenbach, G. (2006). Citation Advantage of Open Access Articles. *PLOS Biology*, 4(5), e157. https://doi.org/10.1371/journal.pbio.0040157

26. Mueller-Langer, F. & Watt, R. (2018). See #12
27. Yang, S., Harlow, L., Puggioni, G., & Redding, C. (2017). A Comparison of Different Methods of Zero-Inflated Data Analysis and an Application in Health Surveys. *Journal of Modern Applied Statistical Methods*, 16(1). https://doi.org/10.22237/jmasm/1493598600
28. Lawrence, S. (2001). See #1
29. Wang, X. *et al*. (2015). See #15
30. Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., Haustein, S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6, e4375. https://doi.org/10.7717/peerj.4375

Around our complex and interconnected world, the research community is advancing discovery for all of us. These illustrations celebrate some of the great minds who have helped advance discovery through history.



Jean-Claude Bradley (1969–2014)

Jean-Claude Bradley was a chemist and passionate proponent of Open Science. Following an early career in patent driven nanotechnology, Bradley came to believe that the work he was doing wasn't having the impact or benefitting mankind in the way he had hoped. At Drexel University, working on antimalarials, he coined the term Open Notebook Science for an approach which aimed to make the details and raw scientific data of every experiment done in the lab freely available within hours of production. Bradley was founding Editor-in-Chief of *Chemistry Central Journal* and a founding Editor of the *Journal of Cheminformatics*. In 2007 he was awarded a Blue Obelisk award for achievements in promoting Open Data, Open Source and Open Standards.