Open access mythbusting: Testing two prevailing assumptions about the effects of open access adoption

Dan Pollock 1*, and Ann Michael 2

1 Silicon Guru Ltd., London, UK
2 Delta Think Inc., Philadelphia, PA, USA

ORCID:
D. Pollock: 0000-0002-0881-7758
A. Michael: 0000-0003-3332-1970

*Corresponding author: Dan Pollock
E-mail: dan@silicon-guru.com

Abstract
This article looks at whether there is evidence to support two prevailing assumptions about open access (OA). These assumptions are: (1) fully OA journals are inherently of poorer quality than journals supported by other business models and (2) the OA business model, that is, paying for publication, is more ‘competitive’ than the subscription journal access business model. The assumptions have been discussed in contemporary industry venues, and we have encountered them in the course of their work advising scholarly communications organizations. Our objective was to apply data analytics techniques to see if these assumptions bore scrutiny. By combining citation-based impact scores with data from publishers’ price lists, we were able to look for relationships between business model, price, and ‘quality’ across several thousands of journals. We found no evidence suggesting that OA journals suffer significant quality issues compared with non-OA journals. Furthermore, authors do not appear to ‘shop around’ based on OA price.

INTRODUCTION
Open access (OA) has seen a steady rise since the late 1990s (Björk & Solomon, 2014; Johnson, Watkinson, & Mabe, 2018). It now accounts for around $570m of the $10bn journal publishing market (Pollock & Michael, 2018). It represents a change in the way that scholarly publishing is practiced and monetized, and discussion is ongoing as to the effectiveness and benefits of these changes (see Crotty, 2017; Laakso & Björk, 2012).

This paper focuses on selected key market dynamics that follow from the increasing adoption of OA. In particular, by cross-referencing pricing data with measures of impact and analysing their correlation over time, it tests suppositions about the effects of adopting commercial OA models. It focuses on common concerns that the authors have heard raised by scholarly communications organizations in the course of their work as independent advisors.

As publishers decide how best to respond to changes, they often work with external advisors to provide outside perspectives and ideas. The data and analysis in this paper have been gathered and conducted by one such advisory firm, Delta Think (see http://oainfo.deltathink.com). As independent advisors, we speak with many organizations while maintaining a neutral perspective.

We have chosen to examine two assumptions about open access (OA) that we commonly encounter. For the purposes of discussion, we have dubbed them as ‘myths’ as it is our intent to provide readers with additional insight that will allow them to evaluate the accuracy of each.

Myth 1: OA journals are inherently, or more inclined to be, of lower quality than subscription journals.

Myth 2: OA is a more competitive model than subscription publishing models.

We will examine each of these ‘myths’ in turn, outlining a method we used to test it and presenting the results. As we are testing contemporary discussions, our references draw from the formal literature and from online discussions, such as industry news and blog posts.
Key points

- There is no evidence that, in the mainstream literature, open access (OA) journals suffer significant quality issues compared with non-OA journals.
- An increasing number of fully OA publications are attaining higher Journal Impact Factors at faster rates than their subscription and hybrid counterparts.
- Researchers do not appear to shop around for the best-value article processing charges (APCs).
- APCs for fully OA journals are slightly more price-sensitive than for hybrid journals but still show only a weak relationship between APC and impact.
- When numbers of papers published are taken into account, megajournals influence the fully OA market and show a mild price sensitivity when included; if their influence is excluded, price sensitivity remains very low.

METHODOLOGY

In general, our approach was to cross-reference several lists of journal metadata, so we could then look for correlations between them. We used impact-based measures (as a proxy for ‘quality’) and publishers’ public pricing information. A custom MySQL database was used to collate, tidy, and cross-reference the lists based on journal ISSNs. Visualizations were created using Zoho Reports, which allowed further custom SQL work to calculate correlation metrics at scale across the tidied data. The sources and nuances specific to each myth are discussed in detail in each myth’s section below.

Myth 1: OA is low quality

Are OA journals inherently of lower quality than subscription journals? The notion here is that commercial imperatives will undermine editorial best practice: papers will be chosen because they make money rather than because they represent quality research. There has been ongoing debate on this. For example, see Shieber (2009) and Davis (2009). Science’s famous sting (Bohannon, 2013), of course, did little to allay concerns (Esposito, 2013). As (OA advocate) Peter Suber commented at that time, ‘Even if the study was biased... and some OA publishers fared well, the perception has stuck’ (Suber, 2013).

Myth 1 method

To test this assumption, we first need a measure of quality. The obvious choice is the Journal Impact Factor (JIF) as we know that authors attach high importance to it as a measure of prestige. For example, in its Author Insights survey in 2015, Nature Publishing Group (now part of Springer Nature) found that the reputation of the journal is ‘very important’ or ‘important’ for 97% of respondents, along with other factors including relevance to the discipline (96%), quality of the peer review (92%), and JIF (90%). Having an OA option ranked fourth from last (36%) (21,377 authors were surveyed; see Nature Publishing Group, 2015).

While we recognize that the JIF is not a measure of absolute ‘quality’ (Clarivate Analytics, 1994), its use in this regard is widespread and consistent (see Swan & Brown, 1999; Van den Eynden et al., 2016). We therefore think it reasonable to examine how JIFs relate to journals’ OA status as a proxy to measure their quality.

Using Clarivate’s Journal Citation Reports, it is possible to filter fully OA journals (there is currently no possible OA in hybrid filtering). We were therefore able to collate average JIFs to determine how they relate to journal type: fully OA, not OA (i.e. not fully OA), and across the whole index. By collating averages per year, we could see if things changed over time.

Myth 1 results

Figures 1 and 2 visualize the results of the analysis.

Figure 1 suggests that JIFs for fully OA journals track just below the average for all journals. So, at first glance, this would suggest that concerns about the ‘lower quality’ of fully OA journals are not unfounded. However, there are two important caveats.

First, although they have been significantly below the average in the past, fully OA journals are clearly catching up. Their average JIF in 2017 was 2.28 compared with 2.33 for the index as a whole, or a 2.1% difference. This compares with a 12.6% difference in 2009 (1.6 fully OA vs. 1.83 for the index). Regression analysis of the data suggests that, in 2018, averages will be almost the same (2.47 vs. 2.48) and that fully OA journals could exceed in the average in 2019 if current trends continue.

Second, the crude averages mask important nuances in dynamics. Figure 2 examines how fully OA journals perform at the top and bottom ends of the index. It teases out whether fully OA journals are more or less likely to be high performers than the average. The grey line shows the number of fully OA journals as a percentage of all journals in the index. This is our baseline. The other colours show what happens if we focus on just journals with the stated measures of impact. For example, the orange line shows the number of fully OA journals with a JIF of 4 or more as a percentage of ALL journals with a JIF of 4 or more. So, by comparing with our grey ‘control’ line, we can see if journals of particular levels of impact account for proportionately more or less of the index than the average.

Myth 1 discussion

The data suggest that there is no significant gap between impact factors of fully OA journals and the average. The proportion of fully OA journals in the index is growing over time, but the proportion of higher-performing fully OA journals is growing even faster. In other words, the high-performing fully OA journals are taking an increasing share of the index.
We know, of course, that JIFs vary by field, and many consider them to be flawed metrics (see Chawla, 2018 for example). However, they are widely used metrics, with large commercial publishers and societies well-represented. We therefore think that analysing JIFs provides a reasonable representation of the mainstream market, and doing so in aggregate provides a reasonable proxy for testing generalized perceptions. In this context, the data simply do not support a notion of compromised quality in fully OA journals. Indeed, if current trends continue, fully OA journals may even begin to outperform the average.

Determining the causes of this observation is difficult. Many studies (see SPARC Europe, 2016) claim the existence of an Open Access Citation Advantage, which would increase citation-based metrics, and therefore ‘quality’, simply by virtue of increased traffic. (There are arguments against this of course; see Davis, 2014.) Alternatively, we could be seeing evidence of fully OA journals maturing, with their general quality catching up to the average and therefore attracting more citations. In addition, given that fully OA journals are on average likely to be younger titles, and that JIFs for new titles are hard to receive, we would expect fully OA journals’ JIFs to take time to become fully established.
Myth 2: APCs show a competitive model

Can we assume that article processing charge (APC)-based OA is more competitive than the traditional subscription business model? The notion here is that APCs operate in a buyer-driven market, within which authors can choose between genuine substitutes (e.g. see Björk & Solomon, 2014; Suber, 2013). This will lead to price sensitivity (‘shopping around’) compared with the supplier-driven subscriber models, which offer complementary products and less choice. (See also Samberg, Schneider, Taylor, & Wolfe, 2018 for a recent example of this proposition.)

Myth 2 method

To test this assumption, we first assume that (as per Shieber, 2009) subscription journals are inherently complementary and so buyers cannot shop around. So, our question becomes whether the author APC model is different to this and whether we see evidence of authors substituting journals based on price. We therefore looked for a correlation between price (list APC) and ‘quality’ of publication.

For price, we gathered information on 13,571 journals from the 21 largest publishers by volume of output. We gathered prices in December 2017 and January 2018 from the publishers’ websites. Different publishers change their prices at different times of year, so we took our pricing data nominally to cover the 2017/2018 pricing year. We used list prices for CC BY licences (or equivalent) to compare like-for-like across publishers and control any tiered pricing variations. (Some publishers discount APCs for less permissive licences, such as the restriction of commercial use or derivative products.)

For ‘quality’, we know that authors value JIFs highly when assessing journals (as discussed above). However, JIFs should only be used within specific subject areas (Clarivate Analytics, 1994). For a market-wide comparison at a per-journal level, we used the Source Normalized Impact per Paper (SNIP) instead as a proxy for quality. Produced by CWTS (http://www.journalindicators.com/), and derived from Scopus data, the SNIP accounts for differences in citation patterns across disciplines. We used SNIP data for the 2016 publication year, covering 23,160 journals. This data were made available in July 2017. We compared it with our 2017/2018 prices on the assumption that publishers considering APC pricing levels would set them for the coming year based on the most recently available data about impact.

We cross-referenced the price list and SNIP data sets by matching ISSNIs and using linking The ISSN International Centre (2018; http://www.issn.org/understanding-the-issn/assignment-rules/the-issn-l-for-publications-on-multiple-media/) to maximize the hits. We used a custom bibliographic mySQL database to tidy the data and run the cross-referencing. We matched 9,896 journals with both price list and SNIP information that we could identify as either fully OA or hybrid journals based on publishers’ own information.

Myth 2 results

The summary of the results is visualized as a scatter plot in Fig. 3, and details are shown in Table 1.

Overall, the shape of the scatter plot chart suggests little correlation between APC and SNIP. The dots do not even loosely cluster around a line. The presence of entries at the bottom right suggests that higher-than-average prices are being asked for journals of relatively modest impact. There are some vertical rows of...
The data suggest low relationship between price and measures of impact. The single biggest predictor of price appears to be the journal business model: fully OA journals cluster around the cheaper end of the spectrum, hybrid the more expensive. There is evidence of some shopping around of fully OA journals but only when the influence of megajournals is taken into account, and even then, it is mild. In other words, megajournals show a greater price sensitivity than average, and the volume of papers within them is affecting figures across the whole market.

Myth 2 discussion

The data suggest low relationship between price and measures of impact. The single biggest predictor of price appears to be the journal business model: fully OA journals cluster around the cheaper end of the spectrum, hybrid the more expensive. There is evidence of some shopping around of fully OA journals but only when the influence of megajournals is taken into account, and even then, it is mild. In other words, megajournals show a greater price sensitivity than average, and the volume of papers within them is affecting figures across the whole market.

The cluster of data points at around $5,000 and above are largely coming from organizations with premium brands and/or premium journals (such as the American Chemical Society or Nature Communications). We should be clear that we are not suggesting that these are ‘over-priced’ or passing any value judgement. As we discussed above, we know that qualitative perceptions of quality influence authors’ decisions, so it is entirely logical that publishers may price journals that meet such perceptions towards the top end of the market.

Comparison with Björk and Solomon’s (2015) study suggests that the correlation between price and impact metrics have reduced since 2011. The caveat here is that we did not try to recreate their exact method, and the data sets differ between the two studies.

However, whatever qualitative factors authors may choose, we know that impact measures are important, and our data cover around 50% of scholarly output. So, we think that it is fair to conclude that authors do not shop around, and that market forces – even for fully OA journals – are not taking hold.

CONCLUSIONS

Fully OA journals are operating comparably with the averages in the all-important JIFs. If current trends continue, they may even perform above average by 2019. The data suggest that an increasing number of fully OA publications are attaining higher impact factors at faster rates than their subscription and hybrid counterparts.

Authors do not appear to be shopping around based on APC levels. Some of the previous studies suggested that this should hold true for fully OA journals in particular. However, although we see a slight strengthening of the relationship between price and impact in this case, the relationship remains weak.

The data set we have analysed (http://oainfo.deltathink.com) offers more possibilities than presented here, and perhaps these
can be explored in future papers. Examining patterns in measures other than the JIF is another way of looking at the quality myth. (Although we have explored SNIP and the SCImago Journal Ranking and found the patterns to be consistent with our findings here – see Michael & Pollock, 2018.) We could further analyse the price versus SNIP data by subject or by publisher (see Pollock, 2018 for some analysis about the latter) to see if there are any differences in shopping around by discipline.

Our two ‘myths’ were chosen to address a positive and a negative perception of OA. Whatever one’s position, neither appears to be supported by our data.

REFERENCES

Björk, B. C., & Solomon, D. J. (2015). Article processing charges in OA journals: Relationship between price and quality. Scientometrics, 103, 373–385. http://doi.org/10.1007/s11192-015-1556-z

Björk, B. C. & Solomon, D. J. (2014). Developing an effective market for open access article processing charges. Retrieved from https://wellcome.ac.uk/sites/default/files/developing-effective-market-for-open-access-article-processing-charges-mar14.pdf

Bohannon, J. (2013). Who’s afraid of peer review? Science, 342(6154), 60–65. https://doi.org/10.1126/science.342.6154.60

Chawla, D. S. (2018, 3 April). What’s wrong with the JIF in 5 graphs. Nature Index [Web log post]. Retrieved from http://www.natureindex.com/news-blog/whats-wrong-with-the-jif-in-five-graphs

Clarivate Analytics. (1994, 20 June). The Clarivate Analytics Impact Factor. Originally published in the Current Contents print editions June 20, 1994, when Clarivate Analytics was known as The Institute for Scientific Information® (ISI®) [Web log post]. Retrieved from http://clarivate.com/essays/impact-factor/

Crotty, D. (2017, 23 October). Open access: A look back [Web log post]. Retrieved from http://scholarlykitchen.sspnet.org/2018/10/22/open-access-a-look-back/

Davis, P. (2009, 23 October). Open access and vanity publishing [Web log post]. Retrieved from http://scholarlykitchen.sspnet.org/2009/10/21/open-access-and-vanity-publishing/

Davis, P. (2014, 21 October). Is open access a cause or an effect? [Web log post]. Retrieved from http://scholarlykitchen.sspnet.org/2009/10/21/open-access-and-vanity-publishing/

Esposito, J. (2013, 23 October). The natural limits of gold open access [Web log post]. Retrieved from http://scholarlykitchen.sspnet.org/2013/11/19/the-natural-limits-of-gold-open-access/

Johnson, R., Watkinson, A., & Mabe, M. (2018, 23 October). The STM report: An overview of scientific and scholarly publishing’. Retrieved from http://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf

Laakso, M., & Björk, B.-C. (2012). Anatomy of open access publishing – A study of longitudinal development and internal structure. BMC Medicine, 10, 124. https://doi.org/10.1186/1741-7015-10-124

Michael, A. & Pollock, D. (2018, 23 October). Evaluating quality in open access journals [Web log post]. Retrieved from http://deltathink.com/news-views-evaluating-quality-in-open-access-journals/

Nature Publishing Group. (2015, 23 September). Author Insights 2015 survey [Web log post]. Retrieved from https://doi.org/10.6084/m9.figshare.1425362.v7

Pollock, D. (2018, 23 October) Open access charges [Web log post]. Retrieved from http://deltathink.com/news-views-open-access-article-processing-charges/

Pollock, D. & Michael, A. (2018, 23 October). Open access market sizing update 2018 [Web log post]. Retrieved from http://deltathink.com/news-views-open-access-market-sizing-update-2018/

Samberg, R., Schneider, R. A., Taylor, A., & Wolfe, M. (2018, 23 October). It’s time for the US to get serious about funding open access [Web log post]. Retrieved from http://www.natureindex.com/news-blog/its-time-for-the-us-to-get-serious-about-funding-open-access

Shieber, S. (2009, 23 October 2018). The occasional pamphlet from the virtual writing-desk of Stuart Shieber, on scholarly communication and other topics [Web log post]. Retrieved from http://blogs.harvard.edu/pamphlet/2009/10/16/is-open-access-publishing-a-vanity-publishing-industry/

SPARC Europe. (2016, 23 October). The open access citation advantage service (OACA) [Web log post]. Retrieved from http://sparc europe.org/what-we-do/open-access/sparc-europe-open-access-resources/open-access-citation-advantage-service-oaca/

Suber, P. (2013, 4 October). New "sting" of weak open-access journals [Web log post]. Retrieved from http://plus.google.com/+PeterSuber/posts/CRHeCATQqGq

Swan, A., & Brown, S. (1999). What authors want, the ALPSP research study on the motivations and concerns of contributors to learned journals. Learned Publishing, 12, 170–172. https://doi.org/10.1087/09531519950145742

The ISSN International Centre. (2018, 23 October) The ISSN-L for publications on multiple media [Web log post]. Retrieved from http://www.issn.org/understanding-the-issn/assignment-rules/the-issn-l-for-publications-on-multiple-media/

Van den Eynden, V., Knight, G., Vlad, A., Radler, B., Tenopir, C., Leon, D., ... Corti, L. (2016). Towards open research: Practices, experiences, barriers and opportunities [Web log post]. Wellcome Trust. https://doi.org/10.6084/m9.figshare.4055448