

Are open access fees a good use of tax payers money?

Graham Kendall, University of Nottingham (UK and Malaysia)

EMAIL: Graham.Kendall@nottingham.edu.my

ORCID: 0000-0003-2006-5103

February 14, 2024

1 Introduction

[Butler et al. \(2023\)](#) reported that five large commercial publishers (Elsevier, Sage, Springer-Nature, Taylor & Francis and Wiley) received \$1.06 billion in publication fees between 2015 and 2018. Another three publishers were mentioned (Frontiers, MDPI and PLOS), but were not analyzed. The revenue was underestimated and the number of open access (OA) articles increased over the period of study. The five publishers analyzed charged (on average) \$1,989 for gold OA articles and \$2,905 for hybrid articles.

In this letter eight publishers are analyzed, over a longer period; nine years (2015-2023). The publishers are the same as those previously analyzed, and also the three that were mentioned but not analyzed. Three of the publishers' revenue is estimated as the data is too large to collect for this short letter, but a project which includes these publishers (and others) is currently underway. The other five publishers have their open access revenue estimated using article level data, article processing charges (APC) and/or the publishers' annual reports.

This letter provides a revenue estimate for eight publishers over a nine year period (2015-2023). Given that the data is difficult to collect, and then it can only be used to give a rough indication of revenue, it is argued that the publishers themselves should provide information, for each article, so that the funders of open access articles (typically the tax payers) can not only access the research it funds (which is the primary motivation for open access) but also know how their funds are supporting open access charges, in much more detail than is available at the present time.

2 Methodology

The data collection process for the eight publishers is given below. The three publishers marked with an asterisk are too large to collect data for this letter, but a larger project is currently being undertaken that will include these, and additional, publishers. For these three publishers, estimates of their revenue (drawing on the figures in [Butler et al. \(2023\)](#)) is given so that an overall figure, for the eight publishers, can be derived.

For four of the publishers, how many papers are published by each of its journals, for each year is captured. This is not an easy task as a bespoke capture plan (using a browser web scraping extension) had to be devised for each publisher. The current APC for each journal is also captured and this is discounted by 5% each year to provide an APC for previous years. The figure of 5% was chosen as random, manual sampling suggests that this is not unreasonable. It would have been better to have captured the APC for each journal, for each of the nine years, but this is challenging to do and the data is unlikely to be available for every journal for all of the nine years, so an estimated figure would still be required.

Once we have the number of articles published by a given journal in a given year, and an APC for each year, it is a simple matter to derive the yearly income for that journal and then sum up each journal to give an annual income for the publisher. We note that waivers are often mentioned on the

web sites of open access journals, but there is no way of knowing if a waiver has been awarded and, if so, how much, so this data could not be collected.

In two cases (PLOS and Wiley), the publishers' annual reports are available. As the article level data is also captured for PLOS, the revenue figure can be compared against the figures from the annual reports. In the case of Wiley, we rely solely on the annual reports as they publish 1,600 journals and have published over eight million articles and collecting that amount of data is outside the scope of this letter.

1. **Elsevier** have an open access FAQ page^[1] from which you can download a list of journals, which indicates whether they are hybrid, open access or traditional, along with the current APCs (as at 23 January 2024). Of the 2,702 journals listed, 834 are fully open access. The number of articles published in these 834 journals between 2015-2023 is 731,054.
2. **Frontiers** is a fully open access publisher, so all of their journals are fully open access. Data was collected from their 228 journals. The current APC's (as at 30 January 2024) were collected for each journal. The number of articles for each year (2015-2023) was also captured. The total number of articles is 459,743.
3. * **MDPI** is a fully open access publisher, publishing about 430 journals. Between 1996 and 2022, they published one million articles^[2]. It is worth noting that if the average APC was \$USD 1,000 (which is almost certainly an underestimate) MDPI's total revenue would be \$USD 1 billion. Using the average APC (\$USD 1,989) from [Butler et al. \(2023\)](#), this would put the total revenue closer to \$USD 2 billion. A conservative estimate of \$USD 500,000 is used as their revenue for 2015-2023.
4. **PLOS** is a fully open access publisher. Data was collected on their 14 journals, although two are new and have not published any articles. The current APC's (as at 29 January 2024)^[3] were collected for each journal, as well as the number of articles (2015-2023). The total number of articles is 207,935.

PLOS provides an annual overview of its accounts^[4], which shows their revenue from open access charges (see Table 3).
5. **Sage** has 211 fully open access journals^[5]. One of the Sage web pages says^[6] "*For pure gold open access journals, APCs vary from journal to journal, so please visit the individual journal's homepage for details*". Therefore, the current APC for each journal was manually collected. Several journals do not charge APCs, as these are supported by other means (e.g. a university or a trade association). For other journals it was not possible to find an APC, so it was set to zero. Of the 211 journals, listed as being fully open access, 61 had their APCs set to zero. The number of articles published in these 211 journals, between 2015-2023, is 121,742.
6. * **Springer-Nature** covers a number of different imprints^[7,8] (BMC, Nature Portfolio, Springer and Palgrave Macmillan). They offer 683 open access journals and more than 2,200 hybrid journals. They have published more than 124,000 fully open access articles^[7]. The data is too large to collect for this letter. The estimated revenue we use for this publisher is \$USD 1,179,348,760. This is double that of the estimate given in Butler et al. This will be an underestimate as i) the figure in Butler et al. was underestimated (and was their lower estimate), ii) their estimate was over a four year period (2015-2018) and this letter includes an additional five years (2019-2023) and iii) the figures provided here assume no increase in the APC's, which is almost certainly not the case.
7. * **Taylor & Francis** have published over five million articles and publish 364 open access journals. The data is too large to collect for this letter. The estimated revenue we use for this publisher is \$USD 153,531,114. The reasons for this are the same as given for Springer-Nature.
8. **Wiley** have published eight million articles, across 1,600 journals^[9]. Accessing their annual reports^[10] provides financial information on open access revenue. The 2015-2023 annual reports indicate that open access revenue was \$USD 505,235 million.

#	Publisher	Butler	Kendall
1	Elsevier	221,441,616	293,832,673
2	Sage	31,576,202	49,305,233
3	Wiley	141,316,332	120,717,000
	TOTAL	394,334,150	463,854,906

Table 1: Estimated revenue of three publishers (2015-2018). Figures are in \$USD

#	Publisher	Kendall
1	Elsevier	1,384,0987,843
2	Frontiers	1,211,017,631
3	PLOS	327,526,349
4	Sage	159,511,574
5	Wiley Limited	505,236,075
	SUB TOTAL	3,587,389,472
6	MDPI	500,000
7	Springer Nature	1,179,348,760
8	Taylor & Francis	153,531,114
	TOTAL	5,420,269,346

Table 2: Estimated revenue of five publishers (2015-2023). Figures are in \$USD

3 Analysis

Table 1 shows the estimated revenue figures for three publishers, for 2015-2018. The table compares the estimated revenue derived here (the ‘Kendall’ column) and the figures presented by Butler et al. (the ‘Butler’ column). For Elsevier and Sage the Kendall revenues are higher than those of Butler et al. The estimated revenue for Wiley, for Kendall, is extracted from the annual reports and is lower than the Butler figure. Overall, the Kendall estimate is higher than the Butler et al. estimate. The main message though is that there is not a reliable way to accurately estimate the revenue income for a given publisher.

Table 2 shows all eight publishers, covering 2015-2023. The revenue estimate for publishers 1-5 (in Table 2) is \$USD3.587 billion. For the eight publishers, the estimated revenue is \$USD 5.420 billion. These figures will not be totally accurate, indeed, they may be quite far from the true values but it does provide a good indication of the scale of spend on open access revenues, across just eight publishers.

Table 3 shows a comparison between the “*Publication fees, nett*” as reported in the PLOS financial statements and the estimated figures using article level data that was collected from their web site. Between 2015 and 2017, the estimates are lower than the income reported by PLOS. For 2018-2021, the estimates are higher than the values reported by PLOS. Looking at 2015-2020, the difference in

Year	From Financial Overview	Kendall	% difference
2015	42,274,910	37,987,182	-10.68
2016	36,772,796	33,790,076	-8.45
2017	34,832,837	33,730,198	-3.22
2018	31,663,670	33,272,712	4.96
2019	29,847,728	33,384,452	8.15
2020	32,428,621	35,861,769	10.05
2021	32,402,344	39,054,764	18.62
TOTAL	240,222,906	246,081,152	2.41

Table 3: The PLOS revenue figures from their annual statements. Only up to 2021 was available at the time of writing. Figures are in \$USD

the values is 0.38%, which is negligible given the scale of the figures involved. The 2021 revenue figure differs by almost \$USD 6 million, a difference of 18.62%. This significant difference has yet to be explained, but it is positive to see the strong correlation across the other eight years (just 0.38% difference) and even an overall difference of 2.41% suggests that the article level calculation broadly agrees with the data in the financial statements, albeit compensating from one year to the next.

4 Discussion

Only eight publishers are included in the analysis in this letter. There are many other (legitimate) publishers that publish open access papers, which would significantly increase the overall spend on open access revenues.

There are other publishers which are more motivated by the financial side of publishing and they have little regard for maintaining the integrity of the scientific archive. These are often referred to as predatory publishers (Kendall, 2021, Kendall and Linacre, 2022, Macháček and Srholec, 2022). OMICS (the only proven predatory publisher), for example, has not been included. Estimating the open access charges from these publishers will significantly increase the reported revenues.

In this analysis only fully open access journals were considered. That is, those journals that offer both traditional and open access options were excluded. If these were included, it would significantly add to the number of articles. It would also significantly increase the data collection challenges not only as there would be significantly more papers to process, but to identify papers that were published under an open access model might be challenging.

The data is incomplete. As evidenced in Butler et al. (2023), as well as noted here, to collect the data is a vast undertaking, and even then many assumptions/estimations have to be made. The estimates will never be fully accurate unless the publishers are willing to provide article level data, which includes all necessary information (see Section 6).

5 Could open access fees provide better value if spent elsewhere?

The catalyst for the open access model of publishing was a requirement for government funded research to be available to those that pay for it, typically the tax payer. A, perhaps unintended, consequence of this is that the tax payer not only pays for the research to be carried out, but pays again to have that research published in a journal. Moreover, the information about how much has been spent on these publication costs is not easily available and those funds goes to commercial organizations who have an eye on the bottom line and have shareholders to consider.

Is this investment of billions of dollars providing value for money? To try and contextualize the amount that is spent on open access fees, we ask what other initiatives these funds could support?

Forbes^[11] provides details of USA Professorial salaries across a range of disciplines. They range between \$92,250 p.a. and \$133,950 p.a.. Taking into account additional employment costs, if it is assumed that the average professorial salary is \$USD 200,000, the \$USD 5,420,269,346 of open access fees reported here could fund 27,101 professors.

A UK government report^[12] says that the average student debt, at the completion of their course, is GBP 45,600 (\$USD 57,583 (using xe.com, 10 February 2024)). The \$USD 5,420,269,346 open access fees could remove this debt for 94,130 students. Another way of framing this is to consider the GBP 9,500 (\$USD 11,996) p.a. fees that are paid by UK students. Over a three year course they would pay GBP 28,500 (\$USD 35,983). The open access revenue reported here would fund 150,634 students.

6 Call to action

1. Elsevier and PLOS should be applauded for providing annual accounts, which includes data about open access fee revenue. We would encourage other publishers to do this.
2. The fee paid to publish each individual paper should be noted on the published article, as well as in the metadata. The actual amount paid and the amount of any waiver should be provided.

3. Who paid the open access fee should be stated so that it is easier to calculate (for example) the investment made by a specific university or country. This is even more important when the authors come from different institutions and/or countries.
4. Governments should reflect whether the money provided by their tax payers is providing value for money by paying open access fees, or whether the funds would be better spent on supporting students, supporting more research or even used in other government departments to support other urgent issues that the country faces.
5. Robust, peer reviewed journals should be established which are run as not for profit entities. The open access fees that are currently spent with commercial publishers should easily cover the costs of these journals. The journals should be under the control of the funding agencies and it should be a condition of accepting a grant that the results of that research are published in those journals. As long as the peer review processes are in line with the highly respected journals, these journals will quickly establish themselves as the leading journals in the discipline. This model would also be another way to combat the threat posed by predatory journals, as publicly funded research could not be published in those journals.

Of course there is also a need to ensure that the fees are proportionate to the service being provided and that publishers are not simply setting their fees as high as they believe they can get away with.

7 Limitations of this study

This is a letter, in response to the Butler et al. article. As such it is not a full research article. This is currently being prepared and will have much more detail than can be presented here. The limitations listed below will need to be considered for a more complete article.

1. Given the widely different ways that the data is presented by the journals/publishers it is challenging to collect the data, and be fully confident that it is robust and complete. Indeed, it is almost certain that any data collection exercise is lacking and new tools may need to be developed to provide datasets which are robust.
2. There is no one “go-to” place for every journal where the data can be collected from, other than the publisher itself. For example, indexes such as Scopus, Web of Science or the Directory of Open Access Journals (DOAJ) do not hold information on every journal.
3. Data is only collected at the journal level. It would be interesting to collect data about authors (noting that author disambiguation is an issue) and also countries, so that an estimate can be derived about how much a given country spends on open access fees.
4. Where there are multiple authors across different institutions/countries it is challenging to know who paid the open access fee and how its costs should be attributed.
5. Only fully open access journals have been considered here. Many open access articles are published in hybrid and/or transformative journals. Collection information about open access articles published in hybrid journals adds another layer of complexity to the data collection task.
6. There are models where institutions can enter an agreement with a publisher, enabling their researchers to publish in their journals. There is no information publicly available that shows the effect of these agreements on the underlying finances. Nor is there any information whether a fee waiver was granted for a given paper.
7. Many publishers are missing from this analysis. The publishers analyzed are some of the largest but there are hundreds (if not thousands) more that can be looked at.

Web Pages

1. <https://www.elsevier.com/open-access#4-faqs>, accessed 28 January 2024 (archived at <https://bit.ly/42eBEr8>)
2. <https://www.mdpi.com/about/announcements/5130>, accessed 15 November 2023 (archived at <https://bit.ly/3MO4jwN>)
3. <https://plos.org/publish/fees/>, accessed 29 January 2024 (archived at <https://bit.ly/3HW0faZ>)

4. <https://plos.org/financial-overview/>, accessed 9 February 2024 (archived at <https://bit.ly/3ujovk5>)
5. <https://journals.sagepub.com/gold-open-access-journals-disciplines>, accessed 28 January 2024 (archived at <https://bit.ly/3HVv3bT>)
6. <https://us.sagepub.com/en-us/sam/author-information>, accessed 3 February 2024 (archived at <https://bit.ly/3OtoKji>)
7. <https://www.springernature.com/gp/open-research/journals-books/journals>, accessed 02 February 2024 (archived at <https://bit.ly/49no8nr>)
8. <https://www.springernature.com/gp/products/journals>, accessed 02 February 2024 (archived at <https://bit.ly/3vZhzJp>)
9. <https://onlinelibrary.wiley.com/library-info/products/journals>, accessed 4 February 2024 (archived at <https://bit.ly/3UtFRW6>)
10. <https://www.annualreports.com/Company/john-wiley-sons-inc>, accessed 4 February 2024 (archived at <https://bit.ly/49njrd6>)
11. <https://www.forbes.com/sites/andrewdepietro/2023/08/24/highest-paying-college-professor-jobs-of-2023/>, accessed 16 November 2023 (archived at <https://bit.ly/3R0KpkA>)
12. <https://commonslibrary.parliament.uk/research-briefings/sn01079/>, accessed 16 November 2023 (archived at <https://bit.ly/47yw44g>)

Declarations

1. **Author Contributions:** As the sole author of this manuscript, Graham Kendall is responsible for all aspects of the paper including conceptualization, software, investigation and writing & revising.
2. **Conflict on interest:** The author has no conflict of interest in the preparation of this manuscript.
3. **Funding:** No funding was received for the purposes of this research.
4. **Data Availability:** The data is not being made available at the present time as this is a letter in response to an article. A more complete paper is being prepared at which point the full dataset will be made available.

References

- Butler, L-A., L. Matthias, M-A. Simard, P. Mongeon, S. Haustein. 2023. The oligopoly's shift to open access: How the big five academic publishers profit from article processing charges. *Quantitative Science Studies* In Press doi:10.1162/qss.a-00272.
- Kendall, G. 2021. Beall's legacy in the battle against predatory publishers. *Learned Publishing* **34**(3) 379–388. doi:10.1002/leap.1374.
- Kendall, G., S. Linacre. 2022. Predatory journals: Revisiting Beall's research. *Publishing Research Quarterly* **3**(38) 530–543. doi:10.1007/s12109-022-09888-z.
- Macháček, V., M. Srholec. 2022. Predatory publishing in Scopus: evidence on cross-country differences. *Quantitative Science Studies* **3**(3) 859–887. doi:10.1162/qss.a-00213.