Article processing charges: Mirroring the citation impact or legacy of the subscription-based model?

Nina Schönfelder [iD]

National Contact Point Open Access OA2020-DE, Library, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany

Check for updates

The MIT Press

ABSTRACT

With the ongoing open-access transformation, article processing charges (APCs) are gaining importance as one of the main business models for open-access publishing in scientific journals. This paper analyzes how much of APC pricing can be attributed to journal-related factors. With UK data from OpenAPC (which aggregates fees paid for open-access articles by universities, funders, and research institutions), APCs are explained by the following variables: (a) the "source normalized impact per paper" (SNIP), (b) whether the journal is open access or hybrid, (c) the publisher of the journal, (d) the subject area of the journal, and (e) the year. The results of the multivariate linear regression show that the journal's impact and hybrid status are the most important factors for the level of APCs. However, the relationship between APC and SNIP is different for open-access journals and hybrid journals. APCs paid to open-access journals were found to be strongly increasing in conjunction with higher journal citation impact, whereas this relationship was observed to be much looser for articles in hybrid journals. This paper goes beyond simple statistics, which have been discussed so far in the literature, by using control variables and applying statistical inference.

1. INTRODUCTION

The open-access transformation of scientific journals is under way. An increasing number of libraries and library consortia enter into transformative agreements with publishers (offsetting, read-and-publish, publish-and-read agreements) that make individual articles open access. Several models aim at flipping entire journals to open access, such as SCOAP³. New open-access journals have been founded by both native open-access publishers and subscription-based publishers. Research funders have tightened the open-access mandates of researchers (e.g., Plan S), which puts pressure on publishers to transform journals. Moreover, they provide funds for paying article processing charges (APCs).

The main motivation behind the open-access transformation is to make publicly funded research results more accessible, which could enhance further research and technological advance. The secondary motivation is to save public money and resolve the serial crisis. However, uncertainty about whether the open-access transformation is financially viable is an obstacle to concrete action, which some policy-oriented reports address.

In the influential "Max Planck Digital Library Open Access Policy White Paper," Schimmer, Geschuhn, and Vogler (2015) indicate that the money spent globally each year for the

research publishing system is sufficient to enable a large-scale open-access transformation. They conclude that current library acquisition budgets used for journal subscriptions are adequate to finance the open-access transformation of journals without risks. Schimmer et al. (2015) made a rough estimate that this holds true on a country level for Germany, the UK, and France. Lundén, Smith, and Wideberg (2018) make the same point for Sweden. Ilva, Laitinen, and Saarti (2016) argue that open-access publishing would be more affordable for Finland than the subscription-based model. In a pioneering report to the Joint Information Systems Committee (JISC), Houghton et al. (2009) identified through economic modeling that gold open access (i.e., open-access publishing in contrast to closed-access publishing) would be a more cost-effective scholarly communication system for the UK at the national level. In a further work, Swan and Houghton (2012) modeled the costs and benefits for four British universities with different characteristics regarding size and research intensity. They found that all universities would make savings from gold open access if APCs were at the then current averages. However, the most research-intensive institutions would face increased costs if the average level of APCs rose above £2,000.[1]

The drawback of all these studies is the dependence on the observed or assumed average APC. The problem with previous or currently observed APC averages is that they might substantially differ from what publishers will charge, on average, in a purely open-access publishing system. There will be differences for the following reasons: (a) The publishing system may shift to directions that cannot be foreseen today, and (b) the characteristics of contemporary open-access journals requesting APCs differ from the characteristics of subscriptions-based journals (e.g., their reputation and profile). If subscription-based journals flip to open access, they will probably charge different APC levels than are observed now.

Although the first issue is not resolvable, the second one is manageable. The main contribution of this paper is to identify publishers' pricing behavior according to some characteristics of their journals. This key finding (i.e., what influences APC pricing today and to what magnitude) can be used to infer what the APC for each journal will be after a hypothetical journal flipping to open access. From this, predictions could be made on both the average APC and their distribution in a purely open-access journal-publishing system, which is of utmost importance for policy recommendations. The "Pay It Forward" study conducted at the University of California Libraries (2016) was the first to break this new ground.

The literature on factors determining APC levels has relied so far on descriptive statistics, such as comparison of means, simple correlation coefficients, or visualizations via scatterplots. The literature suggests that APCs are related to the impact factor (Björk & Solomon, 2014; Solomon & Björk, 2012; University of California Libraries, 2016), the scientific discipline (Solomon & Björk, 2012; University of California Libraries, 2016), the type of publisher (commercial publisher vs. scientific society/university (Morrison et al., 2015; Solomon & Björk, 2012), subscription vs. gold open-access publisher (Björk & Solomon, 2014), and the publishing house (Jahn & Tullney, 2016). Björk and Solomon (2014), Jahn and Tullney (2016), and University of California Libraries (2016) show that APCs in hybrid journals are on average higher than in open-access journals. To my knowledge, Romeu et al. (2014) were the first to show that APCs for publications in open-access journals are much more strongly correlated with the journal impact factor than APCs in hybrid journals. A simple bivariate regression analysis of list-price APCs from 78 open-access journals on their "source normalized impact per

---

[1] Interestingly, the average APC paid in the UK and reported to OpenAPC is currently about this amount, taking into account the euro/pound exchange rate from May 2019.

paper" (SNIP) was first performed by the University of California Libraries (2016), although the regression did not control for any other factors and the statistical significance was not reported. Moreover, the study provides an economic model to explain the rationale for why the perceived quality of a journal is positively related to its APC.

However, all previous literature failed to examine the interdependence between the above-discussed factors. For example, the finding that APCs for publications in hybrid journals are on average more expensive than APCs in open-access journals could be resolved by the citation impact. Publishers could argue that hybrid journals have on average more citation impact than open-access journals, which are mostly market newcomers, and hybrid journals are therefore more valuable. A further problem with the previous literature is that readers less familiar with statistics could infer causality from correlations, which need not to be the case. Therefore, it is of utmost importance to use multivariate regression analysis and statistical inference for the improvement of our understanding on APC levels.

Throughout this study and consistent with the literature, I define an article processing charge as the fee for the publication of an open-access article in an open-access or hybrid journal. Usually, either the author directly or his or her institution is invoiced. Other fees eventually associated with publishing (e.g., submission, page, or color fees) are not considered as being part of APCs. APCs are charged to publish scientific articles in open access. That means "free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, [...] without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself " (BOAI, 2002). Open-access articles may be published either in open-access journals, where the complete content is open access, or in hybrid journals, where only some parts are open access and other parts have closed access and may be accessed via paying a subscription fee. Journals with completely closed-access content are called *subscription-based journals*. The phrase *open-access transformation* refers to the conversion of the publication system from closed access to open access—and within the purpose of this study—for scientific peer-reviewed journals.

This paper investigates how much of APC pricing can be attributed to journal-related factors. With data from OpenAPC, which is part of the INTACT project at the Bielefeld University Library, Germany, the APCs actually paid (in contrast to catalog prices) are explained by the following variables: (a) the SNIP of the CWTS Journal Indicators capturing the citation impact of a journal, (b) whether the journal is open access or hybrid, (c) the publisher of the journal, (d) the subject area of the journal, and (e) the year. I performed a multivariate linear regression on the total OpenAPC data set as well as on a subsample of British data from 2014 to 2017 to circumvent the problem of sample-selection bias.

The study design is limited to journals that cover all their costs associated with publishing articles via an author-facing APC. This paper does not aim at explaining why some journals charge an APC and others do not; neither does it aim to explore the total costs of publishing or the operating costs of publishers. In particular, it does not consider free-of-charge journals, which are issued by universities or research institutions, and run via in-kind support. Therefore, the use of OpenAPC, which does not record any free-of-charge publications, is consequential.

The paper is organized as follows. The OpenAPC data set and the CWTS Journal Indicator SNIP are explained and descriptive statistics are presented in Section 2. Section 3 first describes how to circumvent the issue of sample selection bias and then outlines the statistical model. Section 4 presents the results, the consequences of which are discussed in depth in Section 5. Section 6 concludes.

2. DATA

2.1. Data Sources

2.1.1. The OpenAPC data set

OpenAPC is a unique data set on APCs actually paid. OpenAPC is part of the INTACT project, which was funded by the Deutsche Forschungsgemeinschaft (German Research Funding Foundation) and, since October 2018, by the Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research), Germany. OpenAPC is located at the Bielefeld University Library with contributors from Europe and North America. It aggregates fees paid for open-access articles by universities, funders, and research institutions (see Broschinski and Pieper (2018) for more information on OpenAPC). Among data from numerous German, Swedish, and Norwegian universities and research institutions, OpenAPC aggregates data from the Austrian Fund for Scientific Research (FWF) and the British Wellcome Trust, as well as the Jisc Collections. In version 3.50.3 from January 25, 2019 (Jahn & Broschinski, 2019), which is used in this study, the OpenAPC data set comprises 72,975 observations in total.[2]

The OpenAPC data set is a sample of APCs paid by researchers from 13 countries. However, for most countries, the sample is small (particularly the United States) and clearly not representative of the APCs paid by their researchers (for example, Germany).[3] Nevertheless, it is the most comprehensive data collection on APCs actually paid that is publicly available.[4] Most importantly, OpenAPC offers a rather good representation of APCs that were paid in the UK. For the research question and the method, it is not important that the data at hand cover all or most of the APCs actually paid from a country, but that the sample is not biased or skewed to some factors related to both the APC level and the explanatory variables (SNIP, hybrid or open-access journal, publisher, subject area, period.) To my knowledge, there is no such bias for the UK (see Section 3.2 for an in-depth discussion).

For the purpose of this study, the following indicators were used from the OpenAPC data set:

- Top-level organization which covered the fee (*institution*)
- Year of payment (*period*)
- APC amount paid, including taxes, discounts etc.; excluding submission fees or page/color charges (*euro*)
- A Boolean indicator (*is_hybrid*) on whether the journal is hybrid (true) or gold open access (false)
- Publisher (*publisher*)
- Journal title (*journal_full_title*)

Information on the International Standard Serial Number (*issn*) as well as the linking ISSN (*issn_l*) is used for merging the OpenAPC data set with the CWTS Journal Indicators. An

---

[2] However, there is one reported APC that is out of realistic scope (above €10,000) and most probably the result of a typing error (misplaced decimal points). Therefore, this observation is deleted from the beginning.

[3] A quick glance at the Web of Science Core Collection reveals that there are almost one million gold open-access articles, reviews, and proceedings that were published between 2013 and 2017 in journals by North American and European researchers (including about 150,000 publications of British authors). However, no information is available as to whether these publications were associated with an APC. Although most open-access journals do not charge an APC (as reported by the Directory of Open Access Journals), the most publication-intensive journals, journals that belong to big publishers as well as hybrid journals, largely demand APCs.

[4] Maybe publishers have better data, but they are surely confidential.

institutional mapping table provided by the OpenAPC project is used to retrieve the country of the institution that covered the fee. More information on the OpenAPC data is given in Section 2.2.

2.1.2. *The CWTS journal indicators*

Within the research community, the number of published articles as well as the reputation and quality perception of journals in which the articles were published play a major role for career promotion. Journal citation metrics capture or at least try to capture some aspect of a journal's reputation and quality. Publisher emphasize impact factors of their journals to underline their relevance within the research field. In turn, authors frequently use impact factors to decide where to submit a manuscript. It is not the purpose of this paper to analyze or discuss whether journals citation metrics are suitable for research evaluation, career promotion, or subscription to journals. Moreover, I do not answer the question on whether a subscription or publication fee should be linked—from a normative point of view—to the journal's citation impact. I recognize that it does obviously play a role in scientific publishing. The focus of this study is on whether and how the journal's impact is linked to APCs charged, among other journal-related factors.

The indicator of journal citation impact that is used in this study is the "source normalized impact per paper" (SNIP) (CTWS, 2018). It is regularly compiled by the Centre for Science and Technology Studies (CWTS) at Leiden University. The indicator was introduced by Moed (2010) and further developed by Waltman et al. (2013). The SNIP is based on Elsevier's bibliographic database Scopus and uses a source normalized approach to correct for differences in citation practices between scientific fields. This is the main difference between the best-known indicator "Journal Impact Factor" (JIF) of Clarivate Analytics and the SNIP. The former is based on the Web of Science and is published in the Journal Citation Reports (JCR). Because of disciplinary differences in citation behaviors, it is not appropriate to compare the JIFs of journals between different research fields. The SNIP indicator addresses this problem by taking into account the citation characteristics of the journal's subject field (i.e., frequency that authors cite other papers; rapidity of maturing citation impact; extent to which the database used for the assessment covers the field's literature); see Moed (2010). For this reason, the SNIP—instead of the JIF—is applied within this study. The SNIP score ranges from zero to about 79 points. However, only few journals reach SNIP scores above three or four. By definition, the average SNIP value of the cited journals in a field (weighted by its number of publications) equals one (see Waltman et al., 2013).

The CWTS Journal Indicators were accessed in November 2018, with coverage up to 2017. The variables *Source.title*, *Source.type*, *Print.ISSN*, *Electronic.ISSN*, *ASJC.field.IDs* (to retrieve the subject area of the source), *Year*, and *SNIP* were used for further analysis. The analysis is limited to journals only. The CWTS Journal Indicators were merged with OpenAPC by using the ISSNs delivered by the respective data set and the ISSN-to-ISSN-L file from July 1, 2018 that contains a table matching all assigned ISSNs with their corresponding linking ISSN (CIEPS, 2018). This procedure delivered the highest match between both data sets. Data points without any ISSN could not be processed further. By merging both data sets, the OpenAPC data was enriched with the SNIP and the subject area of the respective journals.

2.2. Descriptive Statistics of the Enriched OpenAPC Data Set

In this section, statistics describing the enriched OpenAPC data set are presented. From this we will learn who mostly paid reported APCs and which publishers and journals received most

Table 1. Summary statistics of discrete variables, UK sample

	Frequency	Percentage
Institution		
UCL	6383	16.33
Wellcome Trust	4992	12.77
University of Cambridge	2986	7.64
University of Oxford	2641	6.76
Imperial College London	2524	6.46
University of Manchester	1885	4.82
(Other)	17678	45.22
Period		
2017	12239	31.31
2016	9970	25.51
2015	7413	18.96
2014	7113	18.20
2018	2175	5.56
2013	175	0.45
(Other)	4	0.01
Publisher		
Elsevier BV	8695	22.24
Springer Nature	6343	16.23
Wiley-Blackwell	4791	12.26
Public Library of Science (PLoS)	2318	5.93
Oxford University Press (OUP)	2058	5.26
BMJ	1402	3.59
(Other)	13482	34.49
Journal		
PLOS ONE	1691	4.33
Scientific Reports	1499	3.83
Nature Communications	812	2.08
BMJ Open	563	1.44
Nucleic Acids Research	260	0.67
NeuroImage	250	0.64
(Other)	34014	87.02

Table 1. (*continued*)

	Frequency	Percentage
Subject area		
Health Sciences	9643	24.67
Life Sciences	15405	39.41
Physical Sciences	8358	21.38
Social Sciences & Humanities	2150	5.50
NA	3533	9.04
Published in journal that is		
Open access	13093	33.50
Hybrid	25996	66.50

APC payments. Moreover, we will see how the observations are distributed over the journals' citation impact, subject area, and years. Table 1 provides summary statistics for the discrete variables. Large British universities as well as research funding and research organizations contributed most APC payments to OpenAPC. The last completed reporting year was 2017.[5] In this year, 12,239 APC-funded articles were registered from the UK. The number of observations is rising each year because an increasing number of institutions record APC payments and report them to OpenAPC. The reports from 2018 were incomplete at that time and therefore disregarded in the regression analysis.

In the OpenAPC UK sample, most APC-funded articles were published by Elsevier, Springer Nature, and Wiley-Blackwell—all of them being traditional subscription-based publishers. OpenAPC data suggests that large, traditionally subscription-based publishers dominate the market for open-access publications. Only the Public Library of Science (PLoS) might have noteworthy market shares. In total, OpenAPC reports APC payments to 211 publishers from UK.

However, APC-funded and reported articles were mostly published in the pure open-access megajournal *PLOS ONE* (about 4% of all articles), followed by *Scientific Reports*, which belongs to Springer Nature. The journals' subject areas confirm the practical experience that social sciences and humanities play a minor role in APC-based open-access publishing. About two thirds of the APCs were paid to publish an article in a hybrid journal and one third for publication in an open-access journal.

Table 2 summarizes the continuous variables APC in euros and SNIP. About half of the articles were published in journals that have SNIP values between 1.1 and 1.8. The average citation impact is about 1.6, which is above the standardized SNIP mean of 1, which is the impact of an average journal in a specific field. Very few articles were published in high-impact journals (see Figure 1). The most prestigious reported journal is *The Lancet* owned

---

[5] There is a delay between the payment of an APC and the inclusion of this APC in OpenAPC. Institutions need some time to publish their APC payment via GitHub or report them to a national aggregator (e.g., JISC in the UK). National aggregators collect and process the data and usually publish them once a year. These collections are successively integrated in OpenAPC.

Table 2. Summary statistics of continuous variables, UK sample

	APC in euros	SNIP
Minimum value	18	0.00
1st quartile	1599	1.10
Median	2167	1.35
Mean	2314	1.58
3rd quartile	2863	1.78
Maximum value	9858	16.12
Number of missing values	0	3292

by Elsevier. Unfortunately, about 3,300 SNIP observations are missing because the CWTS Journal Indicators are not calculated for all journals listed in the OpenAPC data set. In addition, the CWTS Journal Indicators for 2018 were not available at that time.

We now turn to a detailed description of the APCs in euros. The mean APC is about €2,300 and the median about €2,200. As one can see in Figure 2, the distribution is right skewed. There are many observations at the lower range, but some observations with high values raise the average APC. Fifty percent of the APC payments range from €1,599 to €2,863. There are few observations above €6,000 (39 from 39,089 payments).

Summary statistics and histograms are also provided for the total sample and the German sample in the supporting information. Most APC payments are reported from the UK, followed by Germany a long way behind. This proportion reflects different reporting behaviors rather than the true size of all APCs paid in these countries.[6] In addition, Austria, Sweden, and Norway reported actively to OpenAPC, although the number of contributed observations has remained low—most probably due to the size of these countries. The increase from 2013 to 2014 was mainly driven by British data.

There are remarkable differences between the British, German, and total sample for some indicators in the OpenAPC data set. However, it should be kept in mind that comparisons between countries might be misleading as, for example, the German sample is definitely not representative of the population. The average APC is higher in the UK and much lower in Germany and the other countries. German APCs are barely above €2,000, most probably due to the APC-funding rules (price cap). In the total sample, almost half of reported APCs stem from publications in hybrid journals, but only 1% in the German sample. On average, British authors published in journals with more citation impact than German authors did. These differences largely reflect different APC-funding rules in the countries. APC funding in Germany is much more restrictive than in the UK. My interest is in explaining the APC-pricing behavior of publishers in general—not (yet) the influence of funding policies. Therefore, the regression analysis in Section 4.1 builds upon the UK sample.

The rest of the section presents plots showing the relationships between two indicators. The first is a scatter plot between APCs and the associated SNIP values (Figure 3). Each point

---

[6] Science-Matrix (2018, p. 20) provides a table showing the total number of published articles and open-access levels (green vs. gold) for Germany and the UK. In 2014, British authors published slightly more articles in total and immediately in open access (gold route) than German authors.
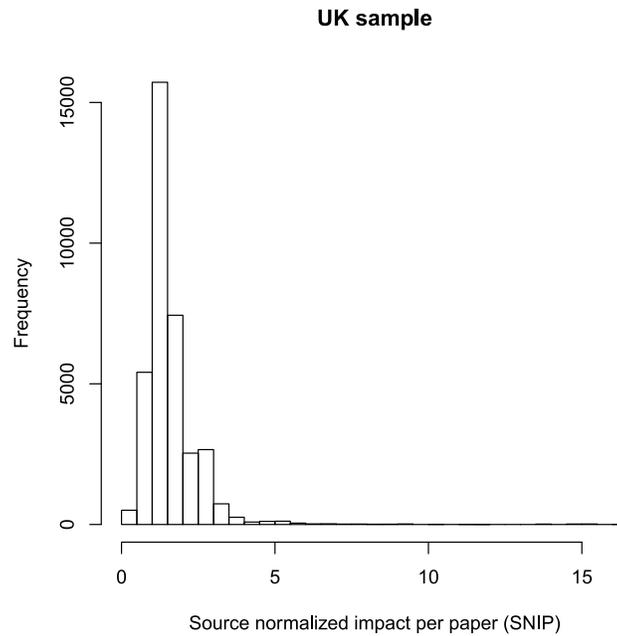
UK sample

Figure 1. Histogram of SNIP for OpenAPC records.

represents an article with its combination of APC and SNIP. The line shows the correlation between the two variables. Although the positive correlation seems to be weak, it is statistically highly significant (test statistic not reported here). Hence, articles in higher impact journals are charged more than in lower impact journals.

Breaking down APC payments for publications in open-access and hybrid journals (Figure 4) highlights that APCs in hybrid journals are much more expensive than in open-
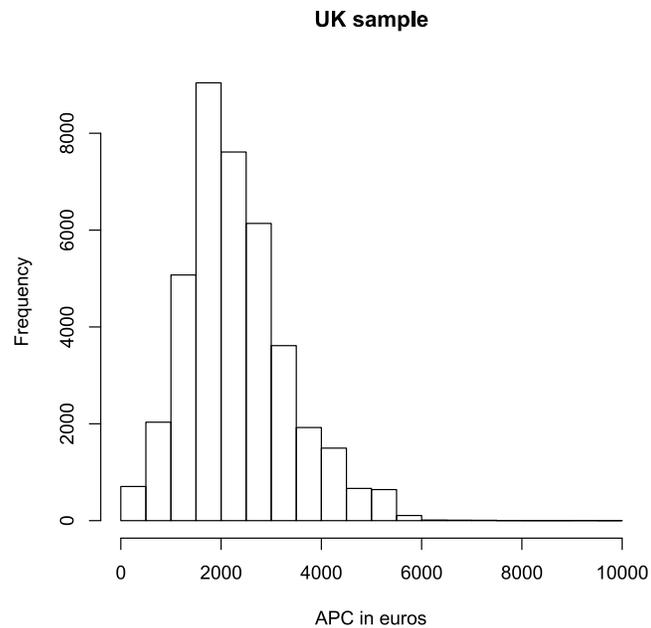
UK sample

Figure 2. Histogram of APC in euros.
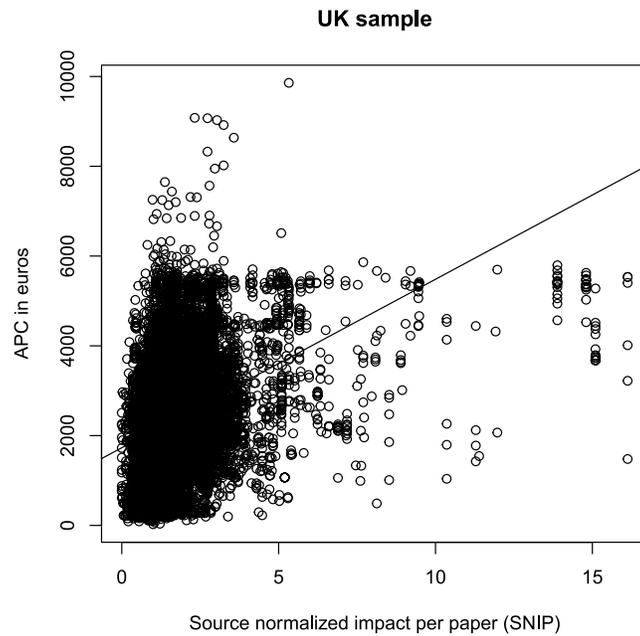
UK sample

Figure 3. Scatter plot APC vs. SNIP.

access journals. Moreover, hybrid journals tend to have a higher citation impact compared to open-access journals (see Figure 5).

Figure 6 presents the share of reported articles published in hybrid or open-access journals for the "big" publishers. Within the group of "other publishers," two thirds of the articles from the UK were released to the public in hybrid journals. Analyzing the shares of each "big" publisher shows a different picture. Either (almost) all articles were published in open-access
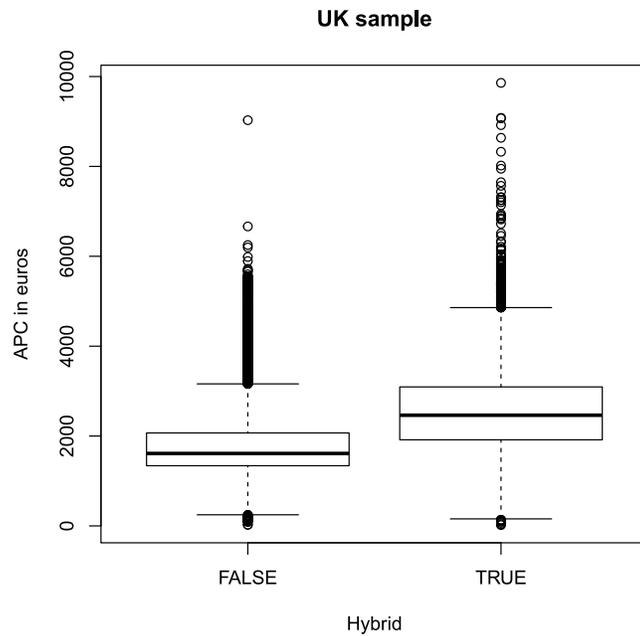
UK sample

Figure 4. Box plots of APC in euro for open-access and hybrid journals.
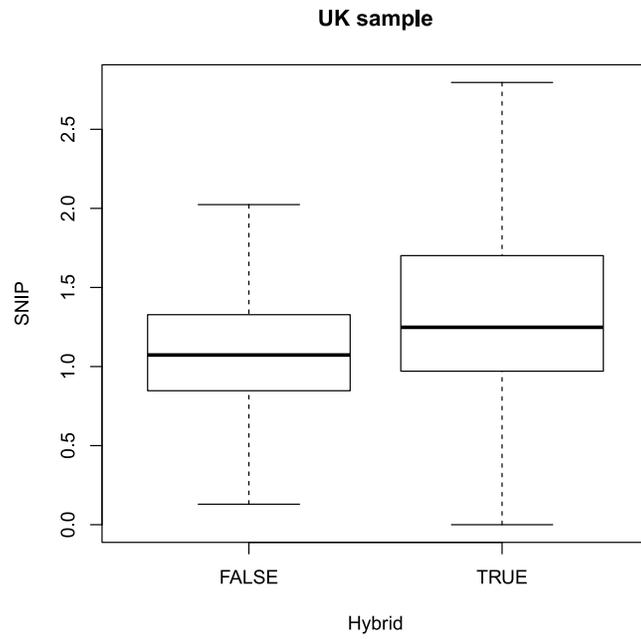
UK sample

Figure 5. Box plots of SNIP for open-access and hybrid journals in 2017. Note: Outliers are not drawn.
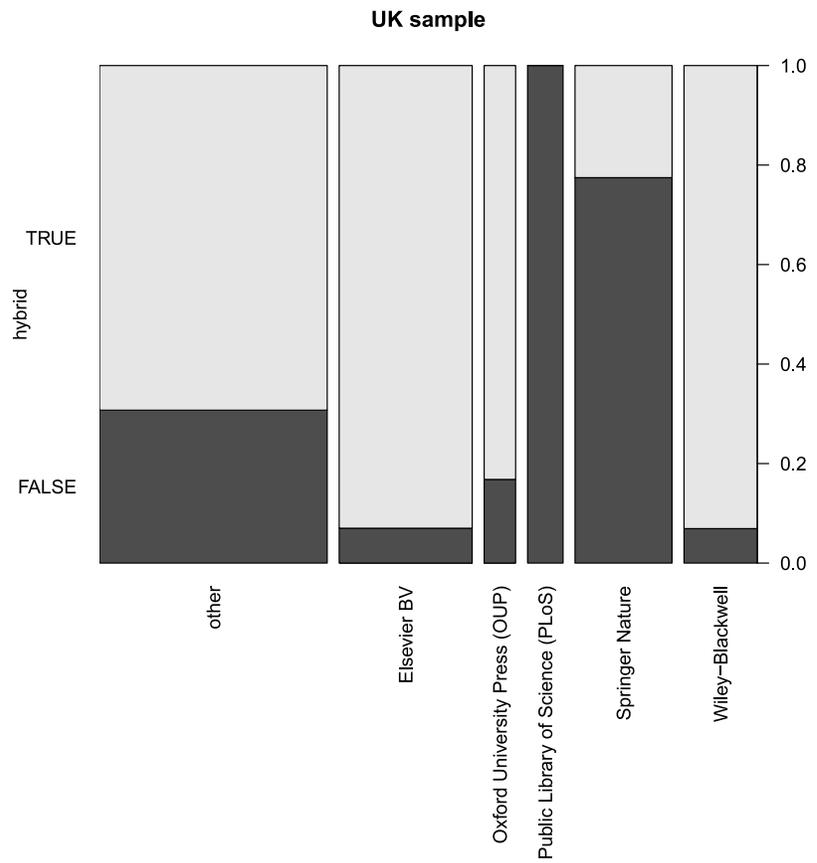
UK sample

Figure 6. Distribution of articles among publishers and open-access vs. hybrid journals.

journals (PLoS and Springer Nature), or almost all articles were published in hybrid journals (Elsevier, Wiley-Blackwell, Oxford University Press).

There are wide differences in APCs levels between the publishers, as one can see in the box plots of Figure 7. The median as well as the upper and lower quartiles of APC payments are the highest for Elsevier, followed by Oxford University Press. This means that these two publishers often charge expensive APCs. APCs are relatively low at PLoS, and they do not vary as much as at the other big publishers.

3. METHOD

3.1. Statistical Model

In this section, we will further investigate the factors behind the different APC levels. A multivariate, linear regression analysis is performed, where the independent variables *SNIP*, *Hybrid*, *Big_publisher*, *Subject_area*, and the year γ explain the dependent variable *APC*:

$$APC_{it} = \alpha_i + \beta_1 SNIP_{it} + \beta_2 Hybrid_{it} + \beta_3 SNIP_{it} \times Hybrid_{it}$$

$$+ \mathbf{Big\_publisher}'_{it}\beta_4 + \mathbf{Subject\_area}'_{it}\beta_5 + \gamma_t + \epsilon_{it}. \tag{1}$$
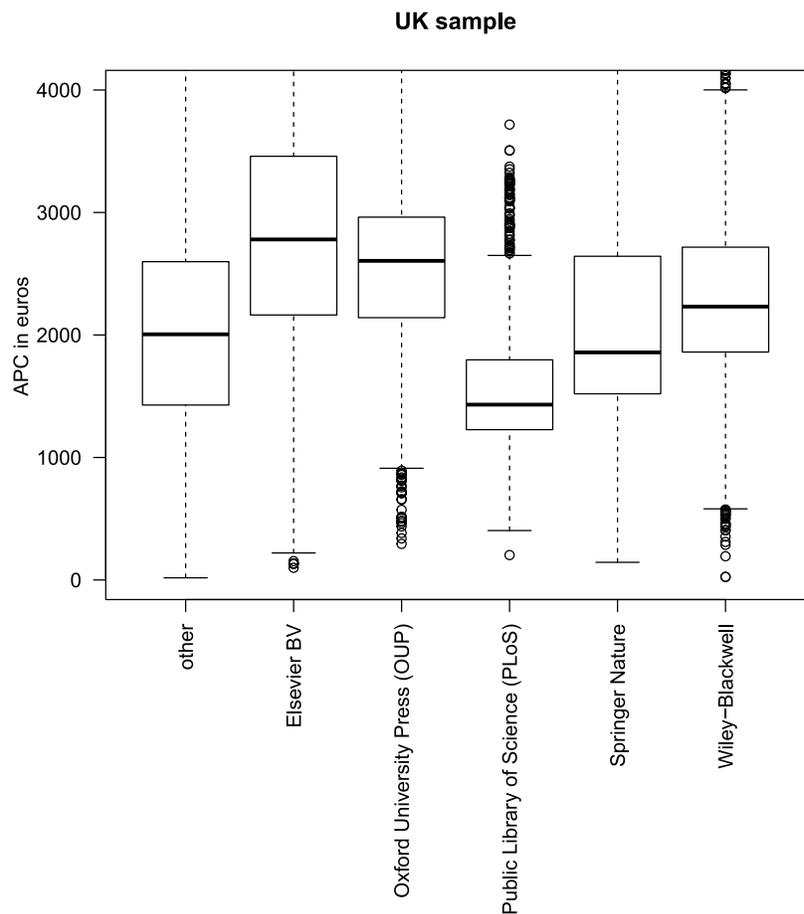
UK sample

Figure 7. Box plots of APC for each publisher. Note: Some APCs are out of scale.

The variable *Big_publisher* is a column vector of dummy variables indicating the six largest publishers according to the respective sample of OpenAPC. The base group contains all other publishers. Likewise, *Subject_area* is a column vector of the four subject areas to which each journal is assigned, where health sciences is the base group. β_4 and β_5 are the corresponding vectors of coefficients, α_i is the individual-specific effect, and ϵ_{it} is the disturbance term. The subscripts i and t denote the ith observation at the tth period. Moreover, I expect that the explanatory power of SNIP is different for hybrid and open-access journals. That is why the estimation equation contains an interaction term between *SNIP* and *Hybrid*.[7]

To illustrate the interpretation of the coefficient of *Hybrid* and its interaction term with *SNIP*, conditional expectations of Eq. (1) are presented. For open-access journals, the conditional expectation is

$$\mathrm{E}[APC_{it}|Hybrid_{it}=0] = \alpha_i + \beta_1 SNIP_{it} \\ + \mathbf{Big\_publisher}'_{it}\boldsymbol{\beta}_4 + \mathbf{Subject\_area}'_{it}\boldsymbol{\beta}_5 + \gamma_t. \tag{2}$$

For hybrid journals, the conditional expectation of Eq. (1) is

$$\mathrm{E}[APC_{it}|Hybrid_{it}=1] = (\alpha_i + \beta_2) + (\beta_1 + \beta_3)SNIP_{it} \\ + \mathbf{Big\_publisher}'_{it}\boldsymbol{\beta}_4 + \mathbf{Subject\_area}'_{it}\boldsymbol{\beta}_5 + \gamma_t. \tag{3}$$

Hence, β_2 induces an intercept shift and β_3 induces a slope shift.

The OpenAPC data set is not a panel but a repeated cross-section. That means that data are obtained by a sequence of independent samples, where the unit of each sample is the article. I performed a static linear regression with random and time effects based on T successive cross-sections. Therefore, heteroskedasticity had to be taken into account and robust standard errors were calculated for hypotheses tests.[8] Eq. (1) is estimated by pooled ordinary least squares (OLS) and the results are reported in Section 4.[9]

3.2. Sample Selection

Two issues arise that could lead to biased coefficient estimates: sample selection and missing data. The first issue arises if the sample at hand is not representative of the population. This would render OLS parameter estimates inconsistent (Cameron & Trivedi, 2006, p. 529). In our case, we observe a sample of APCs, but for some countries, the sample is not randomly drawn from the population, as high APCs are systematically underreported to the OpenAPC project. In Germany, the Deutsche Forschungsgemeinschaft (DFG)—a funding organization—supports publication funds at some universities. If a member of the university is the submitting or corresponding author of an article in an open-access journal, the publication fund can take over the obligation to pay the APC up to €2,000. The APC must not be above this limit to be covered by the DFG-supported publication fund. Otherwise, the author has to pay the APC out of department, third-party, or private funds. Publication funds systematically report to the OpenAPC project, whereas there are almost no ways to report otherwise-funded APCs. To make things worse, authors could choose not to publish in expensive open-access journals at all but to publish in subscription-based journals. Having this in mind, it is possible to infer the determinants for APCs up to €2,000 but not above. The sample selection could be more

---

[7] I also considered nonlinear relationships between APC and SNIP. However, it turned out that linearization is not necessary.

[8] See Cameron and Trivedi (2006, p. 47 and pp. 770–771) for a discussion on repeated cross-sections.

[9] The results were obtained using R 3.4.3 (R Core Team, 2017) with the packages lmtest 0.9-35 (Zeileis & Hothorn, 2002), sandwich 2.4-0 (Zeileis, 2004), car 2.1-6 (Fox & Weisberg, 2011), texreg 1.36.23 (Leifeld, 2013) and xtable 1.8-2 (Dahl, 2016).

or less severe depending on the national conditions for APC funding. The stricter the conditions (e.g., a price cap) the less representative the sample is likely to be. To my knowledge, the conditions for APC funding are least restrictive in the UK. There are no price caps for APCs, and APCs are funded for publications in both open-access and hybrid journals. Fortunately, the OpenAPC data set contains plenty of UK data from 2014 to 2017, so that I can base the entire analysis on the UK sample (see Section 4.1), largely avoiding the problem of sample selection and inconsistent estimates.

The second issue that could lead to biased coefficient estimates arises if the data set has missing observations. In the UK sample, there are approximately 3% of observations with missing citation impact (SNIP) and subject area (e.g., life sciences). I assessed the direction and range of the potential bias on the estimation results due to missing data and report them in the supporting information. Missing data turned out to be a minor problem.

4. RESULTS

4.1. Results of the UK Sample

Turning now to the estimates of Eq. 1, Table 3 shows the results of four models based on the UK sample from 2014 to 2017.[10]

The first model is a bivariate regression of APC on SNIP, which already explains 11% of the total variance. In the second model, APC levels are explained by whether the article was published in a hybrid or open-access journal. Indeed, APCs in hybrid journals are more expensive. This variable explains 11% of the total variance in a bivariate regression. Combining both variables (including their interactions term) represents Model 3, where 24% of the total variance is explained and all coefficients are statistically significant. The coefficient of SNIP is about €888, which means that, on average, an open-access journal with a SNIP value of 2 charges about €888 more than an open-access journal with a SNIP value of 1 (other things being equal). Likewise, a hybrid journal is estimed to charge, on average, about €1,550 more than an open-access journal (again, other things being equal). However, a hybrid journal is less sensitive to its impact. For each additional SNIP score, it charges just about €209 (≈ 888 − 679) more. To sum up, hybrid journals tend to be more expensive and less sensitive to their citation impact than open-access journals. In Model 4, the total set of variables is included to explain APC levels. The dummy variables indicating a big publisher, the subject area, and the year do not add as much to the adjusted R^2. Nevertheless, most coefficients are statistically significant and economically substantial. Publishing in Elsevier journals is quite expensive (on top of the fact that most Elsevier journals are hybrid) and least expensive in PLoS journals. Publishers might follow different price-setting strategies, or some reputation is associated with a publisher label, which is not reflected in the SNIP. Publications in life sciences are much costlier than in social sciences and humanities.

Furthermore, the results seem to indicate a price increase from 2014 to 2015. There are several potential explanations for this finding, which cannot be identified from this research

---

[10] Inspecting residuals (not reported here) shows no serious problems with outliers. However, for economic reasoning, I decided to disregard the lowest and highest 1% of APCs from the UK sample as outliers. On the one hand, the lowest 1% of APCs are likely not stand alone APCs (below €304) because the minimum cost for publishing an article in a reliable journal is well above this amount. Publishing in reliable journals, these APCs could be subsidized by organizations, or discounted APCs because of waivers or personal membership in scientific communities or learned societies. On the other hand, the highest 1% of APCs are most probably the result of typing error (above €5,349). List-price APCs do not exceed this amount, even at the most expensive journals.

Table 3. Statistical models, UK sample

	Model 1	Model 2	Model 3	Model 4
(Intercept)	1747.30***	1854.59***	631.69***	281.31***
	(17.43)	(8.37)	(25.23)	(25.20)
SNIP	360.17***		887.63***	844.24***
	(11.42)		(19.89)	(17.52)
is_hybrid		666.79***	1548.85***	1529.17***
		(10.23)	(28.84)	(26.88)
SNIP × is_hybrid			−678.95***	−641.44***
			(21.57)	(19.42)
Elsevier BV				379.20***
				(13.44)
Oxford University Press (OUP)				−2.95
				(21.16)
Public Library of Science (PLoS)				−233.67***
				(14.27)
Springer Nature				270.10***
				(14.13)
Wiley-Blackwell				2.32
				(13.24)
Life Sciences				201.88***
				(10.94)
Physical Sciences				−147.55***
				(12.59)
Social Sciences and Humanities				−450.28***
				(22.77)
period 2015				324.96***
				(13.12)
period 2016				292.43***
				(11.97)
period 2017				316.80***
				(11.64)
R^2	0.11	0.11	0.24	0.33

Table 3. (*continued*)

	Model 1	Model 2	Model 3	Model 4
Adj. R^2	0.11	0.11	0.24	0.33
Num. obs.	34944	35999	34944	34715
RMSE	906.75	912.77	838.09	787.84

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

design. First, it might be that at least one major or many publishers increased APCs in 2015. Second, exchange rates evolved unfavorably concerning the euro (i.e., the euro devaluated against the pound sterling, the US dollar, or other currencies that publishers use for billing purposes). Third, it cannot be ruled out that the currency conversion made by OpenAPC is slightly inaccurate, as annual average spot rates are used if no information on the day or month of payment is available, and therefore information is lost. However, this would only result in a problem if APC payments were unequally distributed over a year. Of course, APC development over the years indicated by the estimates could have resulted from a combination of these three potential explanations, either cumulative or alleviative. Without doubt, exchange rate shifts have a substantial effect on the level of APC payments if APC bills are denominated in foreign currencies. Because price increases and exchange rate movements are interesting and important topics on their own, I will further investigate them in a follow-up paper. In this study, the period dummies are mainly used for controlling purposes. This ensures that the results concerning the journal-related factors (SNIP, subject area, etc.) are not biased by period-specific effects or trends.

To make the results more clear, estimated APC equations are presented for two publishers (representing two opposite extremes),[11] which both produced journals in life sciences in 2017. Eq. (4) predicts an APC for an open-access article at PLoS depending on the impact of the respective journal:

$$\widehat{APC}_{it} = (281-234+202+317) + 844 \times SNIP_{it}$$
$$= 566 + 844 \times SNIP_{it} \tag{4}$$

Eq. (5) predicts an APC for an open-access article in an Elsevier hybrid journal, other things being equal:

$$\widehat{APC}_{it} = (281+379+202+317+1529) + (844-641) \times SNIP_{it}$$
$$= 2708 + 203 \times SNIP_{it} \tag{5}$$

On the one hand, the APC's component that is not related to the journal's impact is almost four times higher for publications in Elsevier hybrid journals (€2,708) than at PLoS (€566). On the other hand, Elsevier is estimated to charge just €203 for each SNIP score, compared to €844 by PLoS. In the end, it depends on the journal's impact whether an article published by PLoS or by an Elsevier hybrid journal is predicted to be more expensive. Assume a SNIP score

---

[11] After controlling for other effects, Elsevier appears as the most and PLoS the least expensive publisher. These publishers have the highest and the lowest estimated coefficient, respectively.

Table 4. In-sample APC predictions

	PLoS, OA	Elsevier hybrid
SNIP = 1	$\widehat{APC}_{it} = €1410$	$\widehat{APC}_{it} = €2911$
SNIP = 1.35	$\widehat{APC}_{it} = €1705$	$\widehat{APC}_{it} = €2982$
SNIP = 1.78	$\widehat{APC}_{it} = €2068$	$\widehat{APC}_{it} = €3069$
SNIP = 2	$\widehat{APC}_{it} = €2254$	$\widehat{APC}_{it} = €3114$
SNIP = 15	$\widehat{APC}_{it} = €13226$	$\widehat{APC}_{it} = €5753$

Note: The in-sample APC prediction for an open-access journal with a SNIP score of 15 is are rather hypothetical consideration, as no open-access journal has comparable impact.

of one, which is the citation impact of an average journal in a specific field by definition. For example, the journals *PLOS ONE* and *Molecular and Cellular Endocrinology* had a SNIP of approx. one in 2017,[12] both located in life sciences. Then, we can derive the following estimated APCs:

- *PLOS ONE* article: $\widehat{APC}_{it} = 566 + 844 = €1410$
- *Molecular and Cellular Endocrinology* article: $\widehat{APC}_{it} = 2708 + 203 = €2911$

These are examples for in-sample predictions. In Table 4, predicted APCs are presented for PLoS journals and Elsevier hybrid journals with varying levels of citation impact. A SNIP value of one corresponds approximately to the first quartile of the UK sample as well as the total OpenAPC data set. The median of the UK sample is 1.35 and 1.78 is its third quartile. A SNIP value of 15 is about the highest impact a journal has in the OpenAPC data set (*The Lancet*). However, no gold open-access journal has a comparable citation impact. The predicted APCs vary greatly for PloS journals along the citation impact but only slightly for Elsevier hybrid journals. Eighty percent of the reported articles from the UK appeared in journals with a SNIP score below 2. For them, APCs in hybrid journals are predicted to be much costlier than in the open-access counterparts.

To conclude, the journal's impact mirrors APCs in open-access journals, especially at open-access publishers, far better than in hybrid journals, particularly those that are published by the big, traditionally subscription-based publishers.

4.2. Results of the Total Sample

Table 5 presents the regression results of two models based on the total sample. In Model 2, country dummy variables are added to account for country-specific effects (Austria is the baseline country), but their interpretation can be questioned due to the sample-selection problem. The overall findings are the same, but the magnitudes of the coefficients differ somewhat. Because of the sample-selection problem, my conclusions are drawn from the UK sample (Model 4 in Table 3).

5. DISCUSSION

The purpose of this paper was to identify publishers' APC-pricing behavior according to some characteristics of their journals. The results provide evidence that the journal's citation impact

---

[12] Actually, the SNIP values were at 1.1 in 2017 for both journals. For the sake of clarity, the examples are calculated with SNIP values of 1.

Table 5. Statistical models, total sample

	Model 1	Model 2
(Intercept)	638.34 (17.41)***	−445.00 (131.79)***
SNIP	733.68 (14.55)***	700.45 (13.84)***
is_hybrid	1482.00 (21.29)***	1300.46 (21.95)***
SNIP × is_hybrid	−497.34 (16.32)***	−476.57 (15.54)***
Elsevier BV		410.88 (11.80)***
Frontiers Media SA		256.56 (11.78)***
Public Library of Science (PLoS)		−109.97 (8.60)***
Springer Nature		221.38 (8.82)***
Wiley-Blackwell		138.35 (11.31)***
Life Sciences		195.63 (7.98)***
Physical Sciences		−134.56 (9.39)***
Social Sciences and Humanities		−313.78 (16.66)***
period 2006		634.73 (133.77)***
period 2007		696.54 (129.58)***
period 2008		880.87 (131.92)***
period 2009		792.17 (131.79)***
period 2010		971.46 (134.15)***
period 2011		827.03 (129.93)***
period 2012		885.90 (129.36)***
period 2013		873.23 (129.31)***
period 2014		878.92 (129.18)***
period 2015		1172.59 (129.19)***
period 2016		1194.55 (129.14)***
period 2017		1228.91 (129.11)***
country CAN		−322.81 (18.51)***
country CHE		−246.49 (39.83)***
country CZE		−335.54 (63.74)***
country FRA		−256.85 (18.07)***
country GBR		0.78 (12.13)
country DEU		−192.64 (12.46)***
country ESP		−653.24 (25.86)***

Table 5. (*continued*)

	Model 1	Model 2
country GRC		−440.81 (24.01)***
country ITA		−249.17 (65.76)***
country NOR		−265.86 (17.79)***
country SWE		−252.72 (16.47)***
country USA		−500.54 (33.57)***
R^2	0.34	0.43
Adj. R^2	0.34	0.43
Num. obs.	66858	66313
RMSE	787.21	729.58

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

as well as the hybrid status are the most important drivers of APC levels. In line with findings by Solomon and Björk (2012), Björk and Solomon (2014), and the University of California Libraries (2016), my analysis shows that there is a positive and statistically significant relationship between the citation impact and the requested APC—for both open-access and hybrid journals. In fact, two pricing patterns emerge. The journal's impact greatly influences APC levels in open-access journals, whereas it slightly alters APCs in hybrid journals. In open-access journals, each additional SNIP score is associated with a €845 higher APC but only with €203 more in hybrid journals. In this respect, my regression analysis confirms the insights from descriptive statistics of Romeu et al. (2014), who have found that APCs are much more strongly correlated with the JIF in open-access journals than in hybrid journals. The University of California Libraries (2016) were the first to perform a regression analysis, albeit on a small sample, without any controls and no reported significance levels. Their finding that each additional SNIP point is associated with an approximately $710 higher APC in open-access journals fits surprisingly well with the results of my analysis.

Björk and Solomon (2014), Jahn and Tullney (2016), and the University of California Libraries (2016) argue that APCs in hybrid journals are on average higher than in open-access journals. However, they did not control for the journal citation impact. In fact, I present convincing evidence that the fraction of the APC that is not related to the citation impact is much higher for publications in hybrid journals compared to open-access journals (additional €1,530 for the base group). Moreover, my data suggest that the native open-access publisher PLoS tends to charge less than traditional subscription-based publishers (Elsevier and Springer Nature) for comparable journals, which is in line with the conclusions of the literature (Björk & Solomon 2014; Jahn & Tullney, 2016). In addition, I can confirm the influence of the scientific discipline on APCs found by Solomon and Björk (2012) and the University of California Libraries (2016). APCs for publications in life and health sciences are more expensive than in physical sciences and least expensive in social sciences and humanities, even when controlled for other journal-related factors. To sum up, hybrid journals tend to be more expensive and are less sensitive to their citation impact than open-access journals. With reference to the title of this paper, the evidence suggests that APCs are mirroring the citation impact in

Table 6. Actual and predicted total amount of APCs

	Total amount of APCs, in euros
UK, actually paid	82,339,469
UK, as if all open access	70,997,330
UK, as if all hybrid	89,239,866
Total, actually paid	132,395,680
Total, as if all open access	116,774,565
Total, as if all hybrid	157,534,717

Note: Only complete cases as recorded in OpenAPC.

open-access journals, especially at native open-access publishers, but are a legacy of the subscription-based model in hybrid journals, often at Elsevier, Springer Nature, and co.

Overall, this paper largely confirms the previous knowledge obtained from descriptive statistics on the relations between APCs and journal attributes. This paper's main contribution is to control for interdependencies between the above-discussed factors. To isolate the marginal effect of one variable (e.g., citation impact) on APCs, it is necessary to take into account the other relationships that might influence APCs. This was done in the regression analysis. Moreover, with the help of statistical inference, it is possible to calculate confidence intervals and perform significance tests on the observed relationships. Provided the APC equation is correctly specified, this paper

- demonstrates that the relationship between APCs and the other variables is not random, and
- shows the magnitude (in euros) of the marginal effect of each variable on the APC level.

The estimated equation could be used to predict APCs (in euros) for currently closed-access journals or for journals for which we lack APC information. Moreover, the estimated equation can help to answer two questions relevant for policy design and for making strategic decisions in libraries. The first is how much hybrid journals would charge if they flipped to open-access and adopted the open-access price-setting behavior. The second is how much open-access journals would charge if they adopted the hybrid pricing-setting behavior.

To get an idea of what the two pricing patterns imply for the financial aspects of the open-access transformation, I calculated two hypothetical scenarios. What would have been the total APC amount if all articles recorded in OpenAPC had been charged as if they were published in open-access journals? And what would be the sum if they were all published in hybrid journals (leaving other journal characteristics unchanged)? Table 6 presents the hypothetical amounts in euros for the UK sample from 2014 to 2017 and the total sample and compares it with the actual sums. The calculations show that the UK higher education and research system would have saved more than €11 million on OpenAPC-recorded articles if all journals had charged according to the open-access pricing pattern. In contrast, all countries would have spent about €25 million more on APCs if all articles recorded in OpenAPC had been charged according to the hybrid pattern. The effects for all APCs paid from these countries would have been even higher.

Which pricing behavior will dominate in the future after a full journal flipping is crucial. If the pricing behavior of the traditional, subscription-based publishers wins through, the open-access

transformation will come at a much higher cost than expected today from libraries, higher education, and research institutions. Therefore, provisions to introduce competition between publishers and journals are of utmost importance.

The rationale for linking APCs to the journal citation impact is clearly research evaluation. Currently, the evaluation of individual researchers and even entire higher education and research institutions depends much on the use of journal citation metrics as the SNIP or the JIF. Consequently, researchers tend to pursue SNIP/JIF maximizing publishing and probably pay every APC they can afford or their funder takes over. If research funders and higher education and research institutions could make a shift toward assessment based on researchers' own achievements rather than on the basis of the journal in which the research is published,[13] APCs in journals will likely be more level and competitive than observed today.

6. CONCLUSION

APCs are gaining importance as one of the main business models for open-access publishing in journals. By investigating the journal-related factors influencing APC levels, this paper presented key findings that could be used to assess whether the open-access transformation of journals is a financially viable way for individual higher education and research institutions as well as entire countries.

The results show that the journal's impact and the hybrid status are the most important factors for the level of an APC. However, the journal's impact alters the APCs little for publications in hybrid journals, whereas it is crucial for the level of APCs in open-access journals. The journal's subject area and publisher also affect APCs. Moreover, the year of payment influences APCs, although this paper cannot identify whether it is because of price increases or exchange rate movements. To date, it remains an open question how (country-specific) conditions for research and open-access funding interact with APCs.

---

[13] See, for example, the San Francisco Declaration on Research Assessment (https://sfdora.org/read/).

DATA AVAILABILITY

Please see the Supporting Information section above concerning data sets.

REFERENCES

BOAI. (2002). *Budapest Open Access Initiative declaration*. Version 2002-02-14. http://www.budapestopenaccessinitiative.org/read

Björk, B.-C., & Solomon, D. (2014). *Developing an Effective Market for Open Access Article Processing Charges*. Final Report to a consortium of research funders comprising Jisc, Research Libraries UK, Research Councils UK, the Wellcome Trust, the Austrian Science Fund, the Luxembourg National Research Fund and the Max Planck Institute for Gravitational Physics. London: Wellcome Trust. https://wellcome.ac.uk/sites/default/files/developing-effective-market-for-open-access-article-processing-charges-mar14.pdf

Broschinski, C., & Pieper, D. (2018). OpenAPC: A contribution to a transparent and reproducible monitoring of fee-based open access publishing across institutions and nations. *Insights*, *31*. https://doi.org/10.1629/uksg.439

Cameron, A. C., & Trivedi, P. K. (2006). *Microeconometrics: Methods and applications*. Cambridge: Cambridge University Press.

CIEPS. (2018). *The ISSN-L table*. Version 2018-07-01. The International Centre for the registration of serial publications. https://www.issn.org/services/online-services/access-to-issn-l-table/

CTWS. (2018). *CWTS Journal indicators*. Version 2018-05-01. Leiden University's Centre for Science and Technology Studies. http://www.journalindicators.com/Content/CWTS%20Journal%20Indicators%20May%202018.xlsx

Dahl, D. B. (2016). *xtable: Export tables to LaTeX or HTML*. https://CRAN.R-project.org/package=xtable

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*. 2nd ed. Thousand Oaks, CA: Sage. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Houghton, J. et al. (2009). *Economic implications of alternative scholarly publishing models: Exploring the costs and benefits*. A report to the Joint Information Systems Committee (JISC). Victoria University. Loughborough University. http://vuir.vu.edu.au/15222/1/EIASPM_Report.pdf

Ilva, J., Laitinen, A., & Saarti, J. (2016). The costs of open and closed access: Using the Finnish research output as an example. *LIBER Quarterly*, *26*(1): 13–27. https://doi.org/10.18352/lq.10137

Jahn, N., & Broschinski, C. (2019). *Datasets on fee-based open access publishing across German institutions: OpenAPC*. Version 3.50.3. Bielefeld University. https://doi.org/10.4119/UNIBI/UB.2014.18. https://github.com/OpenAPC/openapc-de/releases/tag/v3.50.3

Jahn, N., & Tullney, M. (2016). A study of institutional spending on open access publication fees in Germany. *PeerJ*, *4*, e2323. https://doi.org/10.7717/peerJ.2323

Leifeld, P. (2013). texreg: Conversion of statistical model output in R to HTML tables. *Journal of Statistical Software*, *55*(8): 1–24. http://www.jstatsoft.org/v55/i08/

Lundén, A., Smith, C., & Wideberg, B. M. (2018). National licence negotiations advancing the open access transition—a view from Sweden. *Insights*, *31*(12): 1–7. https://doi.org/10.1629/uksg.413

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, *4*(3): 265–277. https://doi.org/10.1016/j.joi.2010.01.002

Morrison, H. et al. (2015). Open access article processing charges: DOAJ survey May 2014. *Publications*, *3*(1): 1–16. https://doi.org/10.3390/publications3010001

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Romeu, C. et al. (2014). *The SCOAP3 initiative and the Open Access Article-Processing-Charge market: Global partnership and competition improve value in the dissemination of science*. Geneva: CERN. https://doi.org/10.2314/CERN/C26P.W9DT

Schimmer, R., Geschuhn, K. K., & Vogler, A. (2015). *Disrupting the subscription journals' business model for the necessary large-scale transformation to open access*. A Max Planck Digital Library Open Access Policy White Paper. Max Planck Digital Library. https://doi.org/10.17617/1.3

Science-Matrix. (January 2018). *Analytical support for bibliometrics indicators: Open access availability of scientific publications*. http://www.science-metrix.com/sites/default/files/science-metrix/publications/science-metrix_open_access_availability_scientific_publications_report.pdf

Solomon, D. J., & Björk, B. C. (2012). A study of open access journals using article processing charges. *Journal of the American Society for Information Science and Technology*, *63*(8): 1485–1495. https://doi.org/10.1002/asi.22673

Swan, A., & Houghton, J. (2012). *Going for Gold? The costs and benefits of Gold Open Access for UK research institutions: Further economic modelling*. Report to the UK Open Access Implementation Group. http://wiki.lib.sun.ac.za/images/d/d3/Report-to-the-uk-open-access-implementation-group-final.pdf

University of California Libraries. (2016). *Pay It Forward: Investigating a sustainable model of open access article processing charges for large North American research institutions*. Mellon Foundation. http://icis.ucdavis.edu/wp-content/uploads/2016/07/UC-Pay-It-Forward-Final-Report.rev_.7.18.16.pdf

Waltman, L. et al. (2013). Some modifications to the SNIP journal impact indicator. *Journal of Informetrics*, *7*(2): 272–285. https://doi.org/10.1016/j.joi.2012.11.011

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, *11*(10): 1–17. http://www.jstatsoft.org/v11/i10/

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3): 7–10. https://CRAN.R-project.org/doc/Rnews/