QSS

Check for updates

The MIT Press

The APC-barrier and its effect on stratification in open access publishing

Thomas Klebel[1] iD and Tony Ross-Hellauer[1,2] iD

[1]Open and Reproducible Research Group, Know-Center GmbH, Graz, Austria
[2]Open and Reproducible Research Group, Graz University of Technology, Graz, Austria

ABSTRACT

Current implementations of Open Access (OA) publishing frequently involve article processing charges (APCs). Increasing evidence has emerged that APCs impede researchers with fewer resources in publishing their research as OA. We analyzed 1.5 million scientific articles from journals listed in the Directory of Open Access Journals to assess average APCs and their determinants for a comprehensive set of journal publications across scientific disciplines, world regions, and through time. Levels of APCs were strongly stratified by scientific fields and the institutions' countries, corroborating previous findings on publishing cultures and the impact of mandates of research funders. After controlling for country and scientific field with a multilevel mixture model, however, we found small to moderate effects of levels of institutional resourcing on the level of APCs. The effects were largest in countries with low GDP, suggesting decreasing marginal effects of institutional resources when general levels of funding are high. Our findings provide further evidence on how APCs stratify OA publishing and highlight the need for alternative publishing models.

1. INTRODUCTION

Science is central in today's knowledge societies (Stehr, 1994), and is seen as essential to innovation and economic prosperity (Miao, Murray et al., 2022). Yet, science is not conducted in a vacuum. Who performs science has implications on what is studied. For example, diseases and conditions primarily affecting women are strongly understudied in the medical sciences (e.g., Beery & Zucker, 2011; Young, Fisher, & Kirkman, 2019), due to male researchers performing less research directed at women. Similar trends can be found for other factors, such as race and ethnicity (e.g., Deardorff, Hoyt et al., 2019; Turner, Steinberg et al., 2022). For science to meet humanity's needs, scientific research should incorporate knowledge from a diverse range of academics (see also Naik, Sugimoto et al. (2022))[1].

A growing body of literature examines how knowledge generation is structured globally, and highlights how scholarship from outside of the Global North is often deemed less relevant or credible (Albornoz, Okune, & Chan, 2020; Collyer, 2018), or overlooked (Gomez, Herman, & Parigi, 2022). The open access (OA) movement initially raised hopes of leveling the playing

---

[1] This is reflected, for example, in the 2021 UNESCO Recommendations on Open Science, calling for Open Science to "embrace a diversity of knowledge, practices, workflows, languages, research outputs and research topics that support the needs and epistemic pluralism of the scientific community as a whole" (UNESCO, 2021).

field somewhat (Chan, Cuplinskas et al., 2002), by removing the barriers to accessing knowledge posed by journal subscriptions (Matheka, Nderitu et al., 2014). Although more and more research is becoming freely available (Piwowar, Priem et al., 2018), recent research shows that one specific model of OA, publisher-hosted (gold) OA, funded by author-facing charges (article processing charges, or APCs), is erecting a new barrier, preventing authors in less resourced settings (countries, institutions, fields) from contributing to the scientific record (Albornoz et al., 2020; Cabrerizo, 2022; Matheka et al., 2014; Olejniczak & Wilson, 2020; Segado-Boj, Martín-Quevedo, & Prieto-Gutiérrez, 2018; Siler, Haustein et al., 2018; Smith, Merz et al., 2021).

This potential new barrier has been examined on the level of individual authors, institutions, countries, and fields. Focusing on the individual level and investigating data from the United States, Olejniczak and Wilson (2020) found a higher likelihood for publishing OA articles that involve an APC for authors of male gender, from prestigious institutions, with previous federal research funding, or an association with a STEM field. They conclude that "[p]articipation in APC OA publishing appears to be skewed toward scholars with greater access to resources and job security." Along similar lines, Niles, Schimanski et al. (2020) reported that in a survey of Canadian and U.S. academics, publication costs were of higher relevance for women than men in making publishing decisions. The role of institutional support in covering APCs is evidently of urgency for researchers without an affiliation to a research-oriented institution. High APCs and APCs in general might preclude this growing segment of researchers from contributing to the scientific record (Burchardt, 2014; ElSabry, 2017; Gray, 2020, p. 1673).

Investigating the association between institutional characteristics and publishing outcomes, Siler et al. (2018) found a clear hierarchy in publishing access outcomes. Analyzing a set of articles from health research, they found that authors from lower ranked and presumably less wealthy institutions were more likely to publish in toll-access journals, as well as in OA journals with no APC. A key driver for the move towards OA publishing is institutional policies, with Huang, Neylon et al. (2020) finding clear signals for mandates to increase levels of OA publishing. Furthermore, the authors found publisher-mediated OA to be more common in Latin American and African universities, which they attribute to the publishing infrastructure in Latin America (e.g., SciELO) and funder mandates in Africa.

These differences on individual and institutional levels are complemented by inequalities of access to scientific publishing at the level of countries and regions. Researchers from the Global South have more difficulties in paying increasingly high APCs simply due to lower purchasing power parity (Demeter & Istratii, 2020). Waivers for APCs do exist but are not always effective in countering this issue (Burchardt, 2014; Lawson, 2015; cf. Momeni, Dietze et al., 2022). Investigating the geographic diversity of authors across 37,000 articles from Elsevier's "mirror-journal" system, Smith et al. (2021) found a lower geographic diversity of authors for OA articles, and in particular, articles requiring an APC, than for non-OA articles. The authors conclude that their results provide support for the hypothesis that APCs "are a barrier to OA publication by scientists from the low-income countries of the Global South."

In assessing differences in APCs across contexts, scientific fields are an important mediating factor. Studying average APC amounts for Gold and hybrid OA publishing, Björk and Solomon (Björk & Solomon, 2015; Solomon & Björk, 2012b) found higher average APCs journals in STEM, with substantially lower APCs in the Social Sciences and Humanities. These trends might be partially associated with much higher external project funding in STEM than SSH disciplines (Eve, 2014).

The discrepancy in access to publishing is linked to the broader system of knowledge production and its global distribution. Research from the Global North is often focused on phenomena and viewpoints that are relevant to those countries. Research on issues relevant to other regions or communities is commonly not deemed relevant by the Global North and subsequently not accepted for publication in prestigious international journals. In the Global South on the other hand, there is a strong focus on publishing "internationally," that is, in journals published in the Global North (Collyer, 2018; Czerniewicz, 2015). Publications in highly prestigious journals are even sometimes rewarded directly in terms of cash payments (although this practice was abolished in 2020 in China [Mallapaty, 2020]), and are uniformly incentivized indirectly through higher chances to receive promotions (Czerniewicz, 2015). This leads to a situation in which, for researchers from the Global South to publish Open Access in highly regarded journals, they not only have to align their research with that of the North's agenda, but also to pay even higher APCs, as perceived journal prestige (represented by common measures such as the Impact Factor or the DOAJ SEAL) and levels of APCs are linked (Demeter & Istratii, 2020; Gray, 2020; Maddi & Sapinho, 2022; Siler & Frenken, 2019). In economic terms, research money from low-income countries (LIC) partly subsidizes the most prestigious publishing outlets, with researchers from less industrialized countries publishing considerably more frequently in megajournals such as *PLOS ONE* than in the publisher's more prestigious counterparts like *PLOS Biology* (Ellers, Crowther, & Harvey, 2017).

Finally, these tendencies might lead research published in local journals to become less visible. As high-income countries enforce policies to publish OA, research from LIC which might not yet be OA becomes even less visible (Albornoz, Huang et al., 2018; Czerniewicz, 2015). As local journals also usually have lower rankings on common metrics such as the journal impact factor, research published in these journals not only receives less exposure but might be perceived as to be of lesser quality (Gray, 2020).

Given initial evidence that the OA model involving APCs seems to be erecting a new barrier for prospective authors, this paper extends and corroborates previous research in analyzing average APCs and their determinants for a comprehensive set of journal publications across scientific disciplines and world regions, and through time. We pay special attention to the potential effect of institutional resources on APCs and their variation across contexts. In doing so, we provide important evidence to the discussion of how APCs shape publishing outcomes, which we hope will contribute to a more equitable implementation of OA publishing in the future. Our analysis suggests that levels of institutional resourcing and average APCs paid by researchers are related, even when controlling for contextual factors such as academic discipline or country. This APC-Barrier highlights the need for alternative publishing models that are inclusive to researchers irrespective of their institution's level of support for APCs. In Section 2, we discuss how the data set was constructed and explain the methodological steps taken throughout. Section 3 introduces the main findings of the paper, combining an analysis based on descriptive statistics with a formal hierarchical model. Section 4 discusses the findings and highlights implications, while acknowledging some limitations inherent to the analysis.

2. METHODS

For this study, we assembled a large bibliographic data set consisting of 1,572,417 publications. These publications represent all publications published between 2009 and 2019 among journals listed in the Directory of Open Access Journals (DOAJ), where first and/or last authors were affiliated with any institution listed in the 2021 CWTS Leiden Ranking.

OpenAlex (Priem, Piwowar, & Orr, 2022) served as the primary source for bibliographic data. After the decommissioning of Microsoft Academic Graph (MAG) (Microsoft, 2021; Sinha, Shen et al., 2015), OpenAlex now incorporates all data previously present in MAG, which was further enriched by adding data from Unpaywall on Open Access publication status and other identifiers. In using OpenAlex, we found identifiers for publication venues (e.g., journals) to be more reliable than in MAG, significantly improving the potential for matching with further data sources. Scheidsteger and Haunschild (2022) compared MAG with OpenAlex in terms of coverage and metadata accuracy and found that OpenAlex is at least as suited for bibliometric analyses as MAG was. We used the OpenAlex snapshot that was released on April 7, 2022, and converted the .json files to .csv files via python code supplied by the nonprofit organization that built OpenAlex[2]. All further descriptive analysis of the data was conducted with Spark, via the R package *sparklyr* (Luraschi, Kuo et al., 2022).

In line with previous research (e.g., Zhang, Wei et al. (2022); but see Butler, Matthias et al. (2022) on a more thorough approach using historical data), we obtained data on APCs via the public data dump from DOAJ, dated June 3, 2022. As of mid-2022, the DOAJ hosted a community-curated list of close to 18,000 Open Access and peer-reviewed journals. The data from DOAJ contains information on whether the journal imposes an APC and its amount in varying currencies. To match data from DOAJ to OpenAlex, we used the linking ISSN table, which we obtained from the ISSN International Centre on June 13, 2022[3]. Starting from the list of journals in DOAJ, containing 17,717 journals, we were able to match 15,640 (88.3%) venues from OpenAlex. To unify the data on APC charges across currencies, we followed three steps: if an amount was specified in U.S. dollars, we kept this record; if multiple currencies were recorded, we preferred the version in U.S. dollars; and if the amount was not provided in U.S. dollars, we converted it, using the exchange rates at June 4, 2022, following Gray (2020). During data pre-processing, we identified a few journals with erroneous values for APCs, which we subsequently corrected.

2.1. Assigning Publications to Institutions and Fields

To assign publications to institutions, we relied on the information provided in OpenAlex. OpenAlex records the authors of each publication, and parses affiliation information to assign authors to institutions. In the case of single authorships, the publication received a weight of "1" towards the institution of the single author. In the case of multiple authors, we used full counting among authors and fractional counting for authors affiliated with multiple institutions. The following example may illustrate the approach: A given publication P has two authors A and B. Author A has one affiliation (a_1), author B has two affiliations (a_2, a_3). The subsequent weights for publication P were as follows: $w_{a1} = 1$, $w_{a2} = 0.5$, $w_{a3} = 0.5$. We restricted our analysis to first and last authors, following the rationale that decisions on venues and publishing models would usually be taken by the senior and/or the authors that contributed the most (Siler et al., 2018). Recent studies have used first and corresponding authors to attribute publications (Simard, Ghiasi et al., 2022; Zhang et al., 2022). As OpenAlex does not contain information on corresponding authors, we used first and last authors, which contribute more to publications than middle authors (Larivière, Desrochers et al., 2016).

To assign publications to fields, we relied on the "concepts" provided with OpenAlex. Similar to MAG, publications are tagged with concepts. The concepts in OpenAlex are identical on the upper two levels of the hierarchy, whereas OpenAlex has a substantially reduced

---

[2] https://gist.github.com/richard-orr/152d828356a7c47ed7e3e22d2253708d.
[3] https://www.issn.org/services/online-services/access-to-issn-l-table/.

number of concepts on the lower levels (Scheidsteger & Haunschild, 2022). For our analysis, we relied on the top-level concepts only, which consist of 19 unique fields. OpenAlex further provides a "score" for how strongly a given publication is linked to a given concept[4]. We constructed an approach to fractional counting similar to that of institutions, but adapting it to account for the uncertainty around tagging works. For each publication, we calculated the weight towards a single concept c as

$$w_c = \frac{score_c}{\sum_{c=1}^{n} score_c}$$

with $score_1$, $score_2$, ..., $score_n$ denoting all top-level concepts assigned to the given publication.

2.2. Country and Institution-Specific Indicators

2.2.1. Country indicators

Most institutions present in OpenAlex are assigned to a country, via data from MAG or ROR (Research Organization Registry)[5]. To enrich the data from OpenAlex on countries with further information, we used data from the World Bank, via the R package *WDI* (Arel-Bundock, 2022). Specifically, we matched the institutions' countries with the general metadata tables to obtain an indicator for world regions ("East Asia & Pacific," "Europe & Central Asia," etc.). Matching was conducted using the two-digit ISO code provided in both data sets. For data on country income (GDP), we used the indicator "NY.GDP.PCAP.KD," which refers to the GDP per capita in 2015 constant U.S. dollars.

2.2.2. Institutional indicators

For data on levels of institutional resourcing, we used the CWTS 2021 Leiden Ranking (Van Eck, 2021), comprising 1,225 universities across 69 countries. We used the indicator $P_{\text{top 10\%}}$, which is defined as "[t]he number [...] of a university's publications that, compared with other publications in the same field and in the same year, belong to the top 10% most frequently cited"[6]. Previous research (Frenken, Heimeriks, & Hoekman, 2017) has emphasized the role of university age and size when it comes to the level of resources available for supporting research activities (through research equipment, graduate programs, libraries, institutional assistance in securing grant funding, etc.). For this reason, we chose the size-dependent indicator $P_{\text{top 10\%}}$ over size-independent alternatives.

A three-step procedure was undertaken to match records from the Leiden Ranking to OpenAlex. In the first step, we normalized university names and matched based on exact similarity. Normalization included converting to lowercase, unifying encodings, removing commas, and replacing "&" with "and." Duplicate names (e.g., two "University of Heidelberg" in OpenAlex—one in Germany, and one in Ohio) were a rare issue, and were resolved by only retaining matched universities where the countries listed in the Leiden Ranking and OpenAlex also matched. In the second step, we manually matched the remaining universities by searching for a given university name coming from the Leiden Ranking in OpenAlex. Common examples of names that could not be matched automatically were the use of different languages ("Technische Universität Berlin" vs. "Technical University of Berlin"; or "Universidade de Lisboa" vs. "University of Lisbon"), different uses of linking words ("the" or "of"), or the use of different name variants, for example, by using abbreviations (ETH Zürich, VU Amsterdam,

---

[4] https://docs.openalex.org/about-the-data/work#concepts.
[5] https://docs.openalex.org/about-the-data/institution.
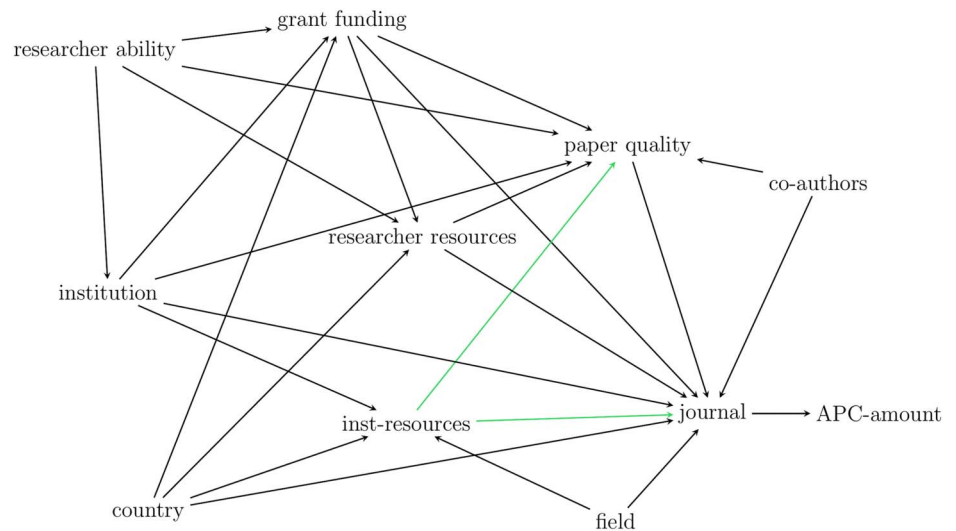[6] https://www.leidenranking.com/information/indicators.

Figure 1. Directed acyclic graph (DAG) of potential causal effects on levels of APCs. The effect of institutional resources ("inst-resources") on the amount of APC charges ("APC-amount") is mediated via the journal. Two main causal paths are marked in green: the direct effect on journal choice, and the indirect effect via paper quality. Our model uses varying slopes for "inst-resources" across countries and fields, and varying intercepts for countries and fields themselves.

KU Leuven, TU Wien, etc.). We used Google searches and Wikipedia to find common and outdated name variants. We further used the links to Wikipedia entries and the institutions themselves, which are provided in OpenAlex, as well as the map provided in the online version of the Leiden Ranking[7]. We were able to match all but one university, resulting in 1,224 (99.9%) matched institutions.

2.3. Hierarchical Modeling

There are many different factors that might lead to differences in publishing outcomes. Figure 1 depicts some of the most salient factors with a directed acyclic graph (DAG). DAGs allow us to represent the causal assumptions of the studied phenomenon visually (Rohrer, 2018). Institutional resources are assumed to contribute to APC levels through two pathways: directly and indirectly via publication quality, where better resources might lead to higher publication quality, which in turn influences where the manuscript might eventually be published. To estimate the total causal effect of institutional resources on APC outcomes, we would need to control for institutions and countries. To estimate the direct causal effect of institutional resources on APC outcomes, we would additionally need to control for publication quality (Pearl, Glymour, & Jewell, 2016). Given that different scientific fields exhibit highly varying publishing cultures (e.g., the different significance of book, journal, and conference publications; varying degrees of acceptance towards preprints, etc.), it is reasonable to assume that relationships might additionally be mediated by scientific fields.

To reduce potential biases in our estimates of the effect of institutional resources on levels of APCs, we constructed a Bayesian multilevel mixture model that controls for scientific field and country. Initially, we planned to also control for institutions. However, our early models

---

[7] For example, https://www.leidenranking.com/Ranking/University2022?universityId=1187&fieldId=1 &periodId=12&fractionalCounting=1&performanceDimension=0&rankingIndicator=pp_top10&minNPubs =100.

suffered from nonidentifiability, because the Leiden Ranking only includes single universities for multiple countries. The mixture components of the model address two particularities of the dependent variable (APC amounts): More than two-thirds of journals in DOAJ charge no APC at all, and this should be incorporated into the model for a comprehensive analysis; and the distribution of APC amounts is multimodal, with a peak for some fields at below U.S. $500, and a main peak for most fields at around U.S. $2,000. We hypothesize that this bimodal distribution stems from differing strategies of publishers and traditions within fields, because the extent to which APCs exhibit the bimodal tendency varies by field (Figure S7, Supplementary material).

Our modeling approach started with a hurdle model, combining a logistic regression for the question of whether a given article had an APC or not, and a lognormal model for the APC amount (conceptually similar to the analysis of Olejniczak and Wilson [2020]). As evidenced by posterior predictive checks on the overall distribution of APC amounts, as well as when making predictions for particular countries, the model did not fit the data well, due to the bimodal distribution of nonzero values (Figures S8 and S9, Supplementary material). For this reason, the model presented in this paper combines two hurdle models in one. The weight given to the two model components is estimated alongside the other parameters, and modeled with respect to the scientific field. Employing multilevel modeling allows us to estimate slopes and intercepts even for countries with only a few universities, by partially pooling information from across the whole data set (Gelman & Hill, 2009; McElreath, 2020). Although these estimates might be more variable, we prefer including all data in the model, as rules for excluding countries based on the number of universities or publications are bound to be arbitrary. Additionally, the exclusion of smaller countries would bias results towards effects present in larger institutions. Further details on the model, including choice of priors and strategies to counter computational difficulties in fitting the model are provided in the supplement.

Because modeling the full data set with Bayesian inference was not feasible, we randomly sampled 8% of articles from the full data set for 2016–2019, which led to a sample size of 76,447. Given its size and the random sampling procedure, the sample is representative of the whole data set.

The model uses log-transformations for both the dependent variable (APC amount) and the independent variable ($P_{\text{top } 10\%}$), which is stratified by field and country. Given the model's complexity in having two hurdle components and two lognormal components, standard measures of directly interpreting coefficients to yield "marginal effects" were not applicable. Equally, we did not deem commonly employed average marginal effects across the whole sample to be informative, given the large share of publications coming from just three countries (China, United States, Brazil). Instead, we constructed comparable effect sizes across the range of $P_{\text{top } 10\%}$ by making predictions at the 20%, 50%, and 80% quantile (*prediction A*), retaining the predicted draws from the posterior distribution. We then made predictions from the model for the value of $P_{\text{top } 10\%}$ at the three cut-offs raised by 1% (*prediction B*) and calculated the ratio as $\beta = \frac{prediction\ B}{prediction\ A}$. This yields values that can readily be interpreted in the standard interpretation of "log-log models," where a 1% change in the independent variable (here: $P_{\text{top } 10\%}$) leads to a change of β% in the dependent variable (here: APC value). The approach can be understood as representing marginal effects at representative values. The Bayesian nature of the model allowed us to construct compatibility regions by visualizing the density of the computed ratios.

To analyze effects for fields, we predicted the ratios for all fields and countries, and then averaged over the effect across countries. In this sense, the effects displayed below are not average marginal effects across the whole sample but average marginal effects at

representative values, weighting each country equally. To analyze effects for countries, we proceeded similarly, averaging over fields for the predictions of each country. This approach combines effects across all model components (i.e., the modeled process for zero and nonzero APCs, as well as the two components predicting the actual size of nonzero APCs).

2.4. Description of the Data

The full data set consists of 1,572,417 publications. The most prevalent field in our data is "Medicine," with the share of fractional weights reaching 30.6% (Table 1, Figure S1). Contrary to the general distribution of fields in MAG and OpenAlex, the second most common field is "Biology" (18.5%). The social sciences are less prevalent in the overall sample, with, for example, publications assigned to "History" amounting to 0.2% of all publications. The high prevalence of "Medicine" and "Biology" and low prevalence of other disciplines can be attributed to two main reasons: First, the general distribution of publications across fields in OpenAlex, and second, the prevalence of certain fields for journals listed in the DOAJ.

Similarly, the proportion of publications assigned to countries and world regions is driven by two processes: first, the general distribution of countries across publications, and second,

Table 1. Fractional publications by field

Field	Fractional papers	Proportion
Medicine	481,047	30.6%
Biology	290,499	18.5%
Chemistry	163,767	10.4%
Computer science	140,438	8.9%
Materials science	106,550	6.8%
Psychology	98,461	6.3%
Environmental science	46,768	3.0%
Physics	44,382	2.8%
Political science	36,339	2.3%
Geography	31,787	2.0%
Sociology	29,340	1.9%
Mathematics	24,093	1.5%
Art	21,573	1.4%
Business	20,992	1.3%
Geology	13,937	0.9%
Philosophy	10,596	0.7%
Economics	5,709	0.4%
History	3,379	0.2%
Engineering	2,760	0.2%

the number of institutions per country that are part of the Leiden Ranking. The number of universities in the Leiden Ranking follows a general division of research productivity, with countries such as China, the United States, Japan, and Germany having many universities in the ranking, and smaller countries or countries with smaller footprints in the international publishing landscape (such as Algeria, Luxembourg, Kuwait, Uganda, and Estonia) only having single universities in the 2019 edition of the ranking.

The distribution of ranked universities across countries also broadly aligns with the overall number of outputs produced in certain countries and world regions. Figure S2, Supplementary material, displays the frequencies and counts across continents. Overall, the largest share of publications in our data set comes from universities in East Asia & Pacific (31.1%), followed by Europe & Central Asia (30.1%), North America (21.2%), and Latin America & Caribbean (12.1%). There are few publications from the Middle East & North Africa (3.0%), South Asia (1.2%), and Sub-Saharan Africa (1.2%).

3. RESULTS

3.1. Descriptive Findings

Taking a high-level view of the data, we find a moderate relationship between levels of institutional resourcing and average levels of APCs for the period 2016–2019 (Figure 2A). Assigning publications to institutions by first and last author does not change the relationship. Given the highly skewed nature of $P_{\text{top 10\%}}$, we show a log-linear relationship. In conceptual terms, this means that a one-unit increase at lower levels of $P_{\text{top 10\%}}$ is considered to be more relevant than at high levels. To investigate how the association develops over time, we analyze the mean APC amounts of the journals for the quartiles of the $P_{\text{top 10\%}}$ distribution (Figure 2B). In line with the cross-sectional view of Figure 1, we find that levels of institutional resourcing are associated with the average levels of APCs of the journals in which the institutions' researchers publish. In particular, the highest quartile (the top 25% universities according to $P_{\text{top 10\%}}$) exhibits a substantially higher mean APC than all other quartiles. The stratification between the quartiles does not change substantially over the observed period, with a slight decrease in the distance between quartiles, and thus a slight reduction in inequality in terms of the APC amount. Given that we use fixed values for APCs across the whole time period, the upward trend most likely represents a shift in publishing patterns and is not driven by an increase in APCs.

3.1.1. Comparing scientific fields

When breaking down the association between institutional resources and levels of APCs across fields, we find that the general pattern holds (Figure 3A): authors from higher ranked institutions publish on average in journals with higher APCs. The strength of the association differs between fields, and there is substantial nonlinearity. It should be noted that there is much less data in smaller fields such as "Philosophy." The estimate of the observed trends is therefore more variable than in fields such as "Biology" and "Medicine."

The differences in changes along $P_{\text{top 10\%}}$ must also be understood in terms of differences in overall levels of APCs, which differ substantially between fields (Figure 2B). Particularly high APCs can be found for journals publishing research in "Biology" (average APC: U.S. $2,118), followed by "Chemistry" (U.S. $1,824) and "Medicine" (U.S. $1,813). At the other end of the spectrum are journals publishing research in the social sciences and humanities, with average APCs of U.S. $59 in "Art," U.S. $71 in "Philosophy," and U.S. $170 in "Sociology." The particularly low average APCs for these fields are to a large extent driven by a high share of publications in journals with no APC.
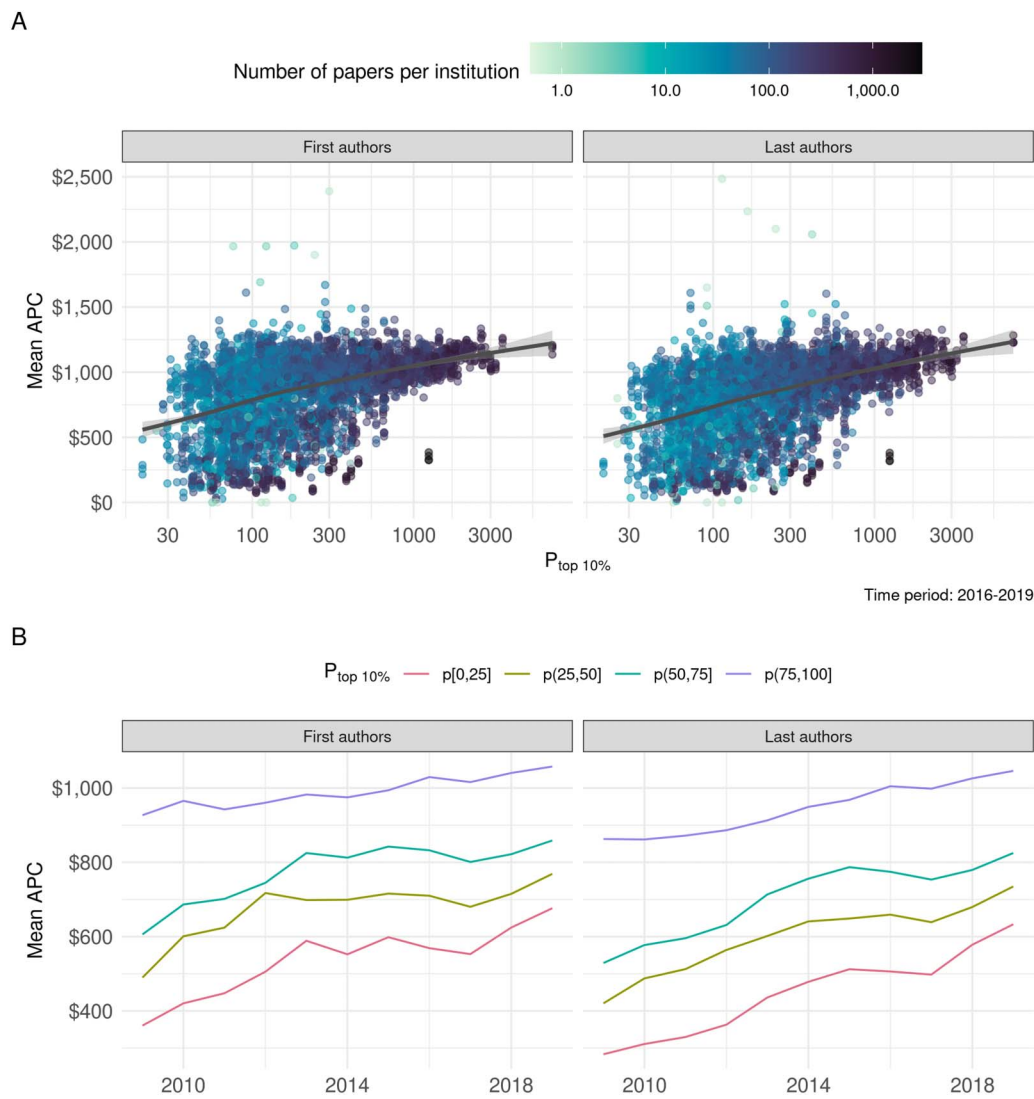
A

B

Figure 2. Association between institutional resources ($P_{top\ 10\%}$) and average APC. The APC is derived from the APC listed for each journal in the DOAJ. Journals with no APC are included with an APC of U.S. $0. (A) displays the association for the time period 2016–2019. The x-axis is on a $\log_{10}$-scale—the displayed relationship is thus nonlinear. The line represents a general additive model fit with cubic splines (default settings from geom_smooth() from the ggplot2 R package). (B) displays the association over time, with $P_{top\ 10\%}$ broken down into quartiles. We use fixed values for APCs across the whole time period.

Analyzing the association across fields over time (Figures S4 and S5, Supplementary material) we observe heterogeneous trends. Although in some fields ("Biology" and "Chemistry") average APCs have been rising from 2009 to 2019, other fields exhibit stable levels of APCs (e.g., "Mathematics," "Geography," "Geology") or downwards trends ("Physics"). Considering the stratification within fields (as evidenced by the spread between quartiles), no clear pattern is discernible. The data suggest a narrowing of the gap between lowest and highest ranked institutions for "Biology" and "Chemistry," and a potential slight increase in stratification for research in "Computer science" and "Sociology." Given that our analysis uses static values for APCs per journal, this most likely represents a shift in where researchers tend to publish (e.g., in journals with or without APCs, or with high or low APCs).
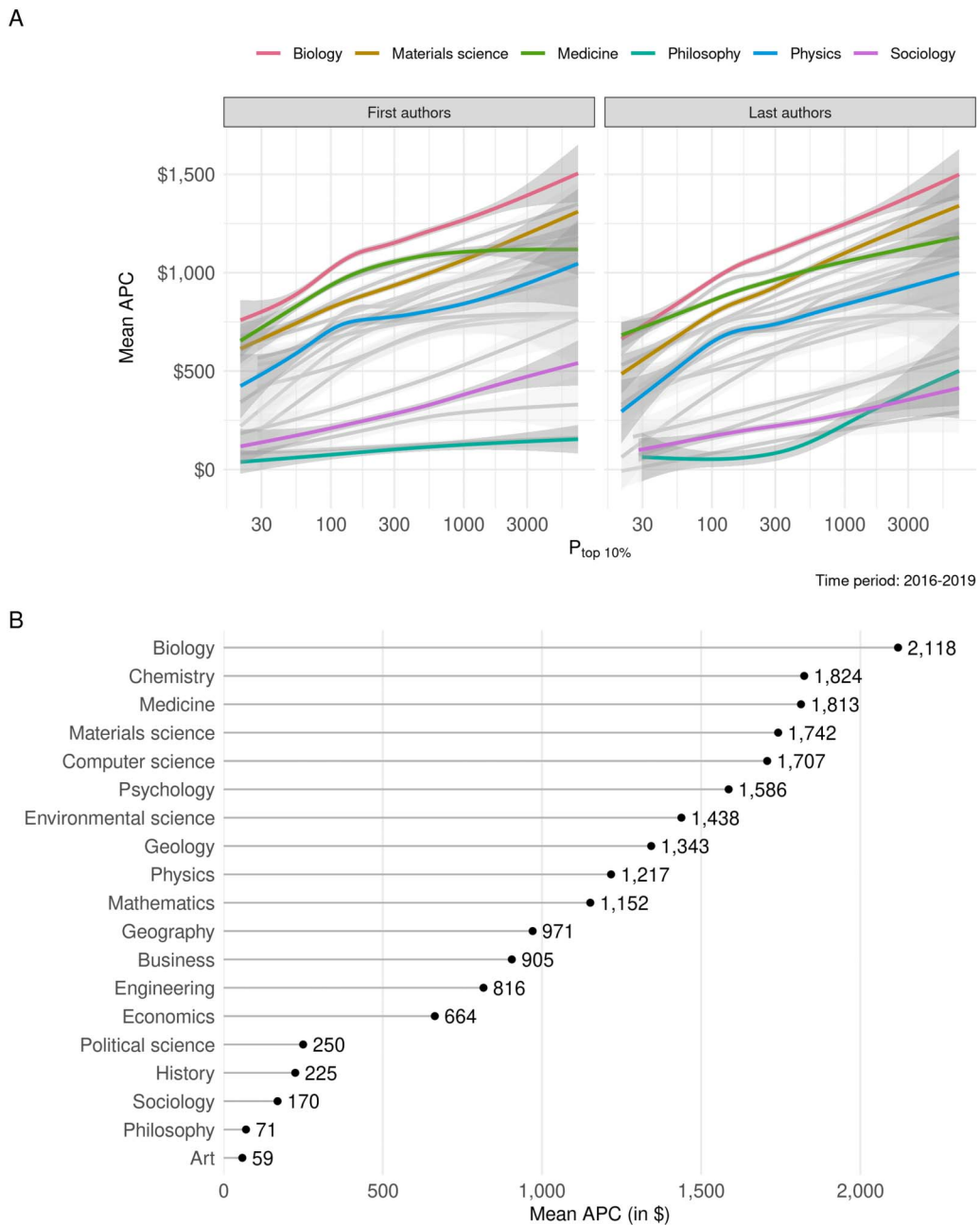
A

B

Figure 3. APC by field. (A) shows the association between APC and $P_{top~10\%}$ per field for the papers published in 2016–2019. Journals with no APC are included with an APC of U.S. $0. To aid interpretation, only selected fields are displayed in color, with the remaining fields plotted in grey. An interactive version of the plot is available in the notebooks published with the code. (B) displays the average APC in the respective fields (2009–2019). The average APC is obtained by multiplying the fractional contribution of each paper towards all fields with the APC of the journal in which it is published and dividing by the sum of fractional contributions within each field. Journals with no APC are included with an APC of U.S. $0.

3.1.2. Comparing countries

Contrasting average APCs between countries, we observe substantial variation (Table S1, Supplementary material). We find the highest average APCs for researchers at institutions in Israel, Switzerland, and Singapore, with average APCs of about U.S. $2,200. In contrast, the
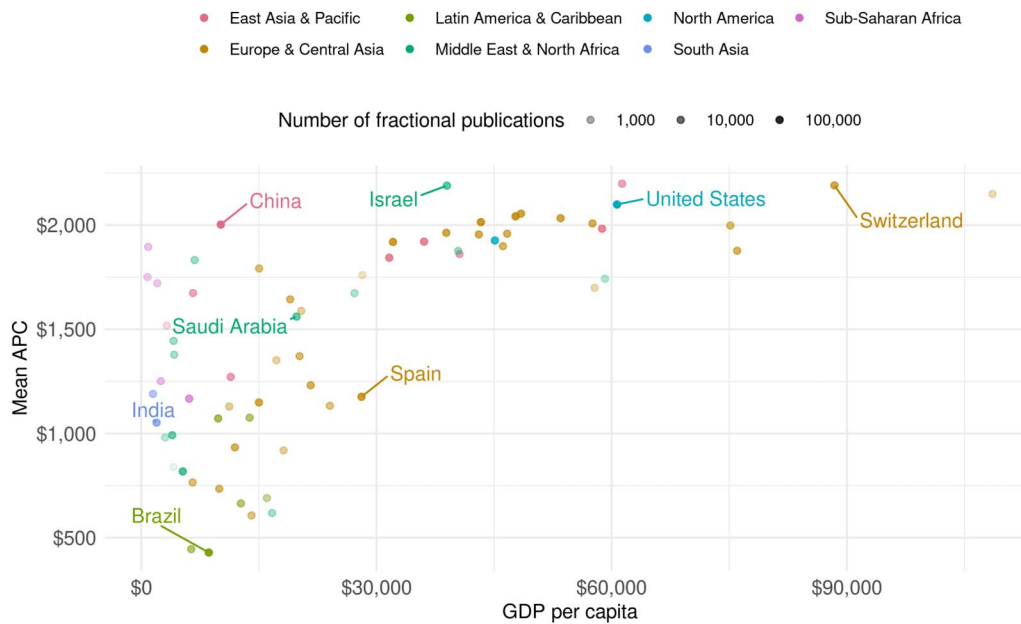
Figure 4. Mean APC per country vs. GDP per capita. APCs are averaged across all years (2009–2016), GDP is expressed in 2015 constant U.S. dollars. Journals with no APC are included with an APC of U.S. $0.

lowest average APCs can be observed for researchers at institutions in Colombia and Brazil, with average APCs below U.S. $500. To further explore the variation, we compare the average APC per country with the country's GDP per capita (Figure 4). We observe high variation in average APCs for authors from countries with low to medium GDP per capita (< U.S. $30,000), ranging from U.S. $429 in Brazil to U.S. $2,002 in China. In contrast, the average APC for authors from countries with a GDP per capita above U.S. $30,000 consistently ranges from U.S. $1,700 to U.S. $2,200. Within the cluster of lower income countries, two key observations can be made. First, the average APC for authors from Latin America and the Caribbean is consistently low, with the highest average APC among these countries in Mexico (U.S. $701). The low levels of APC in these countries likely are a result of local publishing cultures and infrastructure (e.g., SciELO), and potentially the emergence of local journals with low APCs. Second, in contrast, the average APC for authors from Sub-Saharan Africa is relatively high, ranging from U.S. $1,167 to U.S. $1,895. Given that for many countries only a few universities are listed in the Leiden Ranking, we recreated Figure 3 among all institutions present in OpenAlex (Figure S3, Supplementary material). Average APCs across countries are slightly lower, but the association between GDP per capita and average APCs is unchanged.

The relatively high rates of average APCs in countries from Sub-Saharan Africa likely reflect the strong influence of research funding towards these countries, and subsequent mandates to publish OA, which has previously been suggested by Iyandemye and Thomas (2019). An exploratory analysis of the prevalence of field-specific publications across continents indeed reveals that the share of publications in "Medicine" from Sub-Saharan African countries is very high (42.8%), which lends support to this hypothesis. The rate of publications in "Medicine" is even higher in South Asia, whereas the average APC in South Asia is substantially lower. Here, we would suspect that these publications are to a lesser extent driven by third-party funding. The high average APC for China likely reflects its rise to one of the leading nations in science (Gomez et al., 2022; Xie, Zhang, & Lai, 2014; Zhou & Leydesdorff, 2006).
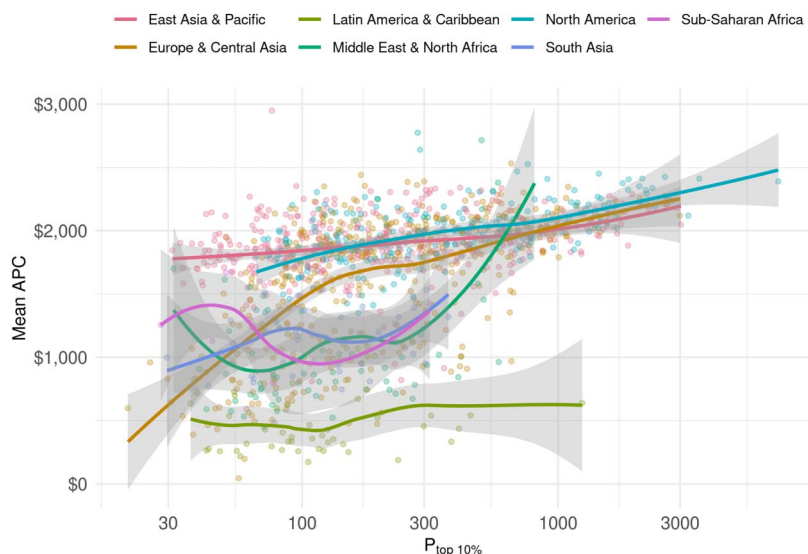
Figure 5. Association between institutional resources and mean APC across world regions. Journals with no APC are included with an APC of U.S. $0.

A purely descriptive analysis of associations between institutional ranking and average APC across countries is not possible, given that some countries have only a few ranked institutions. Figure 5 therefore displays the association broken down by continent, and we estimate individual country effects with a hierarchical mixture model in the next section. The descriptive analysis reveals that there are large disparities in terms of overall levels of APCs. The relationship between institutional ranking and mean APC is strongest in Europe & Central Asia, although the steep slope for low regions of $P_{\text{top } 10\%}$ (from 30 to 100) should be interpreted with caution due to few data points in this region. The trends for all other continents are much more variable, also due to a low amount of data (few universities per continent) and we therefore do not interpret their slopes.

3.2. Modeling the Effect of Institutional Resources Across Fields and Countries

To separate the effect of $P_{\text{top } 10\%}$ on levels of APCs from country and field effects, we used a Bayesian multilevel mixture model. Figure 6 displays the effect of a 1% increase in $P_{\text{top } 10\%}$ on the level of APCs. For all fields except Mathematics, Physics, and Art, higher institutional resources are associated with higher APCs. For most fields, the effect is nonlinear in that it is more pronounced at lower levels of institutional resources than at higher levels. This might suggest support for the hypothesis that institutional resourcing influences submission choices of authors by enlarging or restricting the space of potential venues due to economic reasons. The effect of institutional resources on APCs is strongest in fields from the social sciences ("Political science," "Sociology," "Business"). Estimates for the arts and humanities ("Art," "History," "Philosophy") are not uniform, but exhibit wide credibility regions. The wide intervals are a consequence of these fields exhibiting high rates of zero-APC publishing, which result in a low number of cases to estimate the nonzero component of the model.

Effects in the natural and life sciences ("Biology," "Medicine," "Materials science," "Chemistry") are estimated to be low, with narrow credibility intervals. The negative effect of institutional resources on APC amounts in "Mathematics" is mainly driven by the hurdle component of the model (i.e., authors from higher ranked institutions publish more frequently in journals with no APC at all in these fields, compared to the average of all fields). In contrast, for
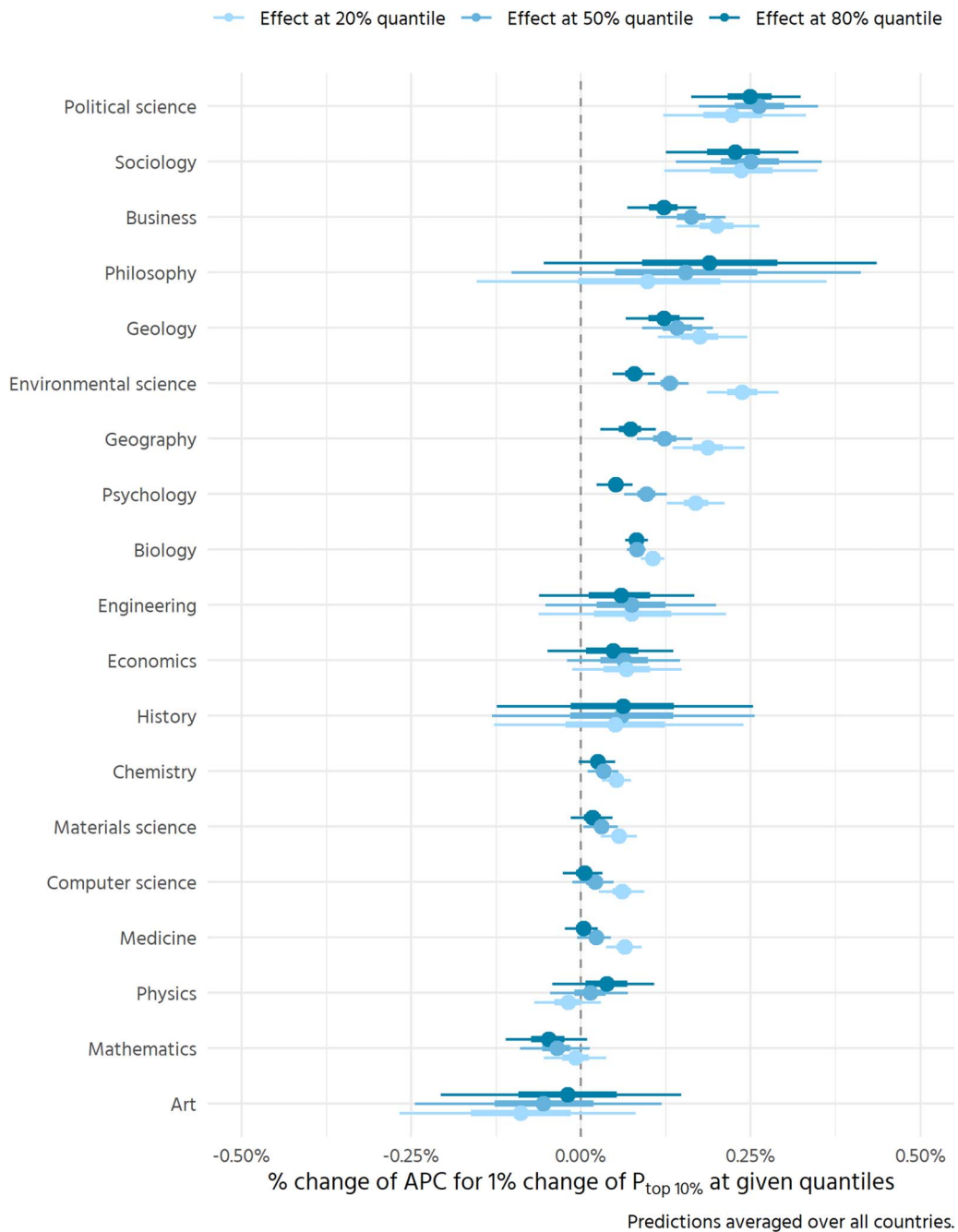
Figure 6. Effect of an increase in $P_{top\ 10\%}$ on the level of APC across fields. The dot represents the median effect averaged over the predicted effects across countries, with the lines representing 50% and 90% highest density intervals.

research published in "Environmental science" and "Sociology," authors from higher ranked institutions tend to publish in journals with no APC *less* than their peers from lower ranked institutions.

Comparing the effect of institutional resources on the level of APCs across countries, we find small to moderate effects, with substantial heterogeneity in the estimates and their variability. Figures S12 and S13 (Supplementary material) show the estimates split by continent,
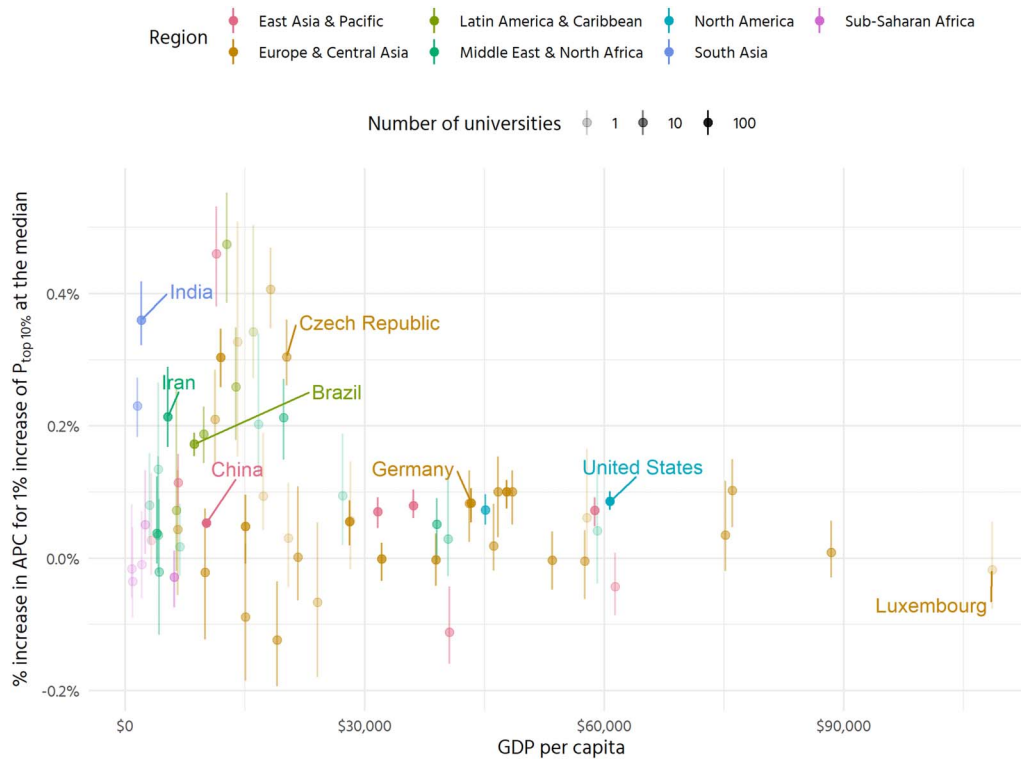
Figure 7. *Post hoc* analysis of the effect of $P_{\text{top 10\%}}$ on APC compared with country GDP, estimated at the median of $P_{\text{top 10\%}}$ and averaged across all fields. Error bars represent the 50% highest density interval.

and by low, middle, and high levels of $P_{\text{top 10\%}}$ (as above). Almost all continents show a spread between countries with moderate negative effects versus those with moderate to large positive effects (e.g., New Zealand–Malaysia, Greece–Slovakia, Tunisia–Iran, Uganda–Nigeria). Analyzing the effects across countries, we hypothesized that the differences in effect sizes might be related to overall levels of wealth. We therefore conducted a *post hoc* analysis, plotting the country estimates along the countries' GDP per capita (Figure 6). This analysis indeed suggests that the effect of institutional resources tends to be stronger in countries with low levels of GDP per capita. Comparing the effect of institutional resources on APCs across countries (Figure 7) with overall levels of APCs per country (Figure 3) suggests a threshold effect: Institutional resources have a small effect on levels of APCs in countries where the overall APC level is high. This might be explained by general levels of resourcing. Alternatively, it could also point to country-specific policies on OA publishing. An additional observation of note is the low estimates for the effect of levels of institutional resourcing on levels of APCs among Sub-Saharan African countries.

4. DISCUSSION

Open Science holds the promise to make research processes more transparent, efficient, rigorous, and inclusive. Yet, current incarnations of OA publishing seem to partly contradict these goals. In this study we investigated what we term the "APC-barrier," finding that higher institutional resourcing is associated with researchers publishing in journals with higher APCs. This linkage is nonlinear and heterogeneous across fields. Although our study has limitations regarding the measurements used, our findings suggest support for the hypothesis that author-facing charges in OA publishing present a barrier to publication and reduce the pool

of knowledge that enters the scientific record. Our results extend and corroborate previous research (Olejniczak & Wilson, 2020; Siler et al., 2018; Smith et al., 2021) on potential factors that influence who is able to publish where.

At a global level, we find substantial heterogeneity in average levels of APCs across fields and countries. We observe high average APCs of above U.S. $1,000 for the natural and life sciences, and low average APCs of up to U.S. $250 for the social sciences and humanities (with the exception of "Economics" and "Business"). Grouping institutions by country, we find a clear economic divide, with high average APCs for countries with a GDP per capita above U.S. $30,000, and very heterogeneous levels of average APCs for less wealthy countries. This heterogeneity can be partly attributed to the effects of policies and alternative publishing models in Latin America (Huang et al., 2020), and targeted research funding in Sub-Saharan Africa (Iyandemye & Thomas, 2019). The case of Sub-Saharan African countries is particularly interesting, as our model's estimates of the effect of institutional resources on levels of APCs are close to zero here. This might further point towards the local importance of third-party funding in driving APC-based OA uptake, as opposed to institutional resources, which are seemingly less important.

There are multiple processes at play that lead to the observed data on the distribution of APCs across fields and countries. Following Olejniczak and Wilson (2020), we jointly modeled the questions of whether a given publication involved an APC or not, and if yes, its magnitude. Our results indicate that this distinction is of relevance, given that for some fields (e.g., "Mathematics") higher levels of institutional resourcing are associated with higher rates of zero APCs, but it is the opposite for other fields (e.g., "Environmental science," "Sociology"). Our assumption is that institutional resources contribute to covering APCs in at least two ways: first through direct funding of APCs, and second through transformative agreements (Borrego, Anglada, & Abadal, 2021), where institutions make deals with major publishers to cover APCs. Direct funding of APCs, if not covered by transformative agreements, is often granted through institutional publishing funds, which commonly also include a cap on the maximum APC that is covered (Click & Borchardt, 2019; Solomon & Björk, 2012a). It can be assumed that such funds are more common among higher ranking institutions with greater resourcing.

Considering the assumed causal pathways depicted in Figure 1, our modeling approach is able to account for some sources of confounding (country and field effects), while other potential confounders have not been incorporated. One important alternative explanation for our results would be the causal path institutional resources → research quality → journal → APC amount. It can be assumed that institutional resources have some effect on research quality (through better infrastructure, higher attractiveness for coauthorships, etc.). Given that there is a moderate link between the perceived quality of journals and the levels of APCs they charge (Demeter & Istratii, 2020; Gray, 2020; Maddi & Sapinho, 2022), this path could account for some of the effects we measure. However, two observations suggest that this might not be a particularly severe issue. First, the correlation between journal prestige (measured via the Impact Factor (IF)) and APC is only moderate, and further, the IF itself is highly debated as a measure for quality (Archambault & Larivière, 2009; Bar-Ilan, 2008; Larivière & Sugimoto, 2019; Lozano, Larivière, & Gingras, 2012; Waltman, 2016; Waltman & Traag, 2020). Second, our estimates of the effect of institutional resources on levels of APCs are highest for the fields "Political science" and "Sociology," which we argue are less resource dependent than other fields with low estimates (e.g., "Physics," "Medicine").

As pointed out by one of our reviewers, it is possible that $P_{\text{top 10\%}}$ actually measures the quality of an institution rather than its resources. Our results would therefore provide an

estimate of the effect of quality → journal → APC amount. Although this is certainly possible, we assume $P_{\text{top 10\%}}$ to be more indicative of institutional resources than of their inherent quality, because it is highly correlated with overall publication output (denoted as "P" in the Leiden Ranking). This points to a crucial issue in studying the effect of institutional resources on OA publishing, which is a lack of more specific data on universities' support budgets for APCs. Further support for initiatives like OpenAPC[8] would enable this.

A further alternative explanation might be that more prestigious journals with high APCs reject research from less prestigious institutions, simply because it is deemed less credible (Albornoz et al., 2020; Collyer, 2018). However, given that the negative effect of APCs on the geographic diversity of authors is substantial (Smith et al., 2021), we don't expect this to be a major source of bias.

Our analysis points to threshold effects when it comes to the effect of institutional resourcing on levels of APCs across countries and fields. In most fields, and particularly those with more observations, the effect of the ranking position on levels of APCs is higher for lower ranking levels. This suggests that resources do make a difference: Once institutions reach a certain level of resourcing, they seem to be able to cover common APCs. Similarly, lower GDP is associated with a stronger effect of institutional resources on APCs in most cases (with the exceptions of Sub-Saharan Africa and, most notably, China). This again suggests that levels of resourcing play a role and points to a threshold effect: In medium- to high-income countries, most institutions can be assumed to be able to support APCs. Above this threshold, ranking differences are only weakly related to levels of APCs. In lower income countries, institutional differences are larger, and structured along the dimension of resourcing.

The observed forces clearly perpetuate the system of cumulative advantage inherent to academia, as well-funded research groups are better able to secure OA publications in prestigious journals with high APCs, leading to citation advantages and further funding down the line. We believe that this demonstrates the impact of APC pricing on the scholarly landscape and that these charges may have a chilling effect on opportunity and equality for researchers from less prestigious or less wealthy institutions. Such stratifications in publishing, favoring traditionally advantaged actors, will only exacerbate historical inequalities (Garuba, 2013) and undermine the wider aims of Open Science (Ross-Hellauer, 2022; Ross-Hellauer, Reichmann et al., 2022). If research is to live up to present and future challenges, it should seek to avoid modes of scholarly publication that exacerbate the marginalization of voices from societies and communities less embedded in the global production of knowledge.

Waivers for APCs do exist, but are, in our estimation, ineffective in countering these issues. Waivers are only applied on request, yet such discount policies are not well communicated (i.e., authors are often unaware). Hybrid journals do not usually offer waivers for OA in their journals, and where waivers are in place, often do not mitigate costs enough to encourage OA authorship (Lawson, 2015; Mekonnen, Downs et al., 2022; Rouhi, Beard, & Brundy, 2022). Most damagingly, however, in our view waivers support rather than challenge the status quo. They are a bandage on the structural inequalities exposed by a system of payment for authorship, not addressing the underlying issues but rather putting the burden on already disadvantaged scholars to appeal for assistance via poorly documented, poorly communicated, *ad hoc* processes whose conditions of application are apt to change.

To take steps towards more equitable publishing models, a range of recommendations have been developed in the ON-MERRIT project (Cole, Reichmann, & Ross-Hellauer, 2022) and

---

[8] https://openapc.net/.

beyond (UNESCO, 2021). Critically important are alternative publishing models that involve no author-facing charges at all (e.g., Diamond OA). Open and sustainable publishing infrastructures should be supported by researchers, institutions, and funders, laying the groundwork of reduced publishing costs and shaping a future where publishing involves no charges, neither for authors nor for readers. In parallel, self-archiving of peer-reviewed works should continue to receive increased attention, given that it can immediately be realized by researchers and institutions. Despite the worrying trends observed in our study, solutions are available that promise to move scholarly communication towards more equity and diversity.

4.1. Limitations and Future Directions

Assembling the data set for this study involved many steps and decisions that could potentially threaten the validity of our results. First, inclusion of universities into the Leiden Ranking can be understood as a marker for resources in itself, as high production of internationally recognized research is a precondition. It is therefore fair to assume that the adverse effect of APC levels on the inclusivity and equity of the scholarly publishing landscape is even stronger for researchers from less resourced institutions. Second, the Leiden Ranking offers a wide range of indicators, and we analyze only one of multiple potential proxy indicators for institutional resources. Third, the definitions of institutions (e.g., which institutes/hospitals/etc. to include) might not completely overlap between OpenAlex and the Leiden Ranking. Additionally, affiliation data from OpenAlex is less complete and reliable for smaller publishers, which might bias our results[9].

Fourth, our analysis uses static values for APCs. Although this is a common approach in the literature (e.g., Zhang et al., 2022), it might introduce uncertainty in the estimates or even bias them, given that journals do in fact change their APCs (Asai, 2020, 2021; Morrison, Salhab et al., 2015). Butler et al. (2022) conducted a similar study to Zhang et al. (2022) but used annual price lists and historical data on APCs to yield more accurate values for actual APCs paid. Their estimate for the revenue of the five largest publishers is hence lower than the estimate of Zhang and colleagues. Because we used the same approach as Zhang et al., it is likely that our analysis also somewhat overestimates the averages for APCs. However, our estimates for the relationship between institutional resources and APCs would then only be biased to the extent that recent price increases differed greatly across disciplines or geographic regions. We leave this as an avenue for future research.

Fifth, we restricted our analysis to publications in fully OA journals, not considering hybrid OA publications. Given that APC charges for hybrid publications are generally higher than for gold OA publications, we assume that the observed trends would be even stronger for OA publications in hybrid journals. Finally, the modeling approach taken to disentangle field and country effects posed computational challenges. We have followed available best practices, but the conclusions should be treated as explorative and be backed up by future research with potentially different approaches.

Our study opens up multiple avenues for further research. Of primary concern should be to find more direct measures of institutional resources. Although we assume that our proxy works well, more direct measures, such as data on institutional support for APCs, general library support on OA, etc., should be considered. Our analysis incorporated the dimension of time for the descriptive results; a more stringent treatment within a suitable model could shed further

---

[9] A point made by Ludo Waltman in personal communication.

light on how the association between institutional resources and levels of APCs changes over time. Lastly, replication attempts using different data sources for bibliographic data (e.g., Web of Science, Scopus, Dimensions) or data on APCs (e.g., OpenAPC) could provide further support for the presented conclusions.

DATA AVAILABILITY

The data sets created in the course of this analysis are available from Zenodo (Klebel & Ross-Hellauer, 2022b). All code to reproduce the creation and analysis of the data is available from Zenodo (Klebel & Ross-Hellauer, 2022a). This includes saved objects from brms, which enables analyzing the model without the need to run the sampler.

REFERENCES

Albornoz, D., Huang, M., Martin, I. M., Mateus, M., Touré, A. Y., & Chan, L. (2018). Framing power: Tracing key discourses in open science policies. In *22nd International Conference on Electronic Publishing - Connecting the Knowledge Commons: From Projects to Sustainable Infrastructure, ELPUB 2018*. Toronto. https://doi.org/10.4000/proceedings.elpub.2018.23

Albornoz, D., Okune, A., & Chan, L. (2020). Can open scholarly practices redress epistemic injustice? In M. P. Eve & J. Gray (Eds.), *Reassembling scholarly communications: Histories, infrastructures, and global politics of open access*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/11885.003.0009

Archambault, É., & Larivière, V. (2009). History of the journal impact factor: Contingencies and consequences. *Scientometrics, 79*(3), 635–649. https://doi.org/10.1007/s11192-007-2036-x

Arel-Bundock, V. (2022). *WDI: World Development Indicators and other World Bank data* (v2.7.6). https://CRAN.R-project.org/package=WDI

Asai, S. (2020). Determinants of revisions to article processing charges for BMC journals. *Publishing Research Quarterly, 36*(1), 63–73. https://doi.org/10.1007/s12109-019-09677-1

Asai, S. (2021). An analysis of revising article processing charges for open access journals between 2018 and 2020. *Learned Publishing, 34*(2), 137–143. https://doi.org/10.1002/leap.1334

Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century—A review. *Journal of Informetrics, 2*(1), 1–52. https://doi.org/10.1016/j.joi.2007.11.001

Beery, A. K., & Zucker, I. (2011). Sex bias in neuroscience and biomedical research. *Neuroscience and Biobehavioral Reviews, 35*(3), 565–572. https://doi.org/10.1016/j.neubiorev.2010.07.002, PubMed: 20620164

Björk, B.-C., & Solomon, D. (2015). Article processing charges in OA journals: Relationship between price and quality. *Scientometrics, 103*(2), 373–385. https://doi.org/10.1007/s11192-015-1556-z

Borrego, Á., Anglada, L., & Abadal, E. (2021). Transformative agreements: Do they pave the way to open access? *Learned Publishing, 34*(2), 216–232. https://doi.org/10.1002/leap.1347

Burchardt, J. (2014). Researchers outside APC-financed Open Access: Implications for scholars without a paying institution. *SAGE Open, 4*(4), 2158244014551714. https://doi.org/10.1177/2158244014551714

Butler, L.-A., Matthias, L., Simard, M.-A., Mongeon, P., & Haustein, S. (2022). The oligopoly's shift to open access. How for-profit publishers benefit from article processing charges. *Zenodo.* https://doi.org/10.5281/zenodo.7057144

Cabrerizo, F. M. (2022). Open access in low-income countries—Open letter on equity. *Nature, 605*(7911), 620. https://doi.org/10.1038/d41586-022-01414-7, PubMed: 35610369

Chan, L., Cuplinskas, D., Eisen, M., Friend, F., Genova, Y., ... Velterop, J. (2002). *Budapest Open Access Initiative.* https://web.archive.org/web/20171213093708/https://www.budapestopenaccessinitiative.org/read

Click, A. B., & Borchardt, R. (2019). Library supported open access funds: Criteria, impact, and viability. *Evidence Based Library and Information Practice, 14*(4), 21–37. https://doi.org/10.18438/eblip29623

Cole, N. L., Reichmann, S., & Ross-Hellauer, T. (2022). Global thinking. ON-MERRIT recommendations for maximising equity in open and responsible research. *Zenodo.* https://doi.org/10.5281/zenodo.6276753

Collyer, F. M. (2018). Global patterns in the publishing of academic knowledge: Global North, Global South. *Current Sociology, 66*(1), 56–73. https://doi.org/10.1177/0011392116680020

Czerniewicz, L. (2015). Opinion: Confronting inequitable power dynamics of global knowledge production and exchange. *Water Wheel, 14*(5), 26–29.

Deardorff, J., Hoyt, L. T., Carter, R., & Shirtcliff, E. A. (2019). Next steps in puberty research: Broadening the lens toward understudied populations. *Journal of Research on Adolescence, 29*(1), 133–154. https://doi.org/10.1111/jora.12402, PubMed: 30869847

Demeter, M., & Istratii, R. (2020). Scrutinising what Open Access journals mean for global inequalities. *Publishing Research Quarterly, 36*, 505–522. https://doi.org/10.1007/s12109-020-09771-9

Ellers, J., Crowther, T. W., & Harvey, J. A. (2017). Gold open access publishing in mega-journals: Developing countries pay the price of western premium academic output. *Journal of Scholarly Publishing, 49*(1), 89–102. https://doi.org/10.3138/jsp.49.1.89

ElSabry, E. (2017). Unaffiliated researchers: A preliminary study. *Challenges, 8*(2), 20. https://doi.org/10.3390/challe8020020

Eve, M. P. (2014). *Open access and the humanities.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316161012

Frenken, K., Heimeriks, G. J., & Hoekman, J. (2017). What drives university research performance? An analysis using the CWTS Leiden Ranking data. *Journal of Informetrics, 11*(3), 859–872. https://doi.org/10.1016/j.joi.2017.06.006

Garuba, A. R. (2013). The prospects of bridging the digital divide in Africa. *Library Philosophy and Practice, 2013*, 882. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085323161&partnerID=40&md5=530601d0b82a6c389c300e61b548f11a

Gelman, A., & Hill, J. (2009). *Data analysis using regression and multilevel/hierarchical models.* Cambridge: Cambridge University Press.

Gomez, C. J., Herman, A. C., & Parigi, P. (2022). Leading countries in global science increasingly receive more citations than other countries doing similar research. *Nature Human Behaviour, 6*(7), 919–929. https://doi.org/10.1038/s41562-022-01351-5, PubMed: 35637294

Gray, R. J. (2020). Sorry, we're open: Golden open-access and inequality in non-human biological sciences. *Scientometrics, 124*(2), 1663–1675. https://doi.org/10.1007/s11192-020-03540-3

Huang, C.-K. (Karl), Neylon, C., Hosking, R., Montgomery, L., Wilson, K. S., ... Brookes-Kenworthy, C. (2020). Evaluating the impact of open access policies on research institutions. *eLife, 9*, e57067. https://doi.org/10.7554/eLife.57067, PubMed: 32924933

Iyandemye, J., & Thomas, M. P. (2019). Low income countries have the highest percentages of open access publication: A systematic computational analysis of the biomedical literature. *PLOS ONE, 14*(7), e0220229. https://doi.org/10.1371/journal.pone.0220229, PubMed: 31356618

Klebel, T., & Ross-Hellauer, T. (2022a). Code for "The APC-Barrier and its effect on stratification in open access publishing". *Zenodo.* https://doi.org/10.5281/zenodo.7198844

Klebel, T., & Ross-Hellauer, T. (2022b). Data for "The APC-Barrier and its effect on stratification in open access publishing" [Data set]. *Zenodo.* https://doi.org/10.5281/zenodo.7014802

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science, 46*(3), 417–435. https://doi.org/10.1177/0306312716650046, PubMed: 28948891

Larivière, V., & Sugimoto, C. R. (2019). The Journal Impact Factor: A brief history, critique, and discussion of adverse effects. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 3–24). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-02511-3_1

Lawson, S. (2015). Fee waivers for open access journals. *Publications, 3*(3), 155–167. https://doi.org/10.3390/publications3030155

Lozano, G. A., Larivière, V., & Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *Journal of the American Society for Information Science and Technology, 63*(11), 2140–2145. https://doi.org/10.1002/asi.22731

Luraschi, J., Kuo, K., Ushey, K., Allaire, J. J., Falaki, H., ... The Apache Software Foundation. (2022). *sparklyr: R Interface to Apache Spark* (v1.7.5). https://CRAN.R-project.org/package=sparklyr

Maddi, A., & Sapinho, D. (2022). Article processing charges, altmetrics and citation impact: Is there an economic rationale?

Scientometrics, *127*, 7351–7368. https://doi.org/10.1007/s11192 -022-04284-y

Mallapaty, S. (2020). China bans cash rewards for publishing papers. *Nature*, *579*(7797), 18. https://doi.org/10.1038/d41586 -020-00574-8, PubMed: 32127708

Matheka, D. M., Nderitu, J., Mutonga, D., Otiti, M. I., Siegel, K., & Demaio, A. R. (2014). Open access: Academic publishing and its implications for knowledge equity in Kenya. *Globalization and Health*, *10*, 26. https://doi.org/10.1186/1744-8603-10-26, PubMed: 24716579

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). Taylor & Francis, CRC Press. https://doi.org/10.1201/9780429029608

Mekonnen, A., Downs, C., Effiom, E. O., Kibaja, M., Lawes, M. J., … Chapman, C. A. (2022). Can I afford to publish? A dilemma for African scholars. *Ecology Letters*, *25*(4), 711–715. https://doi.org /10.1111/ele.13949, PubMed: 34957647

Miao, L., Murray, D., Jung, W.-S., Larivière, V., Sugimoto, C. R., & Ahn, Y.-Y. (2022). The latent structure of global scientific development. *Nature Human Behaviour*, *6*, 1206–1217. https://doi.org /10.1038/s41562-022-01367-x, PubMed: 35654964

Microsoft. (2021). *Next steps for Microsoft Academic—Expanding into new horizons*. Microsoft Research. https://www.microsoft.com/en -us/research/project/academic/articles/microsoft-academic-to -expand-horizons-with-community-driven-approach/

Momeni, F., Dietze, S., Mayr, P., Biesenbender, K., & Peters, I. (2022). Which factors drive open access publishing? A Springer Nature case study. *arXiv*, arXiv:2208.08221. https://doi.org/10 .48550/arXiv.2208.08221

Morrison, H., Salhab, J., Calvé-Genest, A., & Horava, T. (2015). Open access article processing charges: DOAJ survey May 2014. *Publications*, *3*(1), 1–16. https://doi.org/10.3390 /publications3010001

Naik, C., Sugimoto, C. R., Larivière, V., Leng, C., & Guo, W. (2022). Impact of geographic diversity on citation of collaborative research. *arXiv*, arXiv:2204.11713. https://doi.org/10.48550 /arXiv.2204.11713

Niles, M. T., Schimanski, L. A., McKiernan, E. C., & Alperin, J. P. (2020). Why we publish where we do: Faculty publishing values and their relationship to review, promotion and tenure expectations. *PLOS ONE*, *15*(3), e0228914. https://doi.org/10.1371 /journal.pone.0228914, PubMed: 32160238

Olejniczak, A. J., & Wilson, M. J. (2020). Who's writing open access (OA) articles? Characteristics of OA authors at Ph.D.-granting institutions in the United States. *Quantitative Science Studies*, *1*(4), 1429–1450. https://doi.org/10.1162/qss_a_00091

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. Chichester: Wiley.

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., … Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, *6*, e4375. https://doi.org/10.7717/peerj.4375, PubMed: 29456894

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv*, arXiv:2205.01833. https://doi.org/10.48550/arXiv .2205.01833

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42. https://doi.org/10.1177/2515245917745629

Ross-Hellauer, T. (2022). Open science, done wrong, will compound inequities. *Nature*, *603*(7901), 363. https://doi.org/10 .1038/d41586-022-00724-0, PubMed: 35288691

Ross-Hellauer, T., Reichmann, S., Cole, N. L., Fessl, A., Klebel, T., & Pontika, N. (2022). Dynamics of cumulative advantage and threats to equity in open science: A scoping review. *Royal Society Open Science*, *9*(1), 211032. https://doi.org/10.1098 /rsos.211032, PubMed: 35116143

Rouhi, S., Beard, R., & Brundy, C. (2022). Left in the cold: The failure of APC waiver programs to provide author equity. *Science Editor*, *45*, 5–13. https://doi.org/10.36591/SE-D-4501-5

Scheidsteger, T., & Haunschild, R. (2022). Comparison of metadata with relevance for bibliometrics between Microsoft Academic Graph and OpenAlex until 2020. *arXiv*, arXiv:2206.14168. https://doi.org/10.48550/arXiv.2206.14168

Segado-Boj, F., Martín-Quevedo, J., & Prieto-Gutiérrez, J. J. (2018). Attitudes toward open access, open peer review, and altmetrics among contributors to Spanish scholarly journals. *Journal of Scholarly Publishing*, *50*(1), 48–70. https://doi.org/10.3138/jsp .50.1.08

Siler, K., & Frenken, K. (2019). The pricing of open access journals: Diverse niches and sources of value in academic publishing. *Quantitative Science Studies*, *1*(1), 28–59. https://doi.org/10 .1162/qss_a_00016

Siler, K., Haustein, S., Smith, E., Larivière, V., & Alperin, J. P. (2018). Authorial and institutional stratification in open access publishing: The case of global health research. *PeerJ*, *6*, e4269. https:// doi.org/10.7717/peerj.4269, PubMed: 29479492

Simard, M.-A., Ghiasi, G., Mongeon, P., & Larivière, V. (2022). National differences in dissemination and use of open access literature. *PLOS ONE*, *17*(8), e0272730. https://doi.org/10.1371 /journal.pone.0272730, PubMed: 35943972

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., … Wang, K. (2015). An overview of Microsoft Academic Service (MAS) and applications. In *Proceedings of the 24th International Conference on World Wide Web—WWW '15 Companion* (pp. 243–246). https://doi.org/10.1145/2740908.2742839

Smith, A. C., Merz, L., Borden, J. B., Gulick, C. K., Kshirsagar, A. R., & Bruna, E. M. (2021). Assessing the effect of article processing charges on the geographic diversity of authors using Elsevier's "Mirror Journal" system. *Quantitative Science Studies*, *2*(4), 1123–1143. https://doi.org/10.1162/qss_a_00157

Solomon, D. J., & Björk, B.-C. (2012a). Publication fees in open access publishing: Sources of funding and factors influencing choice of journal. *Journal of the American Society for Information Science and Technology*, *63*(1), 98–107. https://doi.org/10.1002/asi.21660

Solomon, D. J., & Björk, B.-C. (2012b). 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

Stehr, N. (1994). *Knowledge societies*. London: Sage.

Turner, B. E., Steinberg, J. R., Weeks, B. T., Rodriguez, F., & Cullen, M. R. (2022). Race/ethnicity reporting and representation in U.S. clinical trials: A cohort study. *The Lancet Regional Health – Americas*, *11*, 100252. https://doi.org/10.1016/j.lana.2022 .100252, PubMed: 35875251

UNESCO. (2021). *UNESCO recommendation on open science*. https://en.unesco.org/science-sustainable-future/open-science /recommendation

Van Eck, N. J. (2021). CWTS Leiden Ranking 2021 [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.4889279

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391. https://doi .org/10.1016/j.joi.2016.02.007

Waltman, L., & Traag, V. A. (2020). Use of the journal impact factor for assessing individual articles need not be statistically wrong.

F1000Research, *9*, 366. https://doi.org/10.12688/f1000research.23418.1

Xie, Y., Zhang, C., & Lai, Q. (2014). China's rise as a major contributor to science and technology. *Proceedings of the National Academy of Sciences*, *111*(26), 9437–9442. https://doi.org/10.1073/pnas.1407709111, PubMed: 24979796

Young, K., Fisher, J., & Kirkman, M. (2019). "Do mad people get endo or does endo make you mad?": Clinicians' discursive constructions of medicine and women with endometriosis. *Feminism & Psychology*, *29*(3), 337–356. https://doi.org/10.1177/0959353518815704

Zhang, L., Wei, Y., Huang, Y., & Sivertsen, G. (2022). Should open access lead to closed research? The trends towards paying to perform research. *Scientometrics*, *127*, 7653–7679. https://doi.org/10.1007/s11192-022-04407-5

Zhou, P., & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy*, *35*(1), 83–104. https://doi.org/10.1016/j.respol.2005.08.006