



Predatory publishing in Scopus: Evidence on cross-country differences

Vít Macháček^{1,2}  and Martin Srholec¹ ¹CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Prague, Czech Republic²Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague, Czech Republican open access  journal

Citation: Macháček, V., & Srholec, M. (2022). Predatory publishing in Scopus: Evidence on cross-country differences. *Quantitative Science Studies*, 3(3), 859–887. https://doi.org/10.1162/qss_a_00213

DOI:
https://doi.org/10.1162/qss_a_00213

Received: 11 February 2022
Accepted: 6 August 2022

Corresponding Author:
Vít Macháček
vit.machacek@cerge-ei.cz

Handling Editors:
Ludo Waltman and Vincent Larivière

Copyright: © 2022 Vít Macháček and Martin Srholec. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



Keywords: academic misconduct, Beall's list, open access, predatory journal, research system, research policy

ABSTRACT

Predatory publishing represents a major challenge to scholarly communication. This paper maps the infiltration of journals suspected of predatory practices into the citation database Scopus and examines cross-country differences in the propensity of scholars to publish in such journals. Using the names of “potential, possible, or probable” predatory journals and publishers on Beall's lists, we derived the ISSNs of 3,293 journals from Ulrichsweb and searched Scopus with them. A total of 324 of journals that appear in both Beall's lists and Scopus, with 164,000 articles published during 2015–2017 were identified. Analysis of data for 172 countries in four fields of research indicates that there is a remarkable heterogeneity. In the most affected countries, including Kazakhstan and Indonesia, around 17% of articles were published in the suspected predatory journals, while some other countries have no articles in this category whatsoever. Countries with large research sectors at the medium level of economic development, especially in Asia and North Africa, tend to be most susceptible to predatory publishing. Policy makers and stakeholders in these and other developing countries need to pay more attention to the quality of research evaluation.

1. INTRODUCTION

“Predatory” (or fraudulent) scholarly journals exploit a paid open-access publication model: The publisher does not charge subscription fees but receives money directly from the author of an article that becomes accessible for free to anyone. However, this entails a conflict of interests that has the potential to undermine the credibility of open-access scholarly publishing (Beall, 2013). Authors are motivated to pay to have their work published for the sake of career progression or research evaluation, for instance (Bagues, Sylos-Labini, & Zinovyeva, 2019; Demir, 2018; Kurt, 2018). In return, predatory publishers turn a blind eye to any limitations of papers during peer review in favor of generating income from authors' fees; the worst of them fake the peer-review process and print almost anything for money, without scruples (Bohannon, 2013; Butler, 2013).

So far, only a handful of studies have examined the geographical distribution of authors published in journals suspected of predatory practices by Beall (2016). On a sample of 47 such journals, Shen and Björk (2015) found that the authors were highly skewed to Asia and Africa, primarily India and Nigeria. Xia, Harmon et al. (2015) examined seven pharmaceutical

journals and also identified the vast majority of authors as being from Southeast Asia, predominantly India, and, to a lesser extent, Africa. Demir (2018) combed through 832 suspected predatory journals and confirmed that by far the greatest number of authors are from India, followed by Nigeria, Turkey, the United States, China, and Saudi Arabia. Wallace and Perri (2018) focused on 27 such journals in economics, in which the authors were most frequently from Iran, the United States, Nigeria, Malaysia, and Turkey.

No matter how insightful these studies are in revealing from where contributors to suspected predatory journals originate, we still know very little about the magnitude of the problem for the respective countries and regions. India appears to be the main hotbed of predatory publishing, but in the context of India's gigantic research system, this may be much ado about little. All the countries cited above are, unsurprisingly, quite large. Could it be that some smaller countries are actually far worse off, though they do not stand out in the absolute figures? Just how large is the propensity to predatory publishing at the national level? Which countries are most and least affected by this problem, and why?

Existing literature provides very scant evidence along these lines and the studies at hand are limited to individual countries and use different methodologies, so the results are not easily comparable. For example, Perlin, Imasato, and Borenstein (2018) found that suspected predatory journal articles accounted only for about 1.5% of publications in Brazil, while Bagues et al. (2019) showed that around 5% of researchers published in such journals in Italy. No study has yet examined the penetration of national research systems by predatory publishing in a broad comparative perspective. Systematic scrutiny of cross-country differences worldwide is lacking.

This paper helps to fill that gap by examining the propensity to publish in suspected predatory journals for 172 countries in four fields of research during the 2015–2017 period. Using the names of journals and publishers on lists by Beall (2016), we derived the ISSNs of 3,293 titles from Ulrichsweb (2016) and searched Scopus (2018a) for them. We identified 324 matched journals with 164,000 indexed articles. Next, we downloaded from Scopus the number of articles by author's country of origin published in these journals and compared the figures to the total number of indexed articles by country and field. The resulting database provides more representative and comprehensive country-level evidence on publishing in the suspected predatory journals than has been available in any previous study.

Our analysis indicates that there is remarkable heterogeneity in the propensity to publish in suspected predatory journals across countries. In line with earlier evidence, the most affected countries are in Asia and North Africa, but they are not necessarily the same ones cited above. In the most affected countries, including Kazakhstan and Indonesia, around 17% of articles were published in the suspected predatory journals, while some countries have no articles in this category whatsoever. India's situation also looks daunting, but it is not the worst off. Econometric analysis of cross-country differences shows that countries with large research sectors at the medium level of economic development tend to be most susceptible to predatory publishing. Arab, oil-rich, and/or eastern countries are also particularly vulnerable. To the best of our knowledge, this is the first systematic attempt to pin down national research systems at the most risk of falling into the trap of predatory publishing.

No doubt, the lists of predatory, questionable, or fake journals are controversial. It should be emphasized that the purpose of this paper is not to evaluate the suspected predatory journals and assess whether they deserve this label or not. Beall (2015, 2016) developed the identification criteria and put his reputation on the line by curating the lists, which in turn became widely used in empirical research on this topic (see, for instance, Bagues et al. (2019), Bohannon (2013), Bolshete (2018), Cobey, Grudniewicz et al. (2019), Demir (2018, 2020),

Downes (2020), Erfanmanesh and Pourhossein (2017), Frandsen (2017), Ibba, Pani et al. (2017), Kurt (2018), Perlin et al. (2018), Shen and Björk (2015), Shamseer, Moher et al. (2017), Wallace and Perri (2018), and Xia et al. (2015)). We use this source of data and our competences in comparative research to throw new light on cross-country differences in the propensity to publish in them. In our view, this helps to deepen understanding of the problem of predatory publishing.

The paper proceeds as follows. Section 2 reviews the existing literature on predatory publishing, introduces Beall's lists, and elaborates on their validity and limitations. Section 3 explains how the data set has been constructed and how it can be used. Section 4 provides an exploratory analysis of differences across countries and relevant country groups and presents econometric tests of the relationships hypothesized. The Section 5 summarizes the key findings and pulls the strands together.

2. TAKING STOCK OF THE LITERATURE

2.1. Predatory Publishing

Jeffrey Beall popularized the term *predatory publishing* on his blog (Beall, 2016). It is used to describe the practice of abusing paid open-access scientific publishing. In contrast to standard subscription-based models, authors publishing via paid open access do business directly with publishing houses. They pay article processing fees directly to the publisher of the journal. Both authors and publishers are motivated to publish articles. Predatory journals perform only vague, pro forma, or (in some cases) no peer review, and allow publication of pseudo-scientific results (Bohannon, 2013; Butler, 2013). Predatory journals have also been accused of aggressive marketing practices, having fake members of editorial boards, and amateur business management (Beall, 2015; Cobey, Lalu et al., 2018; Eriksson & Helgesson, 2017a). However, the latter are only side effects. We use the term *predatory journals* to signify journals suspected of abusing paid open access to extort fees from authors and following significantly flawed editorial practices.

The open-access model, although it is a defining element of predatory journals, is not at fault *per se*. The inherent conflict of interest does not have to be exploited. There are effective means to ensure the quality of the editorial practices of journals. Databases dedicated to supporting open access, such as the Directory of Open Access Journals, are already working to develop operational mechanisms to guarantee quality and to employ transparency measures such as open peer review, which can easily detect fraudulent publishers. Journals not performing peer reviews have admittedly nothing to report here. The existence of predatory journals does not mean that the movement calling for democratizing communication of scientific results is fruitless.

Nevertheless, it is challenging to recognize a predatory journal in practice, because there is no clearly defined boundary between journals that follow ethical editorial standards and those that are merely vehicles for exploiting publication fees. Most often, to facilitate awareness and identification, lists are used to identify suspected predatory journals. The most prominent example is Jeffrey Beall's blog (Beall, 2016), which was shut down at the beginning of 2017 (Straumsheim, 2017)¹. A private company, Cabells, subsequently began to offer a similar list (Silver, 2017), the content of which, however, is locked behind a paywall (Cabells, 2022). China has announced the formation of a list of "poor quality" journals (Cyranoski, 2018),

¹ Anonymous authors continue with Beall's work and regularly update his lists on a new website (Anonymous, 2022).

which was followed by the creation of a list of questionable journals by the National Science Library of the Chinese Academy of Sciences (Zhang, Wei et al., 2022), but this list seems to be far narrower in scope than both of its predecessors.

The inclusion of individual journals on a list should be based on rigid and transparent criteria. Beall (2015) provided a list of criteria that he used to make decisions about journals and publishers. Eriksson and Helgesson (2017a) and Cobey et al. (2018) have also suggested a similar list of characteristics to identify predatory journals. The key set of Beall's criteria points directly to the most salient problem of dubious editorial practices: ("Evidence exists showing that the publisher does not really conduct a bona fide peer-review"; "No academic information is provided regarding the editor, editorial staff, and/or review board members"). However, there is also a group of indicators concerning professionalism and/or compliance with ethical standards ("The publisher has poorly maintained websites, including dead links, prominent misspellings and grammatical errors on the website"; "Use boastful language claiming to be a 'leading publisher' even though the publisher may only be a start-up or a novice organization", etc.).

Grudniewicz, Moher, and Cobey (2019, p. 211) addressed what they perceived as a lack of agreed definition of predatory publishing by convening dozens of experts on this topic, who arrived at the following consensus: "Predatory journals and publishers are entities that prioritize self-interest at the expense of scholarship and are characterized by false or misleading information, deviation from best editorial and publication practices, a lack of transparency, and/or the use of aggressive and indiscriminate solicitation practices," which is arguably well in tune with Beall (2015). But when it comes to criteria for the identification of predatory journals in practice, they argue for relying on easy to detect defects, such as misinformation, spamming, and/or spelling errors, rather than attempting to assess the quality of peer review. By giving up on the latter, however, the identification is likely to miss predatory journals that have become professionalized and manage to avoid the most obvious blunders, while still neglecting peer review to maximize profits. In this regard, we concur with Moed, Lopez-Illescas et al. (2022) that accepting manuscripts without any rigorous form of peer review is the core characteristic of predatory journals that we thus should not leave out.

Kurt (2018) identified four pretexts that are often used to justify publication in predatory journals by the authors: social identity threat; lack of awareness; high pressure to publish; and lack of research proficiency. The common denominator is urgency. Researchers tend to publish in these journals as a last resort and often refer to institutional pressure, a lack of experience, and fear of discrimination from "traditional" journals. Justifications for publishing in predatory journals therefore appears to be a complex mix of factors operating at both personal and institutional levels.

Demir (2018) and Bagues et al. (2019) also argue that the tendency to publish in predatory journals is likely to be related to the quality of research evaluation in the country. The more the research evaluation system relies on outdated routines such as counting articles indexed in Scopus, Web of Science, or Medline, the higher incentive for researchers to publish in fraudulent journals just to clinch points for outputs regardless of merit. In countries where the culture of evaluation and peer pressure push researchers to publish in respectable journals, there is little to no motivation to resort to predatory journals, as such behavior will harm the researcher's reputation.

Predatory publishing can be seen as wasteful of resources. Shen and Björk (2015) estimated the size of the predatory market as high as US\$74 million in 2014, based on article processing fees, and the figure may well have grown significantly since. Perhaps more important than the

direct costs, however, are indirect costs stemming from the fact that the opportunity to bypass the standard peer-review process leads researchers astray. Instead of spending their time producing relevant insights, researchers may be increasingly prone to writing bogus papers that only pretend to be scientific. If this occurs on an increasing scale, research systems are in peril. The fact that research published in scientific journals is predominantly funded from public sources only amplifies these concerns.

2.2. Beall's Lists

Beall (2016) maintained two regularly updated lists of “potential, possible, or probable” predatory journals and publishers, henceforth for the sake of brevity referred to as “suspected predatory”: a) a “list of standalone journals,” which contains suspected individual journals; and b) a “list of publishers,” which contains suspected publishers, most of which print multiple journals.

Crawford (2014b) went through every single item on Beall's lists in late March and early April 2014. He found 9,219 journals, of which 320 were from the list of standalone journals and 8,899 from the list of publishers. Between 2012 and 2014, about 40% of those journals published no or fewer than four articles; in other words, they were empty shells, and a further 20% published only a handful of articles. Another 4% consisted of dying or dormant journals whose publications fell to a few articles in 2014, and 6% were unreachable (the web link was broken, for instance). Overall, fewer than 30% of the identified journals published articles regularly. Fewer than 5% of the journals appeared “apparently good as they stand,” meaning that there was no immediate reason to doubt their credibility, which, however, did not imply that they were in fact credible.

Shamseer et al. (2017) confirmed that Beall's listed journals contained more spelling errors, promoted bogus bibliometric metrics on their websites, and their editorial board members were much more difficult to verify than those of “ordinary” journals. Bohannon (2013) exposed flawed editorial practices by submitting fake scientific articles to journals of publishers from Beall's list. The fake articles were accepted for publication by four-fifths of the journals that completed the review process, which vindicates doubts about their peer review routines. Bagues et al. (2019) showed that journals on Beall's list tend to have low academic impact and cite researchers admitting that the editorial practices of these journals are flawed. Journals from these lists truly seem to be doubtful.

Moed et al. (2022) examined journals from Beall's list of publishers with the help of bibliometric analysis using an updated version of a database published with an earlier version of this paper. First, they found that the article output of a random sample of these journals that were not indexed in Scopus had a strong tendency to dwindle, and two-fifths of them were discontinued, confirming that most of them do not succeed in becoming regular publication venues. Second, they found that the subset of these journals that were indexed in Scopus suffered as a group a strong decline in citation impact and achieved impact levels far below that of a control group of other open access journals indexed in Scopus, which they interpret as a signal that in general their scientific relevance is inferior. Finally, however, they also pointed to variability of the suspected predatory journals and that judging by bibliometric records the inclusion of some publishers on Beall's list might be questionable.

Strinzel, Severin et al. (2019) compared lists of predatory journals originated by Beall (updated by its anonymous continuator) and more recently launched by Cabells Scholarly Analytics (hereafter Cabells) as well as lists of credible journals compiled by the Directory of Open Access Journals (DOAJ) and Cabells using data from December 2018. In terms of journals and publishers indexed, they concluded that there was a considerable overlap

between the lists or predatory journals and even speculated that Cabells' list may have used Beall's list as a source, but that there was essentially no or a very limited overlap between them the lists of credible journals². In terms of inclusion criteria, the analysis revealed that both of the lists of predatory journals most frequently considered business practices, including the business model, misinformation on location, spamming, and boastful language, but that these aspects were far more dominant for Beall than Cabells, and that the main difference was that Cabells used noticeably more criteria than Beall related to peer review and policy. However, as also acknowledged as a limitation by the authors, the comparison relied only on the number of criteria in each category, not reflecting on their relative weight for indexing, which could have differed significantly.

2.3. Limitations

As Eriksson and Helgesson (2017b) state, "the term 'predatory journal' hides a wide range of scholarly publishing misconduct." Some are truly fraudulent, while many others may operate on the margins. However, Beall's lists force us to work with a binary classification in which a journal or publisher is considered either predatory or not. As Beall did not systematically explain his decisions, it is not possible to make a more detailed quantification of "predatoriness," though elaborated criteria exist.

Beall's lists have been strongly criticized for the low transparency of his decision-making process (Berger & Cirasella, 2015; Bloudoff-Indelicato, 2015; Crawford, 2014a). Although the criteria are public, justification of decisions on individual journals and publishers is often not clear and difficult to verify. Beall debated the decisions on his blog or on Twitter in some important instances, but very often a journal or publisher was added to the list without justification being provided. The lack of comprehensive, rigid, and formal justification of Beall's judgments is a major drawback of his list.

In particular, caution is warranted when working with Beall's list of publishers. Classifying an entire publishing house as suspected predatory is a strong judgment, and it cannot be ruled out that some journals which actually apply reputable standards have been listed along the way. The list includes some publishers that maintain broad portfolios of dozens and even hundreds of journals, some of which may not deserve the predatory label, so that using Beall's list may result in overestimations of true "predators." It is likely that the overwhelming majority of

² Strinzel et al. (2019) remain silent on what explains the differences between the lists of predatory journals that they actually identified, but four factors driven by data issues are likely to be in play. First, Beall's lists include only open access journals and publishers, while Cabells' list also includes subscription-based ones. Second, one needs to keep in mind the findings by Crawford (2014b) that most of the journals on Beall's lists are empty shells, the findings by Moed et al. (2022) that they tend to dwindle or become discontinued, and the findings by Siler, Vincent-Lamarre et al. (2021) that they become rebranded and morphed into different outlets. The origins of Beall's lists go back to the early 2010s and he was well known to focus on adding new records rather than deleting possibly irrelevant old ones, while Cabells' list was launched in 2017; therefore many records that remain indexed in the former might not appear in the latter simply because they ceased to exist or changed in the meantime. Third, Beall's blog was shut down at the beginning of 2017 (Straumsheim, 2017) and even though its continuator pledges to update it regularly (Anonymous, 2022), it cannot be taken for granted that the updates are as thorough as could have been in Beall's original endeavor, and therefore that new predatory journals that emerged in the meantime may have been recorded in Cabells' list but not in the updated Beall's list. Finally, Strinzel et al. (2019) did not identify individual journals from Beall's list of publishers, while Cabells' lists of journals and publishers are linked together; hence they compared the restricted list of Beall's standalone only journals with the all-encompassing list of Cabells journals, as a result of which the overlap at the journal level had been underestimated.

these journals are of poor quality, but poor quality is not a crime *per se*. One must, therefore, keep in mind that the list of publishers has been painted with a relatively broad brush.

Nevertheless, respectable publishing houses should have zero tolerance for predatory practices. Just as in the banking sector, academic publishing services are based on trust, and if that is lost, the business is doomed. A single journal with predatory inclinations that are not quickly corrected by the publisher can substantially damage the entire brand. Beall's suspected predatory mark signals serious doubts about the publisher's internal quality assurance mechanisms at the very least.

The greatest controversy was triggered by inclusion of the Frontiers Research Foundation on Beall's list of publishers in October 2015. Beall defended this decision by pointing out several articles that, according to him, should not have been published. According to critics of this move, the Frontiers publisher is "legitimate and reputable and does offer proper peer-review" (Bloudoff-Indelicato, 2015). Frontiers journals appear to be quite different from typical suspected predatory outlets on the face value of their citation rates. Only four journals in Frontiers' portfolio of 29 included in this study are not ranked in the first quartile in at least one field according to the Scimago SJR citation index (Scopus, 2018b). Most Frontiers journals are also indexed in the Web of Science and the Directory of Open Access Journals. Hence, judging by the relevance of Frontiers journals for the scientific community, there is a question mark about their inclusion on Beall's list.

Another concern arises from the time scale. The suspected predatory status used in this study is derived from the content of Beall's lists on April 1, 2016. Jeffrey Beall continually updated his lists. However, the lists always reflect only current status, with no indication of when the journal and publisher may have become suspected to be predatory. When looking back in time, we may run into the problem of including in the predatory category records that do not deserve that label, because the journal became suspected only a short time before its inclusion on the list. In some cases, older articles published in journals that are currently suspected to be predatory may have gone through a standard peer review. Hence, historical data must be used with great caution.

Further, Beall's lists are very likely to suffer from English bias. The lists contain mainly journals that at least have English-language websites. In regions in which a large part of scientific output is written in other languages—such as in Latin America, Francophone areas, and countries of the former Soviet Union—estimates of the extent of predatory publishing based on Beall's lists may be underestimated, because Beall did not identify suspected predatory journals in local languages. Likewise, Scopus covers scientific literature in English far more comprehensively than publications in other major world languages. This bias should be kept in mind when interpreting cross-country differences.

3. DATABASE

Our database was built in three steps. First, we compiled a comprehensive overview of journals suspected of predatory practices by matching the lists of standalone journals and publishers by Beall (2016) with records in the Ulrichsweb (2016) database, which provides comprehensive lists of periodicals. Second, we searched the International Standard Serial Numbers (ISSNs) of the journals obtained from Ulrichsweb in Scopus and downloaded data on authors publishing in these journals by their country of origin. Third, we downloaded the total number of indexed articles by country from Scopus. Ultimately, we obtained not only a full list of suspected predatory journals listed in Scopus but, even more importantly, we also obtained harmonized data on the propensity to publish in these journals by country, which

allows us to shed new light on cross-country patterns (for a brief overview of the data generation process see Table 1).

Beall's lists were downloaded on April 1, 2016. First, we identified all search terms in each item on the lists. For some entries, Beall presented multiple versions of a journal designation; for example, the journal name and its abbreviation. All available versions were used as a search term. Next, we searched the terms in the Ulrichsweb database for the same day, using an automatic

Table 1. Overview of the data generation process

1. Obtaining the ISSNs of suspected predatory journals:

- (a) Beall's lists downloaded on April 1, 2016.
- (b) The names on Beall's lists were searched for using an automatic script in Ulrichsweb on the same day.
- (c) The entries found in Ulrichsweb were manually verified with the help of hypertext links in Beall's lists.
- (d) 4,665 ISSNs of 3,295 individual journals were confirmed to be associated with Beall's lists.

2. Searching for "predatory" ISSNs in Scopus:

- (a) The "predatory" ISSNs were searched for using an automatic script in Scopus on March 19, 2018.
- (b) 439 ISSNs of 324 individual journals that had at least one entry in Scopus during the period 2015–2017 were identified.
- (c) The script downloaded the total number of indexed articles in each journal and the number of these articles by the author's country of origin during the period 2015–2017.
- (d) To avoid double counting articles in journals with ISSNs for both print and electronic versions, duplicates were eliminated.

3. Downloading total number of articles in Scopus by country and field of research:

- (a) The total number of indexed articles by country during the period 2015–2017 was downloaded using the Scopus API on March 19, 2018.
 - (b) The total number of indexed articles by country and field of research during the period 2015–2017 was downloaded using the Scopus API on March 5, 2020.
-

script programmed in Python. When we searched for a standalone journal, the script used the “title” field, and for the publisher, the script used the “publisher” field. In the end, the algorithm saved all search results. The search request in Ulrichsweb was as follows for standalone journals:

```
+(+title:(“Academic Exchange Quarterly”))
```

and for publishers:

```
+(+publisher:(“Abhinav”))
```

The raw search on Ulrichsweb produced a database of 19,141 results linked to individual entries on Beall’s list. Results without ISSNs were removed, as they were most probably not listed in Scopus anyway; this reduced the database to 16,037 search results with 7,568 unique ISSNs. The reduction is due to using multiple search terms related to the same entry and to the “fuzziness” of the Ulrichsweb search³. To make sure that the journals are listed by Beall, the remaining search results were checked manually. Beall’s lists consist of hypertext links, so we compared the ISSN on the journal’s website with the ISSN on Ulrichsweb. If the two ISSNs matched, the entry was retained; if they differed, the entry was removed from our database. A publisher’s identity was confirmed if at least one ISSN listed on its website was found in an entry linked to the publisher’s name on Ulrichsweb.

We confirmed 4,665 unique ISSNs associated with Beall’s lists. Many journals have dual ISSNs, one for its print version and one for its electronic version. The number of individual journals is 3,293, of which 309 featured on the list of standalone journals, 2,952 referred to the list of publishers, and an additional 32 journals appeared on both lists, perhaps because Beall did not recognize that the respective journal was from a publisher already on his list. For simplicity, these journals are considered to belong to the list of publishers.

This is in line with the analysis of Crawford (2014b), which identified fewer than 3,000 journals that published articles regularly, and thus in fact appeared to be continuously in operation. Shen and Björk (2015) found around 8,000 journals that were “active” in the sense that they published at least one article. However, many of these, as per Crawford (2014b), may not publish significantly more than that and are not likely to be registered in databases. Note that there are 1,003 hypertext links on the list of standalone journals, from which it follows that more than two-thirds of these are not included in Ulrichsweb, let alone in more selective databases. Apart from the unverified information on their web pages, there is no information about them. Previous attempts to collect data on suspected predatory journals were far less comprehensive⁴.

In the next step, we searched for the presence of these “predatory” ISSNs in the Scopus (2018a) citation database during the period 2015–2017. Once again, this search was performed using an automatic script programmed in Python. The search was performed on March 19, 2018. For each ISSN detected in Scopus, the script downloaded not only the total number of documents in the “article” category but also more detailed data on the

³ The Ulrichsweb search engine uses a “fuzzy” search that does not require perfect matching of strings. For example, when we searched for *Academe Research Journals*, journals of *Academic Research Journals* were also found. This is beneficial because the search is robust to typos, interpunctuation signs, and small errors written in the search terms. However, it also requires careful manual verification of search results.

⁴ For example, Perlin et al. (2018) found only 1,100 ISSNs from both the list of publishers and the list of standalone journals using an automatic website crawler and Demir (2018) analyzed only the list of standalone journals.

number of these articles by the author's country of origin. The search request in Scopus was as follows:

ISSN(1234-5678) AND DOCTYPE(ar) AND PUBYEAR > 2014 AND PUBYEAR < 2018

We identified 439 ISSNs of 324 individual journals with at least one entry in Scopus, of which 37 appear on the list of standalone journals and 287 on the list of publishers. Thus, nearly 10% of the journals in our database were indexed in Scopus. We detected 164,073 articles published in these journals, of which 22,235 occur in standalone journals and 141,838 come from the list of publishers, jointly making up 2.8% of all articles indexed in Scopus during the period under consideration. Hence, the list of publishers, which was rather neglected in previous empirical studies of predatory publishing, is the dominant source. The journals were assigned to four broad fields of research: Health Sciences; Life Sciences; Physical Sciences; and Social Sciences, based on the Scopus Source List (Scopus, 2018b). If a journal is assigned to multiple fields, it is fully counted in each of them. The database is available for download in Zenodo (Macháček & Srholec, 2022b).

Finally, we obtained data on the total number of articles in Scopus by author's country of origin and field of research during the period 2015–2017, which is the denominator required to compute the penetration of suspected predatory journals in the article output of each country. The download was performed on March 5, 2020. The search was performed using the following request:

AFFILCOUNTRY(country) AND SUBJAREA(field) AND DOCTYPE(ar) AND PUBYEAR > 2014 AND PUBYEAR < 2018

In the Scopus database, an article is fully attributed to a country if the affiliation of at least one of its authors is located in that country. Joint articles by authors from different countries are counted repeatedly in each participating country. Hence, the data measure article counts, not fractional assignments. If articles in suspected predatory journals have fewer coauthors than other articles, the predatory articles penetration is underestimated and vice versa; this can be uneven across countries⁵. For some articles, Scopus reports the country of origin as "undefined"; these are excluded from our analysis⁶.

How come there are suspected predatory journals in Scopus? Journals need to fulfil a number of selection criteria to become indexed in the database (Scopus, 2019). However, these criteria are either formal, such as having an ISSN, online availability, and English language abstracts and titles, or derived from bibliometrics, such as a minimum threshold of citations, article output and diversity of authors by country; or rely on policies in the sense of what the journal declares what it does, for instance, in terms of peer review, rather than what it does in practice. If these boxes are ticked, the journal is very likely to be accepted into Scopus. Yet predatory journals that have managed to professionalize their business operation might look like regular scientific outlets on the outside, their bibliometric profile might not differ that much from other fringe journals, and they do not shy away from lying about their editorial

⁵ Unfortunately, the Scopus database does not directly provide harmonized data on the number of authors by country that published in a journal. However, we can count the number of countries to which at least one author of an article is affiliated by journal. Based on data for 324 suspected predatory journals and 23,387 other Scopus journals, the average numbers of country affiliations turns out to be 1.20 and 1.23, respectively; hence there is not a significant difference and the bias is likely to be rather small.

⁶ Only 1,069 suspected predatory journal articles had an "undefined" country of origin. Hence, the overwhelming majority of the articles found are included in our analysis.

practices; deception is their defining feature. So this filter is not likely to be effective in keeping out predatory journals that are good pretenders.

Scopus (2021) in a reaction to the earlier publication of this article (Macháček & Srholec, 2021) acknowledged that the database included the matched journals that have been identified in our analysis, re-evaluated them and discontinued coverage of more than two-thirds of them:

All of the 137 suspicious titles mentioned in the paper have gone through the re-evaluation process and as a result for 97 titles (71%), the decision was made to discontinue coverage in Scopus. Also, all other journals listed by Beall that are mentioned in the paper have gone through the re-evaluation process and as a result 65% of these titles were also discontinued. (Scopus, 2021, p. 5)⁷

Moed et al. (2022) also reported that the indexing of about 60% of the suspected predatory journals that they found in Scopus using the updated version of our database was discontinued and that 2016 was a peak year in this respect. Scopus thus validated in hindsight that most of these journals were problematic and probably should not have been included in the database in the first place. In the meantime, however, papers published in these journals before they were discontinued remain included in the database, possibly misleading unsuspecting readers by its content⁸.

At the same time, it needs to be emphasized that the fact that a matched journal has not been discontinued by Scopus does not signify that it should be absolved from the suspected predatory status. Scopus selection criteria and by extension the re-evaluation criteria, as discussed above, rely heavily on bibliometric indicators and a journal's declared policies and only partly check for the attributes that have been proposed to identify predatory journals (Beall, 2015; Grudniewicz et al., 2019; Strinzel et al., 2019), especially with regard to differences between what the journal claims to be the case and reality. Scopus (2021) noted that formal complaints that have been raised about publication standards are reflected in the re-evaluation process but does not provide any details (i.e., how many complains are collected, what they are typically about, how they are fact-checked, and what represents an offense that is serious enough to discontinue coverage). Until this mechanism of data collection becomes more transparent and widely used by the research community, one cannot rule out that predatory journals fall through the cracks.

Admittedly, the suspected predatory journals that are indexed in Scopus represent only the tip of the iceberg, which is not representative of the whole business. No matter how imperfect the entry filter of Scopus turns out to be, the journals that made it through represent probably the least ugly part. Nevertheless, from the research evaluation perspective, suspected predatory journals indexed in respected citation databases are more dangerous than ordinary bogus journals that few take seriously, because the indexation bestows a badge of quality⁹. All too often, evaluations at various levels rely on this badge and blindly assume that whatever is indexed counts and deserves to be supported by taxpayer's money. Scopus-listed journals are in practice considered "scientific" by many institutions and even national evaluation systems, such as in the Czech Republic (Good, Vermeulen et al., 2015), Italy (Bagues et al., 2019),

⁷ Macháček and Srholec (2021) identified 324 suspected predatory journals with at least one entry in Scopus during the period 2015–2017, which does not correspond with the number of journals cited by Scopus (2021), but the statement makes clear that most of those that were re-evaluated have been discontinued.

⁸ An early version of this paper came out in March 2017 (Macháček & Srholec, 2017) and was presented at the Scopus Content Selection and Advisory Board Meeting in Prague on November 3, 2017.

⁹ We use Scopus rather than the Web of Science because it covers substantially more journals (Mongeon & Paul-Hus, 2016) and is more vulnerable to suspected predators (Demir, 2020; Somoza-Fernández, Rodríguez-Gairín, & Urbano, 2016).

and probably many other countries as well. In particular, evaluation systems that do not check the actual content using their own peer-review assessment are most exposed, but such assessment tends to be expensive and difficult to organize, and thus is relatively rare in exactly the environments that need this check most.

4. CROSS-COUNTRY PATTERNS

Out of more than 200 countries for which the data are available, we excluded dependent territories and countries with fewer than 300,000 inhabitants. The analysis considers evidence from the period between 2015 and 2017, because, as noted above, using older data risks that some of the journals currently featured on Beall’s lists were not yet predatory at an earlier time. However, we use data from 3 years to increase the robustness of the results. Only countries generating at least 30 articles during this period are included in the analysis. As a result, the final sample consists of 172 countries, which together account for the overwhelming majority of the world’s research activity.

The outcome variable used throughout the analysis is the share of articles linked to Beall’s lists out of all articles by authors from the given country, hence the share of articles published in suspected predatory journals out of total articles. First, we look at the global picture and examine which countries are most and least affected by predatory publishing. Then, we attempt to pin down the most salient patterns by considering differences between groups of countries. Finally, we investigate how these patterns differ by broad fields of research.

Figure 1 displays the results on a world map. The darker the color, the higher the national propensity to publish in suspected predatory journals. The main pattern is visible at a quick glance: The darkest areas are concentrated in Asia and North Africa. In contrast, Europe, North and South America and Sub-Saharan Africa are relatively pale. Hence, generally speaking, both the most and least developed countries tend to be relatively less affected, while developing countries with emerging research systems, excepting those in South America, appear to be most in harm’s way.

Table 2 shows figures for the top and bottom 20 countries. Kazakhstan and Indonesia appear to be the most badly affected, with roughly every sixth article falling into the suspected predatory category. They are followed by Iraq, Albania, and Malaysia, with more than every tenth article appearing in this category. Some of the most severely affected countries are also among the

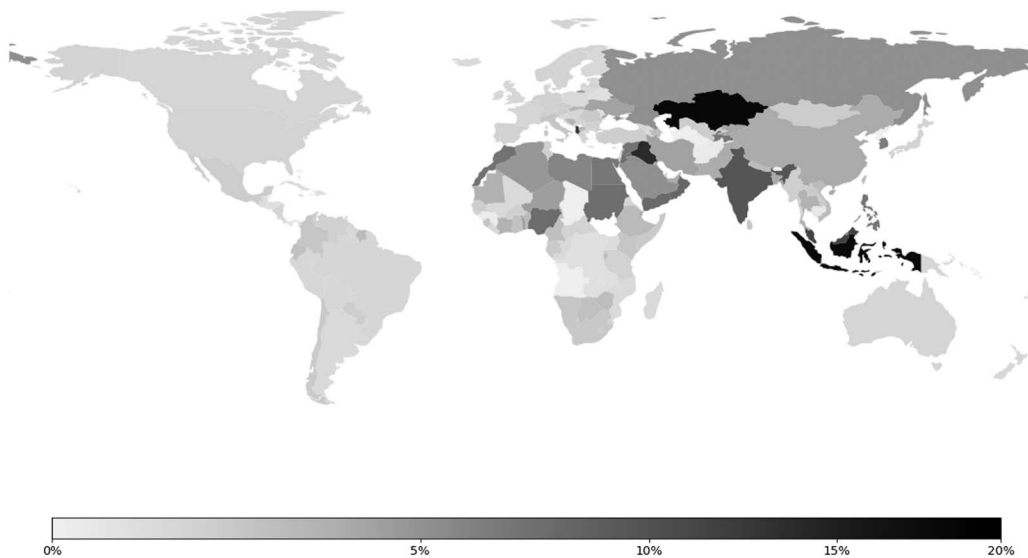


Figure 1. Percentage of suspected predatory journal articles in total articles, 2015–2017. Source: Scopus (2018a), author’s calculations.

Table 2. Percentage of suspected predatory journal articles in total articles, top and bottom 20 countries, 2015–2017

Top 20		Bottom 20	
Kazakhstan	17.00	Guatemala	0.74
Indonesia	16.73	Solomon Islands	0.74
Iraq	12.94	Bahamas	0.74
Albania	12.08	Angola	0.72
Malaysia	11.60	Honduras	0.72
India	9.65	Belarus	0.70
Oman	8.25	Congo, Dem. Rep.	0.68
Yemen	7.79	Moldova	0.67
Nigeria	7.31	Afghanistan	0.57
Sudan	7.20	Panama	0.56
Jordan	7.19	Cambodia	0.40
Morocco	6.95	Haiti	0.35
Syria	6.88	Guinea	0.10
Philippines	6.68	Belize	0.00
Egypt	6.65	Bhutan	0.00
Palestine	6.56	Cape Verde	0.00
Tajikistan	6.48	Chad	0.00
South Korea	6.37	Maldives	0.00
Libya	6.06	North Korea	0.00
Brunei	5.44	Turkmenistan	0.00

Source: Scopus (2018a), author's calculations.

largest in terms of population: India, Indonesia, Nigeria, the Philippines, and Egypt, which underlines the gravity of the problem. However, small countries that might have been difficult to spot on a world map, such as Albania, Oman, Jordan, Palestine, and Tajikistan are also seriously affected. South Korea is by far the worst among advanced countries. All countries on the top 20 list, excepting only Albania, are indeed in, or very near, Asia and North Africa.

Surprisingly, the opposite end of the spectrum, with the lowest penetration of suspected predatory journal articles, is also dominated by developing countries, including some of even the least developed. In several, for instance Bhutan, Chad, and North Korea, there are no authors published in suspected predatory journals whatsoever. This is a rather diverse group of countries scattered across continents. Nevertheless, they have one additional feature in common: Most are small countries with underdeveloped research systems. In fact, 13 countries on the bottom 20 list produced fewer than 100 articles per year, on average. It may well be that these research systems are small enough to make direct oversight of the actual content of the manuscripts feasible, in which case predatory journal articles would have nowhere to

hide. In large research systems with thousands of articles produced every year, predatory publishing may more easily fly under the radar of the relevant principals.

Table 3 summarizes the main patterns by presenting average propensities to publish in suspected predatory journals by country groups, and provides details by the source list. First, we

Table 3. Percentage of suspected predatory journal articles in total articles by country group and source list, 2015–2017

Country group	Number of countries	Total	Source list			
			Standalone	Publishers excl. Frontiers	Frontiers	Total excl. Frontiers
<i>Geography</i>						
Europe	40	1.96	0.32	0.95	0.68	1.27
America	28	1.22	0.10	0.53	0.59	0.63
Asia	49	4.22	0.86	3.01	0.35	3.87
Africa	50	2.33	0.41	1.27	0.64	1.68
Oceania	5	1.14	0.04	0.43	0.67	0.47
<i>Language</i>						
English spoken	37	2.64	0.41	1.65	0.58	2.06
French spoken	21	2.41	0.35	1.22	0.84	1.57
Spanish spoken	21	1.24	0.11	0.43	0.71	0.53
Arabic spoken	21	5.13	1.17	3.52	0.44	4.69
Other language spoken	86	2.42	0.45	1.49	0.48	1.94
<i>Natural resources rents</i>						
Oil and natural gas	24	3.90	0.80	2.68	0.41	3.49
Other natural resources	39	1.77	0.23	0.87	0.67	1.10
Other countries	108	2.51	0.45	1.50	0.56	1.95
<i>Income per capita</i>						
High income	48	2.10	0.22	1.11	0.76	1.33
Upper middle income	44	2.92	0.55	1.95	0.41	2.51
Lower middle income	48	3.28	0.78	2.08	0.42	2.86
Low income	30	1.63	0.16	0.76	0.71	0.92
<i>Size of the research sector</i>						
Large size	43	2.56	0.35	1.48	0.73	1.83
Medium large size	43	3.49	0.75	2.25	0.49	3.00
Medium small size	43	2.62	0.47	1.69	0.46	2.16
Small size	43	1.59	0.25	0.77	0.58	1.01
All countries	172	2.56	0.46	1.55	0.56	2.00

Source: Scopus (2018a), author's calculations.

reiterate the geographical dimension by continents, which confirms that the center of predatory publication is in Asia, while the problem is relatively limited in North and South America. In fact, Suriname, the most affected country in the latter, only ranks 50th in a worldwide comparison. On average, Europe and Africa fall in between the two extremes, but this masks relatively large national differences within these continents along the east–west and north–south axes, respectively. Oceania is also little involved, but there are few countries in the region¹⁰.

Next, we examine differences by major language zones using indicators obtained from the GeoDist database which measure whether the language (mother tongue, lingua franca, or a second language) is spoken by at least 20% of the population of the country (Mayer & Zignago, 2011). Only English, French, Spanish, and Arabic are recognized separately, as other languages are not spoken in a sufficient number of countries. Note that, in contrast to geography, assignment to language zones is not mutually exclusive, as more than one language can be frequently spoken in the same country¹¹.

Admittedly, language zones partly overlap with geography. This is most apparent in South America, which is dominated by Spanish-speaking countries and thus, not surprisingly, the propensities are very similar in both country groups. More revealing is perhaps the fact that Arabic-speaking countries, which are concentrated in North Africa and the Middle East, are the primary hotbeds of predatory publishing. English- and French-speaking countries are far more geographically scattered across the globe.

As noted above, Beall's lists may suffer from English bias. Nevertheless, our results only partially support this expectation. English-speaking countries do not display significantly higher propensities towards suspected predatory publishing than Francophone areas or countries speaking other languages. Spanish-speaking countries turn out to be different, perhaps because we miss predatory journals published in Spanish by relying on Beall's lists and/or Scopus data, but speaking English specifically does not make much difference. Of course, more scholars speak English than do general populations, so tentatively the key takeaway from these figures should be that, for the most part, language does not seem to be a serious entry barrier into predatory publications.

Language zones, in turn, reflect broader differences related to religion, culture, and history, including past colonial links, which often translate to shared institutions and principles of governance. Arabic countries are likely to appear, on average, highly prone to suspected predatory publishing due to a bundle of these factors that affect how research is organized, evaluated, and funded far more than the impact of the language itself. In any case, the language zones are a handy tool to account for broad differences along these lines, especially because such data are available for a very large sample of countries.

Third, it is notable that the top 20 list (Table 2) includes oil-rich countries such as Brunei, Iraq, Kazakhstan, Libya, Nigeria, and Oman, and a closer look at the data reveals that a few more, including Algeria, Bahrain, Iran, Russia, and Saudi Arabia, line up just short of the top 20. Why could there be a connection between oil riches and susceptibility to predatory publishing at the national level? In countries that benefit from oil-related revenues, the fiscal constraints of the governments are eased, so that they can spend on whatever suits them,

¹⁰ More detailed stratification, such as dividing Asia into South, East, Central, and West, or Africa into North and Sub-Saharan, runs into the problem of too few countries in some subgroups, which would make averages unreliable.

¹¹ For example, there are four countries in which both English and French are spoken by at least 20% of the population (Canada, Cameroon, Israel, and Lebanon). Nevertheless, the vast majority of countries are assigned to a single language zone.

including the support of academic research, more than other otherwise similar countries. It may not be coincidental that some of the oil-rich countries, particularly in the Middle East but also elsewhere (Sarant, 2016; Schmoch, Fardoun, & Mashat, 2016), began to invest their resource windfalls in developing indigenous university sectors, while lacking a strong research evaluation culture, which takes time to develop. Although this strategy could be beneficial for the long-term development of these countries, if fortunes are perhaps hastily poured into supporting research, there could be undesirable side effects, such as a spike in predatory publishing.

To check whether there is a systematic pattern, we draw on indicators for rents from natural resources in the World Development Indicators database (World Bank, 2018), specifically from oil and natural gas, and also for a comparison of rents from other resources, including coal, minerals, and forests. Countries are classified as intensive on the respective resources if their resource rents account for more than 5% of GDP; this may sound low, but in practice constitutes a healthy boost to the government budget. The results confirm that countries with an economy intensive on rents from oil and natural gas are on average noticeably more susceptible to suspected predatory publishing than the rest of the world. Moreover, interestingly, this seems to be specific to oil and natural gas, as countries rich in other types of natural resources display even less tendency to this kind of publishing than countries that are not particularly endowed with any of the natural resources considered here.

Fourth, we examine whether there are differences along the level of economic development. For this purpose, we use the World Bank (2016) classification that divides countries into four groups according to gross national income per capita. In line with the anecdotal evidence discussed above, high- and low-income countries appear to be the least affected¹². The worst situation is in middle income countries, many of which recognize the role of research for development, and therefore strive to upgrade, but lag significantly behind advanced countries not only in technology but also in their ability to effectively evaluate and govern their emerging research systems. Yet the largest difference in the proclivity to suspected predatory publishing is between lower middle-income countries, such as Indonesia, India, and the Philippines, and low-income countries. Overall, therefore, there seems to be a nonlinear, specifically inverse U-shaped, relationship.

Finally, as already mentioned above, the low tendency towards suspected predatory publishing in low-income (the least developed) countries may be related to the small size of their public research sectors. To examine whether size matters, we divide the sample into quartiles according to the total number of articles published. Countries with small research sectors do not fall into the most frequent contributors to suspected predatory journals, with the single exception of Tajikistan. In fact, their vast majority rank well below the world average. More than half of low-income countries indeed fall into the small size category, and thus it is not surprising that the propensity to suspected predatory publishing proves to be similarly low in both country groups. Again, there seems to be an inverse U-shaped relationship, albeit with a different shape of the distribution.

Next, results are reported by the source list we used to identify predatory journals using three categories: Beall's list of standalone journals; Beall's list of publishers excluding

¹² The high-income group includes Persian Gulf countries, namely Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and United Arab Emirates, which are rich primarily thanks to oil drilling in the region and in which, with the exception of Qatar, the propensity to suspected predatory publishing is significantly above the world average. If these countries are excluded, the average propensity to suspected predatory publishing in the high-income group drops further to 1.74%.

Frontiers; and Frontiers. The latter is analyzed separately to account for the controversy surrounding the inclusion of Frontiers Research Foundation on Beall's list of publishers, as already discussed above. Frontiers does exhibit a noticeably different pattern from the other two sources. Authors publishing in Frontiers journals are distributed far more evenly across the country groups and in some respects, such as along income per capita, even display an opposite tendency compared to the other sources lists. The top 20 list of countries with the highest propensities to publish in Frontiers journals features Austria, Switzerland, Netherlands, Belgium, Germany, and Israel, and in these as well as most other advanced countries Frontiers is the dominant source in the total figures¹³. As a result, the main patterns identified above are even more pronounced in the total figures excluding Frontiers. From this perspective, Frontiers truly does not look like a typical predatory publisher.

The absolute numbers of articles in suspected predatory journals are also worthy of consideration. In countries with large research systems, predatory publishing can be quite extensive, even if the proportion relative to the total number articles does not seem problematic. The main case in point is China, which does not stand out in relative terms, with 3.66% of suspected predatory journal articles in the total national article count, but around 44,000 articles published in suspected predatory journals had at least one coauthor from China; this is by far the largest number worldwide. This means that nearly every fourth suspected predatory journal article has a Chinese coauthor. Next are India and the United States, with almost every sixth and ninth suspected predatory journal article coauthored by a researcher from that country, respectively. In these countries, there are legions of researchers who are willing to pay to have their work published in suspected predatory journals.

Table 4 provides details of the top 20 most affected countries and the averages across all countries by field of research. The latter indicate that the worldwide propensity to publish in suspected predatory journals is almost two times higher in Social and Life Sciences than in Health and Physical Sciences. Social Sciences are particularly ravaged by this problem in a number of countries: In seven countries, including the relatively large research systems of Malaysia, Indonesia, and Ukraine, more than one fifth of articles appear in suspected predatory journals, and in 14 countries more than one tenth of articles fall into this category. Arguably, the credibility of the whole field is at stake here.

Indonesia, Iraq, and Oman feature on the top 20 lists in all four fields and Egypt, Iran, Kazakhstan, Libya, Malaysia, Nigeria, Palestine, Sudan, and Yemen in three. In these countries, predatory publication practices may have become a systemic problem at the national level, not limited to particular clusters. In contrast, and perhaps even more interestingly at this point, there are countries in which only specific fields have gone rogue. For example, China is by far the worst in Health Sciences, but does not appear on any other field list¹⁴. Albania stands out in Social Sciences only. Likewise, India only looks disreputable in Life and Physical Sciences, Russia in Life and Social Sciences, and Ukraine in Social Sciences¹⁵.

¹³ Approximately two-thirds of suspected predatory journal articles from advanced countries are published by Frontiers. South Korea is a major outlier among advanced countries, not only because of its high overall penetration of this kind of publishing but also in the fact that the vast majority of these articles are not in Frontiers journals. Taiwan and Slovakia are similar but to a lesser degree.

¹⁴ Nevertheless, one must not forget the caveat repeatedly mentioned above that the data predominantly includes journals published in English. China not only has a different language but also its own writing system; thus local problems with the predatory model of publication may largely escape our attention.

¹⁵ In general, there are far more former socialist countries, especially former members of the Soviet Union, on the top 20 list in Social Sciences than in other fields. Social Sciences were particularly isolated, indoctrinated, and devastated during the communist era, so it is not surprising that this is the case.

Table 4. Percentage of suspected predatory journal articles in total articles by field of research, top 20 countries 2015–2017

Health Sciences		Life Sciences		Physical Sciences		Social Sciences	
China	11.72	Kazakhstan	28.10	Indonesia	22.31	Albania	37.04
Libya	6.20	Iraq	16.55	Malaysia	11.77	Malaysia	29.15
Taiwan	4.87	Syria	14.29	Philippines	10.90	Yemen	28.89
Egypt	4.84	India	13.59	Iraq	10.66	Indonesia	27.21
South Korea	4.73	Algeria	10.99	Jordan	9.19	Tajikistan	25.64
Algeria	4.58	Egypt	10.94	India	8.65	Ukraine	22.63
Luxembourg	4.57	Togo	10.37	Yemen	8.36	Kazakhstan	21.78
Suriname	4.55	Palestine	10.09	Sudan	8.05	Russia	17.54
Saudi Arabia	4.54	Libya	9.39	Morocco	7.86	Brunei	12.60
Nigeria	4.48	Indonesia	9.11	Oman	7.70	Oman	12.39
Iraq	4.36	Nigeria	9.10	South Korea	7.54	Iraq	12.24
Palestine	4.13	Oman	8.77	Kazakhstan	7.17	Azerbaijan	12.15
Indonesia	4.05	Morocco	8.42	Bahrain	6.70	Iran	11.32
Sudan	4.01	Sudan	7.91	Liberia	6.45	Syria	10.11
Iran	3.83	Iran	6.93	Palestine	6.31	Thailand	9.94
Malaysia	3.79	Russia	6.61	Nigeria	6.31	Nigeria	9.28
Chile	3.76	Yemen	6.49	Brunei	5.96	Slovakia	9.27
Italy	3.63	Macedonia	6.19	Egypt	4.99	Bahrain	9.04
United Arab Emirates	3.62	Niger	6.02	Saudi Arabia	4.85	Jordan	8.13
Oman	3.56	Mauritania	6.00	Libya	4.62	Kyrgyzstan	8.06
All countries	1.98	All countries	3.39	All countries	1.96	All countries	3.99

Note. Journals can be assigned to multiple fields of research. Only countries with at least 30 total articles in the respective field of research are included.

Source: Scopus (2018a), author’s calculations.

Overall, we have identified a handful of factors that seem to be relevant for explaining cross-country differences in the propensity to suspected predatory publishing, and which beg for more elaborate examination. Nevertheless, tabulations of the data can only get us so far in isolating their individual effects. Due to limited space and because a combination of several factors appears to be in play, we do not delve deeper into descriptive evidence by field of research but rather explore these patterns using a multivariate regression framework in the next section. The full results at the country level in total and by field of science are available for download in Zenodo (Macháček & Srholec, 2022b)¹⁶.

¹⁶ Note that most of the patterns by country groups identified in the total data also apply by field of research, as also vindicated by the regression results in Section 4.

5. REGRESSION ANALYSIS

In this section, we explore the cross-country differences with the help of an econometric model. The main focus of the analysis is on testing the hypothesized relationships between the level of economic development measured by GDP per capita, size of the (public) research sector measured by the total number of articles and the propensity to suspected predatory publishing, while controlling for other relevant factors. The empirical model to be estimated is as follows:

$$Y_{ij} = \alpha + \beta \text{GDP}_i + \gamma \text{SIZE}_i + \delta X_i + \mu_j + \varepsilon_c \quad (1)$$

where the outcome variable Y is the proportion of articles published in suspected predatory journals, variously defined, GDP per capita represents the level of economic development, SIZE represents the size of the research sector, X is the set of country-level control variables, δ is a fixed effect for the field of research represented by respective dummies, i denotes a country, j denotes a field of research, and ε is the standard error term. Hence, the basic unit of analysis is a field of research in a given country. Because differences between fields of research are fully accounted for by the fixed effects, the estimated coefficients of the country-level variables explain exclusively within-fields variability.

The dependent variable is a proportion that falls between zero and one. The ordinary least squares (OLS) estimator tends to produce predicted values outside of this range and assumes linear relationships. Both problems are addressed by using a fractional logit (binomial) in the generalized linear models (GLM) framework. Robust standard errors derived from Huber-White sandwich estimators are reported. Only observations with at least 30 total articles in the respective country-field and with full data available for the explanatory variables are included in the estimation sample. As a result, the econometric analysis is limited to 630 observations in 163 countries¹⁷. All estimates are performed in Stata/MP 15.1.

Whenever possible, we use continuous variables to measure the explanatory factors, as although the number of observations is essentially quadrupled by using the field-specific data, the sample is still relatively small. As envisaged above, GDP per capita (PPP, constant 2011 international dollars) is used to measure the level of economic development and the total number of articles indexed in Scopus is used as a rough proxy for the size of the research sector. Oil and natural gas rents (% of GDP) are used to control for the availability of extra resources that ease the fiscal constraints of the governments to invest in research. Latitude and longitude of the country's centroid, instead of plain continental dummies, are used to account for geography. However, the only way to control for the language zones is to use dummies. GDP per capita and the size of research sector variables are used in logs to curtail the impact of outliers. All variables refer to (if applicable averages over) the reference period 2015–2017. For descriptive statistics, definitions and sources of the variables entering the regression analysis, see Tables A1 and A2 in the Appendix.

The regression analysis is used as a descriptive tool in this paper. The purpose of the regression model is to test whether the broad cross-country patterns identified above hold in a multivariate framework, when the possible influence of other relevant factors is accounted for. It should be emphasized that the cross-sectional nature of the data does not allow for testing of

¹⁷ Cuba, Eritrea, North Korea, Somalia, and Syria are excluded due to missing data on GDP per capita. Comoros, Djibouti, Timor-Leste, and Turkmenistan are eliminated because they did not generate more than 30 total articles in any of the fields of research.

causality, the estimated relationships indicate correlations, and the results should therefore be interpreted with caution.

Table 5 provides results for the benchmark outcome variable of total suspected predatory publishing (Column 1), then the results are replicated separately by the source list (Columns 2–4) and finally estimated for the total, excluding Frontiers (Column 5). As the descriptive overview revealed that there could be a nonlinear relationship between the propensity to suspected predatory publishing on the one hand and the level of economic development as well as the size of the research sector on the other hand, we test for this possibility by including the respective variables in squared terms.

GDP per capita has a significantly positive main effect, but the negative squared term indicates that there is indeed an inverse U-shaped relationship. The results confirm that the proclivity to suspected predatory publishing has a tendency to increase with the level of economic development, but only up to a point, after which the relationship turns negative. Hence, countries at a medium level of development are the most vulnerable. Likewise, the size of the research sector comes out with a significantly positive main effect and a negative squared term; thus the same interpretation applies, albeit the relationship is estimated to be far less curvilinear¹⁸.

Some of the control variables prove to have even more statistically significant coefficients. First, more reliance on oil and natural gas rents, which, *ceteris paribus*, loosens the fiscal constraints of governments, is strongly positively associated with suspected predatory publishing. Of course, this is not to say that such resources should not be used to fund research, but there is a catch. Second, Arabic countries are confirmed to be particularly prone to suspected predatory publishing, even after oil and natural gas rents and other factors are accounted for, so there is something special about this area. Further, English is assumed to primarily control for the suspected language bias of Beall's lists and Scopus, but this worry is not supported by the results. Finally, longitude has a significantly positive coefficient, so being farther east of the Greenwich meridian implies higher inclinations towards suspected predatory publication.

As far as the comparison by source list is concerned, the results confirm that Frontiers has a different *modus operandi* than the rest of the pack. If only articles in Frontiers journals are considered, for instance, GDP per capita has statistically significant but opposite signs from the benchmark results. In fact, the model explains this outcome variable quite poorly, from which it follows that a different approach is needed to get to the bottom of what is up with this publisher. Although there is no evidence in the data presented upon which we can judge whether the inclusion of Frontiers on Beall's list was justified or not, the results at the very least clearly indicate that Frontiers is atypical. Henceforth, therefore, we focus on the outcomes excluding Frontiers¹⁹.

Figure 2 gives graphical representations of the estimated relationships of main interest, which provide a handy platform for discussing the results in more detail. The figures clearly illustrate that these relationships follow an inverse U-shaped curve. The propensity to suspected predatory publishing increases with GDP per capita up to approximately the level of countries such as India, Nigeria, and Pakistan, after which, however, there is a steep decline. Along the size measure there is initially a steady increase of suspected predatory publishing until a turning point at the level of countries with relatively large research systems such as

¹⁸ If the squared terms are excluded from the model, both coefficients come out highly statistically significant, but GDP per capita has a negative sign while the size of research sector has a positive sign.

¹⁹ It needs to be emphasized that the authors of this article have never had any connection to the Frontiers Research Foundation or any of their journals in any capacity.

Table 5. Explaining propensity to suspected predatory publishing; GLM with logit link for binomial family, 2015–2017

	(1)	(2)	(3)	(4)	(5)
	Total	Standalone	Publishers excl. Frontiers	Frontiers	Total excl. Frontiers
Constant	−6.405*** (0.877)	−11.227*** (1.941)	−7.690*** (1.393)	−5.991*** (0.778)	−7.936*** (1.270)
GDP per capita	0.308* (0.182)	0.838*** (0.284)	0.450 (0.285)	−0.301* (0.158)	0.535** (0.255)
GDP per capita squared	−0.100*** (0.034)	−0.296*** (0.068)	−0.149*** (0.054)	0.113*** (0.027)	−0.180*** (0.049)
Size of the research sector	0.405** (0.188)	1.042** (0.408)	0.446 (0.298)	0.174 (0.178)	0.588** (0.272)
Size of the research sector squared	−0.017* (0.010)	−0.050** (0.021)	−0.019 (0.016)	−0.005 (0.009)	−0.027* (0.015)
Oil and natural gas	0.019*** (0.005)	0.027*** (0.007)	0.023*** (0.007)	−0.011* (0.006)	0.024*** (0.007)
English spoken	−0.095 (0.115)	−0.171 (0.179)	−0.190 (0.178)	0.022 (0.114)	−0.183 (0.157)
French spoken	−0.088 (0.119)	−0.321 (0.234)	−0.179 (0.178)	0.245** (0.106)	−0.215 (0.173)
Spanish spoken	−0.145 (0.188)	−0.544 (0.408)	−0.481 (0.323)	0.246 (0.180)	−0.498* (0.280)
Arabic spoken	0.532*** (0.175)	0.648** (0.258)	0.681*** (0.215)	0.102 (0.124)	0.686*** (0.209)
Latitude	0.003 (0.003)	0.013** (0.006)	0.001 (0.004)	−0.001 (0.002)	0.003 (0.001)
Longitude	0.005*** (0.001)	0.005*** (0.002)	0.008*** (0.002)	−0.001 (0.001)	0.007*** (0.001)
Field of research	Included	Included	Included	Included	Included
AIC	153.03	60.38	108.63	70.63	125.86
BIC	219.72	127.07	175.32	137.32	192.55
Number of research fields	4	4	4	4	4
Number of countries	163	163	163	163	163
Number of observations	630	630	630	630	630

Note. Only countries with at least 30 total articles in the respective field of research are included. The dependent variable is the proportion of suspected predatory journal articles in total articles. Robust standard errors are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

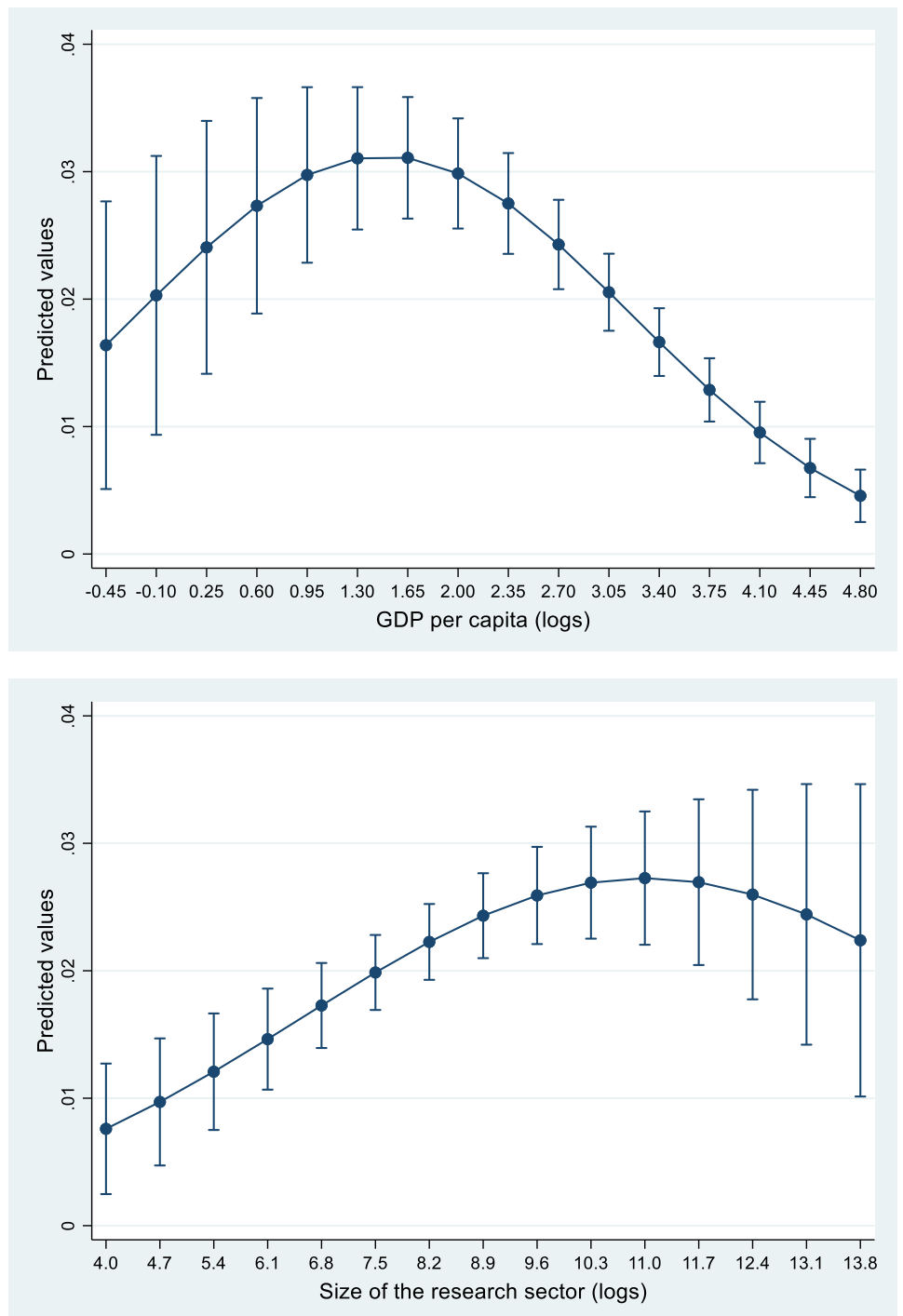


Figure 2. Estimated effects of GDP per capita (upper panel) and size of the research sector (lower panel) on the propensity to suspected predatory publishing (total excluding Frontiers); GLM with logit link for binomial family, 2015–2017. *Note.* Based on results in Column 5 of Table 5. Predictive margins with 90% confidence intervals are displayed.

Malaysia and Saudi Arabia, which is followed by only a slight decrease for the largest ones. The overlapping confidence intervals indicate that, for GDP per capita, the relationship differs most significantly between medium and highly developed countries, while for the size measure the difference is mainly between small and medium research sectors. So what does this mean?

GDP per capita is used for a lack of better measurements that are more intimately related to how a research system is organized and that would be available for a broad sample of countries, including many developing ones. Nevertheless, GDP per capita tends to be highly correlated to many other salient measures. What is likely to make the key difference between medium and highly developed countries that drives the results presented in this study is the capability to perform meaningful research evaluation, including advanced scientometrics and peer review of the actual content of published papers, that does not fall back on only counting the number of articles indexed in Scopus or elsewhere, regardless of quality and merit. If the government is not able to set the right mix of incentives to the public research sector, which is arguably very difficult even in advanced countries, those who do not shy away from predatory publishing have free rein.

Size is an important consideration, as noted above, because large research systems are more complex and therefore notoriously more difficult for governments to evaluate, manage, and steer than small systems. If two countries maintain equally primitive research evaluation frameworks, one with a large research sector composed of dozens of diverse institutions will tend to be more susceptible to predatory publishing than one with a tiny research sector composed of perhaps only a few easy-to-oversee workplaces. Large research systems suffer from a certain degree of anonymity, blind spots, and dark corners, in which predatory publishing flourishes. Around the turning point, however, the system becomes large enough to warrant investment in advanced research evaluation capabilities, which makes life more difficult for those exploiting the loopholes, so that the relationship between predatory publishing and size flattens and even curves slightly down.

6. CONCLUSIONS

Taken at face value, the evidence presented in this paper indicates that countries at a medium level of economic development and with large research sectors are most prone to publishing in suspected predatory journals. This should be a dire warning for developing countries that devote large resources to support research but which may not pay sufficient attention to upgrading their research governance capabilities, including their research evaluation framework. Moreover, the evidence suggests that oil-rich and/or Arabic and/or eastern countries tend to be particularly vulnerable, which completes the picture of who should be primarily on the lookout for predators.

Nevertheless, the general patterns are from a bird's-eye view, so there are exceptions driven by idiosyncratic factors. The prime example of an outlier appears to be Albania, which does not feature most of the high-risk characteristics but is still among the most affected countries. Predatory publishing is a truly global phenomenon, from which no emerging research system is entirely safe. Policy makers in developing countries that do not fit the description of the main risk group should not be fooled into thinking that the problem does not concern them, because if they flinch in their vigilance, their homeland may end up on the list of the most affected countries next time.

The results are broadly in line with previous estimates by Shen and Björk (2015), Xia et al. (2015), and Demir (2018), as well as Wallace and Perri (2018), in the sense that Asia and North

Africa provide the most fertile grounds for predatory publishing and that in particular India and Nigeria belong to the main sources. However, this paper has not only gathered one of the most comprehensive databases of suspected predatory journals, and used far more complete evidence than previous studies, but also provided a much higher level of granularity on the cross-country differences. In fact, a number of countries not mentioned in previous studies are shown here to be likely to suffer substantially from the problem of predatory publishing. In addition, this paper is the first to study the cross-country differences systematically in an econometric framework.

A major limitation of this study is that we can only speculate that the way in which research is evaluated in each country makes the primary difference, whether this includes research organizations at the national level, projects supported by funding agencies, and/or even individuals working on career progression. Ideally, we would like to take the characteristics of the research evaluation framework directly into account, including whether evaluation primarily concerns quantity or quality, whether formulae based on quantitative metrics are used, how advanced the underlying bibliometric approach is, whether insights from peer review assessment are factored in, and, consequently, what principles are applied when allocating research funding. Unfortunately, indicators of this kind are not available for more than a handful of advanced countries, which are not the most relevant here. To pin down the impact of these factors on the propensity to predatory publishing remains an important challenge for future research on this topic.

Another limitation is the cross-sectional nature of the analysis, which, as explained above, stems from the fact that historical data is not reliable. Longitudinal data would allow for more elaborate tests, particularly with respect to causality, than those employed in this paper. There are also likely to be lags in the cause–effect relationships that could be detected when long time series become available. In any case, the 3-year period studied here is rather short, as predatory publishing is a relatively recent and fairly dynamic phenomenon. This may have influenced the results and the list of most affected countries may look somewhat different if a similar exercise is repeated in a few years, which would be desirable.

It should be stressed that the results of this paper should not be interpreted to mean that developing countries should invest less in research, because this would undermine their emerging and often fragile national innovation systems and ultimately thwart productivity growth (Fagerberg & Srholec, 2009). However, it is fair to issue a cautionary note that predatory publishing has the potential to complicate research evaluation and therefore effective allocation of research funding greatly in many corners of the world. Developing countries aiming to embark on a technological catch-up trajectory need to take these intricacies more seriously than ever.

Scopus needs to stay focused on discontinuing coverage of questionable journals and most importantly step up efforts to prevent their indexing in the first place, as once they are allowed in, the content they publish before their potential discontinuation remains in the database forever. Scopus should strive to find a way to fact check whether the journal adheres to the declared editorial practices, most prominently how the peer-review process is performed in practice. It should possibly engage the research community more actively with regard to collecting data on complaints about publication standards in both currently and prospectively indexed journals. Unless the selection criteria are upgraded to reflect not only the declared policy but also reality on the ground and/or the bar for inclusion in terms of bibliometric criteria is raised significantly, new questionable journals will keep creeping in to the database, including rebranded and transformed business operations that have been flagged as predatory

in the past. In the meantime, evaluators, research managers, or university rankings that use Scopus data as inputs in their decisions need to be mindful about it.

Last, but not least, as already discussed above, Beall's lists no doubt have limitations. Beside the lack of transparency of the decisions to list some of the journals and publishers, the most obvious one has become the fact that Jeffrey Beall stopped curating the lists under his name at the beginning of 2017 (Straumsheim, 2017), as the result of which the data have become gradually outdated. Even though his lists continue to be maintained by someone on a new website (Anonymous, 2022), their updates have not been authorized, which makes their use problematic for research purposes. Future research on more recent evidence on this topic ought to look for different data sources. Cabells Predatory Reports, which have been developed in the meantime, seem promising (Cabells, 2022), but this list is proprietary and locked behind a paywall, and the database as a whole is not easily available²⁰. Clearly, the research community needs to continue efforts to improve identification, measurement, and understanding of the problem of predatory publishing.

ACKNOWLEDGMENTS

Earlier versions of the paper were presented at the IDEA think-tank seminar Predatory Journals in Scopus, Prague, November 16, 2016, the Scopus Content Selection and Advisory Board Meeting, Prague, November 3, 2017, and the 17th International Conference on Scientometrics & Infometrics, Rome, September 2–9, 2019. We thank the participants at these events as well as Ludo Waltman and Vincent Larivière, editors of *Quantitative Science Studies*, for their useful comments and suggestions. Martin Srholec also thanks his beloved wife Joanna for her support of the preparation of a revised version of the manuscript during the heat of the Covid-19 crisis. All the usual caveats apply.

The paper was first published in *Scientometrics* in February 2021 (Macháček & Srholec, 2021). Following pressure by Frontiers, the Editor-in-Chief of *Scientometrics* decided to retract the paper based on dubious claims that some of the findings are unreliable (Macháček & Srholec, 2022a). We refuted these claims and disagreed with this decision. The retraction was also strongly condemned by prominent members of the scientometric research community (Retraction Watch, 2021; Srholec, 2021). Later on, Akadémiai Kiadó and Springer Nature—the owner and publisher of *Scientometrics*, respectively—reverted to us the rights to publish the paper. The paper has undergone a minor revision before its republication in *Quantitative Science Studies*, mostly by extending some of the discussion and reflecting on the most recent development in this line of research but not by addressing the alleged flaws that were used to justify the retraction.

AUTHOR CONTRIBUTIONS

Vít Macháček: Conceptualization; Data curation; Methodology; Software; Validation; Visualization; Writing—original draft. Martin Srholec: Conceptualization; Formal analysis; Funding acquisition; Methodology; Validation; Visualization; Writing—original draft; Writing—review & editing.

Both authors contributed equally to this work.

²⁰ For example, in July 2021, we discussed with Cabells the possibility of obtaining access to an early version of the list that underlies their Predatory Reports for the purpose of running a replication study with regard to evidence presented in this paper and comparing the results. After a detailed discussion, however, Cabells concluded that this would not be desirable to do and the access has not been granted to us.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

Financial support from the Czech Academy of Sciences for the R&D&I Analytical Centre (RaDIAC) and from the Czech Science Foundation (GAČR) project 17-09265S is gratefully acknowledged.

DATA AVAILABILITY

Data files that provide a list of journals linked to Beall's lists indexed in Scopus (22_QSS_MachacekSrholec_Supp1.xlsx) and country-level data (22_QSS_MachacekSrholec_Supp2.xlsx) are available in Zenodo (Macháček & Srholec, 2022b). Scopus data for individual journals are proprietary and hence cannot be made available for legal reasons; these data can be accessed directly from <https://www.scopus.com>.

REFERENCES

- Anonymous. (2022). Potential predatory scholarly open-access publishers. <https://beallslist.net> (accessed June 24, 2022).
- Bagues, M., Sylos-Labini, M., & Zinovyeva N. (2019). A walk on the wild side: "Predatory" journals and information asymmetries in scientific evaluations. *Research Policy*, 48(2), 462–477. <https://doi.org/10.1016/j.respol.2018.04.013>
- Beall, J. (2013). Predatory publishing is just one of the consequences of gold open access. *Learned Publishing*, 26(2), 79–84. <https://doi.org/10.1087/20130203>
- Beall, J. (2015). Criteria for determining predatory open-access publishers. Available from <https://beallslist.weebly.com/uploads/3/0/9/5/30958339/criteria-2015.pdf> (accessed May 19, 2018).
- Beall, J. (2016). Scholarly open access: Critical analysis of scholarly open-access publishing (Beall's blog). Available from <https://scholarlyoa.com>; shut down January 2018, archived at <https://archive.org/web/> (accessed April 1, 2016).
- Berger, M., & Cirasella, J. (2015). Beyond Beall's List: Better understanding predatory publishers. *College & Research Libraries News*, 76, 132–135. <https://doi.org/10.5860/crln.76.3.9277>
- Bloudoff-Indelicato, M. (2015). Backlash after Frontiers journals added to list of questionable publishers. *Nature*, 526, 613. <https://doi.org/10.1038/526613f>
- Bohannon J. (2013). Who's afraid of peer-review? *Science*, 342(6154), 60–65. https://doi.org/10.1126/science.2013.342.6154.342_60, PubMed: 24092725
- Bolshete, P. (2018). Analysis of thirteen predatory publishers: A trap for eager-to-publish researchers. *Current Medical Research and Opinion*, 34, 157–162. <https://doi.org/10.1080/03007995.2017.1358160>, PubMed: 28722493
- Butler, D. (2013). Investigating journals: The dark side of publishing. *Nature*, 495(7442), 433–435. <https://doi.org/10.1038/495433a>, PubMed: 23538810
- Cabells. (2022). *Predatory Reports*. Cabells Scholarly Analytics. <https://www2.cabells.com/about-predatory> (accessed June 29, 2022).
- Cobey, K. D., Grudniewicz, A., Lalu, M. M., Rice, D. B., Raffoul, H., & Moher, D. (2019). Knowledge and motivations of researchers publishing in presumed predatory journals: A survey. *BMJ Open*, 9(3), e026516. <https://doi.org/10.1136/bmjopen-2018-026516>, PubMed: 30904874
- Cobey, K. D., Lalu, M. M., Skidmore, B., Ahmadzai, N., Grudniewicz, A., & Moher, D. (2018). What is a predatory journal? A scoping review. *F1000Research*, 7, 1001. <https://doi.org/10.12688/f1000research.15256.2>, PubMed: 30135732
- Crawford, W. (2014a). Ethics and access 1: The sad case of Jeffrey Beall. *Cites & Insights*, 14(4), 1–14.
- Crawford, W. (2014b). Journals, "journals" and wannabes: Investigating the List. *Cites & Insights*, 14(7), 1–24.
- Cyranoski, D. (2018). China awaits controversial blacklist of "poor quality" journals. *Nature*, 562(7728), 471–472. <https://doi.org/10.1038/d41586-018-07025-5>, PubMed: 30353153
- Demir, S. B. (2018). Predatory journals: Who publishes in them and why? *Journal of Informetrics*, 12(4), 1296–1311. <https://doi.org/10.1016/j.joi.2018.10.008>
- Demir, S. B. (2020). Scholarly databases under scrutiny. *Journal of Librarianship and Information Science*, 52, 150–160. <https://doi.org/10.1177/0961000618784159>
- Downes, M. (2020). Thousands of Australian academics on the editorial boards of journals run by predatory publishers. *Learned Publishing*, 33, 287–295. <https://doi.org/10.1002/leap.1297>
- Eriksson, S., & Helgesson, G. (2017a). The false academy: Predatory publishing in science and bioethics. *Medicine, Health Care, and Philosophy*, 20(2), 163–170. <https://doi.org/10.1007/s11019-016-9740-3>, PubMed: 27718131
- Eriksson, S., & Helgesson, G. (2017b). Time to stop talking about "predatory journals." *Learned Publishing*, 31(2), 181–183. <https://doi.org/10.1002/leap.1135>
- Erfanmanesh, M., & Pourhossein, R. (2017). Publishing in predatory open access journals: A case of Iran. *Publishing Research Quarterly*, 33, 433–444. <https://doi.org/10.1007/s12109-017-9547-y>
- Fagerberg, J., & Srholec, M. (2009). Innovation systems, technology and development: Unpacking the relationship(s). In B.-A. Lundvall, K. J. Joseph, C. Chaminade, & J. Vang (Eds.), *Handbook of innovation systems and developing countries: Building domestic capabilities in a global context* (pp. 83–115). Cheltenham: Edward Elgar. <https://doi.org/10.4337/9781849803427.00010>
- Frandsen, T. F. (2017). Are predatory journals undermining the credibility of science? A bibliometric analysis of citers.

- Scientometrics*, 113, 1513–1528. <https://doi.org/10.1007/s11192-017-2520-x>
- Gallup, J. L., Sachs, J. D., & Mellinger, A. (1999). *Geography and economic development*. CID Working Paper no. 1/1999, Harvard University. <https://doi.org/10.1177/016001799761012334>
- Good, B., Vermeulen, N., Tiefenthaler, B., & Arnold, E. (2015). Counting quality? The Czech performance-based research funding system. *Research Evaluation*, 24, 91–105. <https://doi.org/10.1093/reseval/rvu035>
- Grudniewicz, A., Moher, D., & Cobey, K. D. (2019). Predatory journals: No definition, no defence. *Nature*, 576, 210–212. <https://doi.org/10.1038/d41586-019-03759-y>, PubMed: 31827288
- Ibba, S., Pani, F. E., Stockton, J. G., Barabino, G., Marchesi, M., & Tigano, D. (2017). Incidence of predatory journals in computer science literature. *Library Review*, 66, 505–522. <https://doi.org/10.1108/LR-12-2016-0108>
- Kurt, S. (2018). Why do authors publish in predatory journals? *Learned Publishing*, 31(2), 141–147. <https://doi.org/10.1002/leap.1150>
- Macháček, V., & Srholec, M. (2017). *Predatory journal in Scopus*. IDEA study 2 / 2017. https://idea-en.cerge-ei.cz/files/IDEA_Study_2_2017_Predatory_journals_in_Scopus/mobile/index.html#p=1
- Macháček, V., & Srholec, M. (2021). Predatory publishing in Scopus: Evidence on cross-country differences. *Scientometrics*, 126, 1897–1921. <https://doi.org/10.1007/s11192-020-03852-4>, PubMed: 33583977
- Macháček, V., & Srholec, M. (2022a). Retraction Note to: Predatory publishing in Scopus: Evidence on cross-country differences. *Scientometrics*, 127, 1667. <https://doi.org/10.1007/s11192-021-04149-w>
- Macháček, V., & Srholec, M. (2022b). Predatory publishing in Scopus: Evidence on cross-country differences. *Zenodo*. <https://doi.org/10.5281/zenodo.7006761>
- Mayer, T., & Zignago, S. (2011). *Notes on CEPII's distances measures: The GeoDist database*. CEPII, Working Paper No 2011-25. <https://www.cepii.fr/CEPII/en/publications/wp/abstract.asp?NoDoc=3877>. <https://doi.org/10.2139/ssrn.1994531>
- Moed, H. F., Lopez-Illescas, C., Guerrero-Bote, V. P., & de Moya-Anegón, F. (2022). Journals in Beall's list perform as a group less well than other open access journals indexed in Scopus but reveal large differences among publishers. *Learned Publishing*, 35, 130–139. <https://doi.org/10.1002/leap.1428>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Perlin, M. S., Imasato, T., & Borenstein, D. (2018). Is predatory publishing a real threat? Evidence from a large database study. *Scientometrics*, 116(1), 255–273. <https://doi.org/10.1007/s11192-018-2750-6>
- Retraction Watch. (2021). Authors object after Springer Nature journal cedes to publisher Frontiers' demand for retraction. <https://retractionwatch.com/2021/09/07/authors-object-after-springer-nature-journal-cedes-to-publisher-frontiers-demand-for-retraction/> (accessed June 14, 2022).
- Sarant, L. (2016). The Middle East: An end to oil dependency. *Nature*, 537, S6–S7. <https://doi.org/10.1038/537S6a>, PubMed: 27580141
- Scopus. (2018a). *Scopus on-line database*. <https://www.scopus.com> (accessed March 19, 2018).
- Scopus. (2018b). *Scopus source list* (May 2018 version). Current version is available at <https://www.scopus.com/sources> (accessed May 21, 2019).
- Scopus. (2019). *Content policy and selection*. <https://www.elsevier.com/solutions/scopus/content/content-policy-and-selection> (accessed May 19, 2019).
- Scopus. (2021). *The importance of high-quality content: Curation and reevaluation in Scopus*. https://www.elsevier.com/_data/assets/pdf_file/0004/891058/ACADLIBSCARTImportance-of-high-quality-contentWEB.pdf (accessed June 13, 2022).
- Shen, C., & Björk, B.-C. (2015). “Predatory” open access: A longitudinal study of article volumes and market characteristics. *BMC Medicine*, 13(230), 1–15. <https://doi.org/10.1186/s12916-015-0469-2>, PubMed: 26423063
- Schmoch, U., Fardoun, H. M., & Mashat, A. S. (2016). Establishing a world-class university in Saudi Arabia: Intended and unintended effects. *Scientometrics*, 109, 1191–207. <https://doi.org/10.1007/s11192-016-2089-9>
- Shamseer, L., Moher, D., Maduekwe, O., Turner, L., Barbour, V., ... Shea, B. J. (2017). Potential predatory and legitimate biomedical journals: Can you tell the difference? A cross-sectional comparison. *BMC Medicine*, 15(1), 28. <https://doi.org/10.1186/s12916-017-0785-9>, PubMed: 28298236
- Siler, K., Vincent-Lamarre, P., Sugimoto, C. R., & Larivière, V. (2021). Predatory publishers' latest scam: Bootlegged and rebranded papers. *Nature*, 598, 563–565. <https://doi.org/10.1038/d41586-021-02906-8>, PubMed: 34703002
- Silver, A. (2017). Pay-to-view blacklist of predatory journals set to launch. *Nature News*. <https://doi.org/10.1038/nature.2017.22090>
- Somoza-Fernández, M., Rodríguez-Gairín, J.-M., & Urbano, C. (2016). Presence of alleged predatory journals in bibliographic databases: Analysis of Beall's list. *El Profesional de la Información*, 25(5), 730–737. <https://doi.org/10.3145/epi.2016.sep.03>
- Srholec, M. (2021). RETRACTION REBUKE: Predatory publishing in Scopus: Evidence on cross-country differences. <https://inovacnipolitika.blogspot.com/2021/09/retraction-rebuke-predatory-publishing.html> (accessed June 14, 2022).
- Straumsheim, C. (2017). No more “Beall's List”. *Inside Higher Ed*, January 18. <https://www.insidehighered.com/news/2017/01/18/librarians-list-predatory-journals-reportedly-removed-due-threats-and-politics> (accessed March 27, 2017).
- Strinzel, M., Severin, A., Milzow, K., & Egger, M. (2019). Blacklists and whitelists to tackle predatory publishing: A cross-sectional comparison and thematic analysis. *mBio*, 10(3), e00411-19. <https://doi.org/10.1128/mBio.00411-19>, PubMed: 31164459
- Ulrichsweb. (2016). *Ulrichsweb—Global Serials Directory*. <https://ulrichsweb.serialssolutions.com/> (accessed April 1, 2016).
- Wallace, F. H., & Perri, T. J. (2018). Economists behaving badly: Publications in predatory journals. *Scientometrics*, 115, 749–766. <https://doi.org/10.1007/s11192-018-2690-1>
- World Bank. (2016). *How does the World Bank classify countries?* <https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries> (accessed October 10, 2016).
- World Bank. (2018). *World development indicators* (last updated July 2018). New York: World Bank.
- Xia, J., Harmon, J. L., Connolly, K. G., Donnelly, R. M., Anderson, M. R., & Howard, H. A. (2015). Who publishes in “predatory” journals? *Journal of the Association for Information Science and Technology*, 66(7), 1406–1417. <https://doi.org/10.1002/asi.23265>
- Zhang, L., Wei, Y., Sivertsen, G., & Huang, Y. (2022). The motivations and criteria behind China's list of questionable journals. *Learned Publishing*, 35(4), 467–480. <https://doi.org/10.1002/leap.1456>

APPENDIX

Table A1. Descriptive statistics of the variables, 2015–2017

	Mean	St. dev.	Min	Max	N
<i><u>Dependent variables:</u></i>					
Total	0.028	0.039	0	0.370	630
Standalone	0.005	0.013	0	0.216	630
Publishers excl. Frontiers	0.016	0.033	0	0.370	630
Frontiers	0.007	0.009	0	0.057	630
Total excl. Frontiers	0.021	0.039	0	0.370	630
<i><u>Explanatory variables:</u></i>					
GDP per capita	2.341	1.211	-0.443	4.773	163
Size of the research sector	8.355	2.323	3.989	14.071	163
Oil and natural gas	2.676	7.124	0	48.318	163
English spoken	0.221	0.416	0	1	163
French spoken	0.129	0.336	0	1	163
Spanish spoken	0.123	0.329	0	1	163
Arabic spoken	0.117	0.322	0	1	163
Latitude	20.454	24.752	-41.814	67.470	163
Longitude	20.439	58.960	-112.10	177.97	163

Note. GDP per capita and the size of research sector in logs. N – number of observations.

Table A2. Definition and sources of the variables

Variable	Definition	Source
Predatory journal articles	The proportion of articles in journals linked to Beall's lists by authors from the respective country in total articles from that country recorded in the Scopus database.	Scopus (2018a)
GDP per capita	Gross domestic product (GDP) converted to constant 2011 international dollars using purchasing power parity (PPP) rates.	World Bank (2018)
Size of the research sector	Counts of total articles by authors from the respective country recorded in the Scopus database.	Scopus (2018a)
Oil and natural gas	The difference between the value of crude oil and natural gas production at regional prices and total costs of production as percentage of GDP.	World Bank (2018)
English spoken	Dummy with the value 1 if more than 20% of population speaks English.	Mayer and Zignago (2011)
French spoken	Dummy with the value 1 if more than 20% of population speaks French.	Mayer and Zignago (2011)
Spanish spoken	Dummy with the value 1 if more than 20% of population speaks Spanish.	Mayer and Zignago (2011)
Arabic spoken	Dummy with the value 1 if more than 20% of population speaks Arabic.	Mayer and Zignago (2011)
Latitude	Latitude of country centroid measured in degrees from the equator, with positive values going north and negative values going south.	Gallup et al. (1999)
Longitude	Longitude of country centroid measured in degrees from the Prime Meridian with positive values going east and negative values going west.	Gallup et al. (1999)