



INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

$$\frac{n!}{(n-1)!} p^{m-1} (1-p)^{n-m} = p \sum_{\ell=0}^{n-1} \frac{\ell+1}{n} \frac{(n-1)!}{(n-1-\ell)! \ell!} p^{\ell} (1-p)^{n-1-\ell}$$
$$= p \frac{n-1}{n} \sum_{\ell=0}^{n-1} \left[\frac{\ell}{n-1} + \frac{1}{n-1} \right] \frac{(n-1)!}{(n-1-\ell)! \ell!} p^{\ell} (1-p)^{n-1-\ell} = p^2 \frac{n-1}{n} +$$

$$\frac{\ell!}{(n-1)!} p^{m-1} (1-p)^{n-m} = p \sum_{\ell=0}^{n-1} \frac{\ell+1}{n} \frac{(n-1)!}{(n-1-\ell)! \ell!} p^{\ell} (1-p)^{n-1-\ell} = p \frac{n-1}{n} \sum_{\ell=0}^{n-1} \left[\frac{\ell}{n-1} + \frac{1}{n-1} \right] \frac{(n-1)!}{(n-1-\ell)! \ell!} p^{\ell} (1-p)^{n-1-\ell} = p^2 \frac{n-1}{n} +$$

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
<http://ies.fsv.cuni.cz>

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
<http://ies.fsv.cuni.cz>

Disclaimer: The IES Working Papers is an online paper series for works by the faculty and students of the Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Czech Republic. The papers are peer reviewed. The views expressed in documents served by this site do not reflect the views of the IES or any other Charles University Department. They are the sole property of the respective authors. Additional info at: ies@fsv.cuni.cz

Copyright Notice: Although all documents published by the IES are provided without charge, they are licensed for personal, academic or educational use. All rights are reserved by the authors.

Citations: All references to documents served by this site must be appropriately cited.

Bibliographic information:

Macháček M. and Srholec M. (2019): "Predatory Publications in Scopus: Evidence on Cross-Country Differences" IES Working Papers 20/2019. IES FSV. Charles University.

This paper can be downloaded at: <http://ies.fsv.cuni.cz>

Predatory Publications in Scopus: Evidence on Cross-Country Differences

Vít Macháček^{a,b}

Martin Srholec^b

^aInstitute of Economic Studies, Faculty of Social Sciences, Charles University
Opletalova 26, 110 00, Prague, Czech Republic

^bCERGE-EI, a joint workplace of Charles University and the Economics Institute of
the Czech Academy of Sciences, Politických vězňů 7, Prague 1, 111 21, Prague,
Czech Republic

Email (corresponding author): vit.machacek@cerge-ei.cz

July 2019

Abstract:

The paper maps the infiltration of so-called “predatory” scholarly journals into the citation database Scopus. Using the names of “potential, possible, or probable” predatory journals and publishers on Beall’s lists, we derived ISSNs of the respective journals from Ulrichsweb and searched Scopus with it. A total of 324 matched journals with 164 thousand documents indexed in Scopus over 2015-2017, making up a share of 2.8 % of the total articles have been identified. An analysis of cross-country differences in the tendency to publish in these journals reveals that overall the most affected are middle-income countries in Asia and North Africa. Kazakhstan is the country with the largest tendency to publish in predatory journals (18 %). More than 5 % is reported in 20 countries, including large countries such as Indonesia (18 %), Malaysia (11 %), India (10 %), or Nigeria (7 %). Neither developed countries are resistant to predatory publishing. More than 16 000 “potentially predatory” articles were published by authors from United States (0.67 %).

JEL: I28, I29, O38

Keywords: Predatory journals, Beall’s list, open access, academic misconduct

Acknowledgements: Financial support from the research programme Strategy AV21 of the Czech Academy of Sciences and from the Czech Science Foundation (GAČR) project 17-09265S is gratefully acknowledged. Earlier versions of the paper were presented at the IDEA think-tank seminar Predatory Journals in Scopus in Prague, November 16, 2016 and on the Scopus Content Selection and Advisory Board Meeting in Prague, November 3, 2017. We thank the participants at these events for their useful comments and suggestions. All the usual caveats apply.

1. Introduction

The business model of so-called “predatory” scholarly journals is based on a paid open-access (OA) publication model: the publisher does not charge for a subscription but receives money directly from the authors for publication of the article. As a result, the content is accessible for free to anyone. However, the predatory practice also entails a conflict of interests that has the potential to undermine the credibility of scholarly publishing. Authors are motivated to pay to have their work published for the sake of evaluation and career progression. In return, fraudulent publishers turn a blind eye to limitations of the submitted papers during peer-review. Predators' primary goal is to generate income from authors' fees. The worst of them fake peer-review and print anything for money, without scruples.

Why do researchers participate and offer their publications to predatory journals? Some of the predatory publication can be attributed to the low experience of young researchers (Xia et al. 2015). They send their texts in good faith that their text is going to be properly processed, but get cheated by the fraudulent journals. However, predatory publications can also be a result of researchers' strategic behavior (Bagues et al. 2018; Kurt 2018; Demir 2018a). Authors can send their manuscripts to predatory journals to make their scientific results look better and are willing to pay the price of participating in the fraudulent scheme. If the local research environment accepts such results as a piece of solid scientific work and can help the researcher to climb the ladder in the hierarchy, the motivation to pay for publishing the pseudo-scientific results grows.

Predatory publishing can be seen as wasting of resources. Shen and Bjork (2015) estimate the size of the predatory market to 74 million USD in 2014, and the figure might have grown since. However, these are only direct costs associated with Article Processing Costs (APCs). Perhaps more important than direct costs are the indirect opportunity costs. The opportunity to flaw the

standard peer-review process distracts the concentration of researchers. Instead of fully focusing on generating the internationally competitive science, the researchers are motivated to write bogus pseudo-science that only pretends scientific relevance. If this happens on a massive scale on the national level, the efficiency of the local research evaluation system is questioned. The fact that research is often funded from the public sources only amplifies the relevance of these concerns.

This study aims to comprehensively and systematically map the penetration of predatory journals in the Scopus citation database and present representative evidence on cross-country differences in the propensity of scholars to publish therein. The cross-country differences can be useful for understanding the phenomenon of predatory publishing itself and for delivering effective means to fight it. Demir (2018a), Xia et al. (2015) and Shen and Bjork (2015) were analyzing the country distribution of authors in predatory journals. However, no study has measured the penetration of research systems by predatory journals. The results suggest which countries and regions should concern about the effectivity of their research systems and consider reforms enforcing publications comparable with international standards.

The first section explains the phenomenon of predatory publishing and reviews the existing literature studying the content of the Beall's lists. The second section then explains how the dataset has been constructed and elaborates on the limitations of the data. The third section provides an exploratory analysis of the differences across countries and of the geographical distribution of tendency to publish in predatory journals. The concluding section summarizes the key findings and pulls the strands together.

2. **Predatory journals, Beall's lists, and literature**

Predatory journals

Jeffrey Beall coined the term predatory publishing on his blog (Beall 2016). It is used to describe the practice of abusing paid OA scheme of scientific publishing. In difference to standard subscription-based models, the authors publishing via paid OA interact directly with the publishing houses. They pay Article Processing Costs directly to the publisher of the journal. Hence both the author and the publisher are financially motivated in the publication of articles. Predatory journals are such that perform only vague, formal, or even none peer-review and allow for publication of pseudo- or even non-scientific results (Bohannon 2013, Butler 2013). Other attributes of predatory journals such as aggressive marketing practices, fake members of editorial boards or amateur business management are often mentioned (Beall 2015, Cobey et al. 2018, Eriksson and Helgesson 2017a). However, these are only side-effects of the core defining elements. We use the term predatory journals to mark journals that are suspect of abusing paid open-access to attract fees from authors with a significantly flawed editorial practice that fails in enforcing the minimum scientific quality requirements.

The OA model, while being one of defining elements of predatory journals, is not at fault per se. The inherent conflict of interest does not necessarily have to be exploited. The world research infrastructure already offers effective means to ensure the quality of the editorial practice of the journal. Databases dedicated to supporting open-access such as DOAJ are already trying to develop operational mechanisms guaranteeing quality and employing transparency measures such as open peer-review can easily detect the fraudulent publishers. Journals not performing peer-review would have nothing to reveal. The existence of fraudulent journals does not mean that the movement calling for democratizing communication of science should be abandoned.

The predatory journals are not the exhaustive way the bogus science is published. There is no guarantee that the journal based on the standard subscription model will be of high quality. In these journals, the editorial practice can be flawed and corrupted for example by exploiting personal links. However, the direct financial interaction between the journal and authors allows the OA journals to bend editorial practice more easily than in the standard subscription model. Moreover, up to our knowledge, there is no list of bogus subscription-based journals available for exploring.

We believe that the tendency to publish in predatory journals is related to the quality of research evaluation in the country. In countries where the culture of evaluation forces researchers to publish in respectable journals, the motivation to publish in predatory journals is rather low if any. Publication in a dubious predatory journal can harm the researcher's reputation instead of boosting it.

Harder than to define the predatory journal is to recognize it in practice. There is no clear boundary between journals that take their editorial practice seriously and those that are just a vehicle for exploiting publication fees. Most often, black-lists are used to identify suspect predatory journals. The prime example is already mentioned Jeffrey Beall's blog (Beall 2016), however it was shut down at the beginning of 2017 (Straumsheim 2017). On the new website <https://beallslist.weebly.com/> anonymous authors continue in the Beall's work and regularly update his list. New black-list is offered by a private company Cabell's (Silver 2017), but its costs do not allow for in-depth study. Also, China has recently announced the formation of a blacklist of 'poor quality' journals (Cyranski 2018).

The inclusion of individual journals into the black-list should be based on rigid criteria. Beall revealed the list of criteria that he admittedly uses for decision over journals and publishers (Beall 2015). Also, Eriksson and Helgesson (2017a) and Cobey et al. (2018) suggest a list of

characteristics to identify predatory journals that resembles the Beall's criteria. We will concentrate on Beall's criteria as his list is used in our analysis. One set of indicators point directly to the credibility of editorial practice: ("*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*"). Another group of indicators is indirect as they do not concern editorial policy, but other measures of 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*").

The Beall's lists

Jeffrey Beall maintained two regularly updated lists of "potential, possible, or probable" predatory journals and publishers, henceforth for the sake of brevity referred to as "predators" only: i) a "*list of standalone journals*", which contains individual journals suspected of predatory practices; and ii) a "*list of publishers*", which highlights questionable publishers, most of which print multiple journals. Beall's blog went off-line on January 15th, 2017 (Straumsheim 2017).

Several studies investigated the content of the Beall's list. Crawford (2014b) went through every single item on Beall's lists (in late March and early April 2014). He found 9,219 journals in total, of which 320 were from the list of standalone journals and 8,899 from the list of publishers. Between 2012 and 2014 almost 40 % of those journals published fewer than four articles or none, in other words, were empty shells, and that a further 20 % published only a handful of articles. Another 4 % consisted of dying, or dormant journals with a quick drop to a

few articles in 2014, and 6 % were unreachable (the web link was broken, for instance). Hence only approximately one-third of the identified journals published articles regularly.

The lack of trustworthiness of Beall's listed journals and gaps in their editorial practice has also been demonstrated in Shamseer et al. (2017). The systematic comparison of the predatory journals with "ordinary" OA journals and subscription-based journals proved that predatory journals much more often than other contained spelling errors, but also promoted bogus bibliometric metrics on their website. The authors also had much more often trouble to verify the connection of editorial board members with the journals. Famous field experiment by Bohannon (2013) proved the flawed editorial practice by submitting fake scientific articles to 304 predatory journals. More than half of them accepted the fake article for publication. Majority of Beall's listed journals raise very serious doubts about the quality of their editorial practice.

Only a few studies attempted to estimate the regional distribution of authors in predatory journals. Shen and Bjork (2015) analyzed a stratified sample of 613 titles. Xia et al. (2017) analyzed a sample of 68 journals in the field of biomedical science. Demir (2018a) went through websites of 832 journals on the list of standalone journals and identified 24,840 publications published in 2017. All papers point towards South Asian and African countries as the primary origin of authors in predatory journals. Neither developed countries are resistant. A recent study of Italian publications showed that 5 % of researchers in Italian academia have published in journals included on the Beall's list (Bagues et al. 2019). Also, Demir (2018a) found many authors originating from the USA or Europe. However, these studies provide only rough estimates.

Kurt (2018) identified 4 topics that are repeatedly used to justify publication in predatory journals: "*social identity threat, unawareness, high pressure, and lack of research proficiency.*"

The common denominator of these reasons is emergency – authors do not publish in these journals because they want to. They either blame their inexperience or institutional pressure. The researchers publish in these journals because they are afraid of discrimination from the “traditional” journals or due to lack of experience. The justification of predatory publishing is a complex mix of personal motivations, experience, quality, but last but not least, the research evaluation that researchers face (Demir 2018a, Bagues, et al. 2019).

Previous papers assessed the journals' perspective of the distribution of authors. They only ask where the journal authors originate from. We add a systemic perspective by including the number of publications in the country. As such, we can assess the degree to which the particular country is hit by the phenomena of predatory journals. This information can be very valuable for researchers, research managers, and policy-makers in a particular country. However, up to our knowledge, only Perlin (2018) measured the penetration of Brazil by predatory journals.

Caveats of Beall's lists

As Eriksson and Helgesson (2017b) put it, “the term ‘predatory journal’ hides a wide range of scholarly publishing misconduct.” Some are truly fraudulent, while many others may be on the margin. However, the form of Beall's list forces us to work with a binary classification only, in which a journal is labeled either predatory or not. As Beall did not explain his decisions systematically, it is not possible to make any quantification of “predatoriness” of the journal, even though elaborated criteria exist. For this study, the journal either is listed by Beall, or it is not.

Caution is warranted when working with Beall's list of publishers. Classifying the entire publishing house as predatory is a strong judgment, as it cannot be ruled out that alongside with truly fraudulent journals have also been blacklisted those that are fine. The list includes

publishers maintaining broad portfolios of dozens and even hundreds of journals so it might be that some of them strictly speaking do not deserve the predatory label and that using this list could result in an overestimation of the true “predators.” It is likely that the overwhelming majority of these journals are of poor quality, but this is not a crime per se. After all, many non-predatory but poor quality journals have also found their way into the main citation databases. One must, therefore, keep in mind that the list of publishers is a relatively rough 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 label. Beall's predatory mark implies at the very least doubts about the publisher's internal quality assurance mechanisms. Most often, the publishers listed by Beall is raising serious doubts about the professionalism of the journal's editorial practice.

The strong criticism of Beall's list emphasizes the low transparency of the decision-making process (Berger and Cirasella 2015; Crawford 2014a; Bloudoff-Indelicato 2015). Although the criteria are public, the justification of the decision over individual journals and publishers is often not clear and hardly verifiable. While on some important cases Beall published at least a few sentences on his blog or Twitter, very often the journal or publisher is only added to list, without any reasoning. The lack of comprehensive, rigid, and formal justification of the Beall's judgment is a major drawback of his list.

The largest controversies were triggered by the inclusion of the Frontiers Research Foundation in October 2015 on the list of publishers. According to critics of this decision, the Frontiers publisher is “legitimate and reputable and does offer proper peer-review” (Bloudoff-Indelicato 2015). Frontiers journals are quite far from a typical predatory outlet with negligible relevance

for the scientific community. The citation rates rank them fairly well. SJR index reveals that Frontiers is much more cited in comparison with other journals listed by Beall's. Most Frontiers journals are ranked in the first quartile (Q1) in at least one field, according to Scimago SJR (Scopus 2018). Only 4 journals in Frontiers portfolio of 29 evaluated did not get into highest quartile. 3 of those ranked in Q2. Only one journal is ranked in the last quartile only. Most of their journals are also indexed in Web of Science and Directory of Open Access Journals. Beall, on the other hand, defended his decision by pointing out several articles that, according to him, should not have been published (Bloudoff-Indelicato 2015).

Another concern arises from the timescale. The predatory status is derived from the content of Beall's lists on 1st April 2016. Jeffrey Beall continuously updated his lists. However, the lists always reflect only the current status, with no indication of when the journal or publisher became 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 switched to the predatory regime only recently. In other words, it well might be that older articles in journals that are currently considered to be predatory may have gone through a standard peer-review. Hence, historical data must be used with caution.

Jeffrey Beall is American, and his lists are probably subject to English bias. The lists contain mainly journals that at least have their websites in English. In regions where a large part of scientific output is written in other languages - such as in Latin America, China or Francophone world - the estimates of the extent of predatory publishing based on Beall's list might be underestimated since Beall did not identify predatory journals in local languages. Since the majority of internationally relevant science is published in English, this does not constitute a severe problem for the estimation strategy. However, this bias has to be kept in mind when interpreting cross-country differences.

3. Database

Using the two-step methodology we have constructed to our knowledge most comprehensive database of all journals listed on Beall's list. Most empirical studies studying Beall's list focus on only a subset of journals (Shen and Bjork 2015; Xia et al. 2017). Demir (2018a) analyzed the whole list of standalone journals but neglected the list of publishers. The content of Beall's list of publishers is rather neglected with the exception of Perlin et al. (2018) who constructed a database of 1100 ISSNs from both the list of publishers and list of standalone journals using an automatic website crawler.

Scopus, rather than Web of Science, is used because it covers substantially more journals (Mongeon and Paul-Hus 2016) and it is more susceptible to predators (Demir 2018b; Somoza-Fernández 2016). Journals indexed in Scopus should fulfill minimum quality criteria, which among other things include performing a peer-review (Scopus 2019). Scopus-listed journals are also considered to be trustworthy by research evaluation and funding systems in many countries, such as, for example, the "coffee grinder" in the Czech Republic Good et al. (2015) or in Italy (Bagues et al. 2019). If a predatory journal becomes indexed in Scopus, the motivation to publish in it gets a clear boost for scholars. The journals indexed in Scopus might represent only a tip of the iceberg of predatory publishing, but it is the most dangerous part as they are successful in pretending scientific quality.

The literature suggests that only minor part of all the listed journals is expected to be indexed in Scopus, but yet it might be quite a substantial number of journals. The authors should most often originate from developing countries, especially from the countries with a strong influence of English culture (due to language bias). The list of standalone journals is expected to be more reliable than a list of publishers, but yet the list of publishers should follow similar patterns. We

also expect that the tendency to publish is somewhat related to the culture of research evaluation in the country.

In the first step of constructing the database, we built a comprehensive overview of journals suspected of predatory practices by matching the two lists by Jeffrey Beall with records in the Ulrichsweb database (Ulrichsweb 2019). The Ulrichsweb database is considered one of the most comprehensive lists of periodicals. The International Standard Serial Numbers (ISSNs) of these journals were then searched in Scopus, and their bibliometric data were downloaded. As such, not only a full list of predatory journals enlisted in Scopus but also unique data on the composition of authors publishing in these journals by the country of origin.

Beall's lists were downloaded on 1st April 2016. First, all search terms were identified in each item on the lists. For some entries, Beall presented more versions of its name, for example, the journal name and its abbreviation. All available versions were used as a search term. The identified search terms were searched for in the Ulrichsweb database on the same day, using an automatic script programmed in Python, while keeping the information about the item on Beall's list that is searched for. When searched for the standalone journal, the '*title*' field was used by the script. When searched for the publisher, the '*publisher*' field was used. The search term was always enclosed in quotes. In the end, the algorithm saved all search results. The search request in Ulrichsweb looked as follows for standalone journals:

```
+(+title:"Academic Exchange Quarterly")
```

And for publishers:

```
+(+publisher:"Abhinav")
```

Ulrichsweb search engine uses 'fuzzy' search and does not require perfect matching of strings. For example, when searched for publisher *Academe Research Journals*, also journals of

Academic Research Journals were found. This is beneficial because the search is robust to typos, interpunction signs, and small errors written in the search terms. However, it also requires careful manual verification of the search results.

The raw search on Ulrichsweb resulted in the database of 19 141 search results linked to individual entries on the Beall's list. In the next step, the results without ISSN were removed as they are most probably not listed in Scopus neither. 16 037 search results were left after removal, but those are only 7 568 unique ISSNs. This is due to the fuzziness of the Ulrichsweb search mentioned above and the searching for multiple search terms related to the same entry.

Further filtration was needed to exactly identify the journals truly listed by Beall. To resolve any inconsistencies, for example, due to inaccurate names, the remaining search results were checked manually. Beall's lists consist of hypertext links. To check our entries the ISSN we compared the ISSN on the journal's website with the ISSN found on Ulrichsweb. If the two ISSNs matched, the entry was retained; if they differed, the entry was removed from our database. Publisher's identity was confirmed if at least one ISSN listed on its web site was found in an entry linked to the publisher's name on Ulrichsweb. If a matching ISSN was found, all other journals of the same publisher were also retained in our database. If no matching ISSN was found, and the entries, therefore, referred to a different publisher, they were removed.

In total, we confirmed 4 665 ISSNs associated with Beall's lists. Many journals have dual ISSNs, one for a printed and one for an electronic version. The number of individual journals is 3 295, of which 310 feature on the list of standalone journals and 2 984 refer to his list of publishers. Additional 31 journals were both in the list of standalone journals and the list of publishers. For simplicity, we refer to journals from both lists as the journals from the list of publishers only from now on.

This is roughly in line with the previous analysis of Crawford (2014b). There are 1 003 hypertext links on the list of standalone journals, from which follows that more than two-thirds of them are not covered in Ulrichsweb, let alone in more selective databases, and so apart from the unverified information on their web pages, there is no information about them. There are many predatory journals not registered in existing databases.

Table 1. The Data Generation Process

1) Obtaining the “predatory” ISSNs

- a. Beall’s lists downloaded on the 1st April 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 ISSNs in Scopus

- a. The “predatory” ISSNs were searched for using an automatic script in Scopus in March 2018; the script downloaded the number of indexed documents and their distribution by author's country of origin.
- b. 439 ISSNs of 324 individual journals with at least one entry in Scopus were identified.
- c. To avoid double-counting of documents in journals with both printed and electronic ISSNs, the duplicates had to be eliminated.

The next step was to search for these “predatory” ISSNs in the Scopus citation database. Once again, this search was performed using an automatic script programmed in Python. The search was performed in March 2018. For each ISSN detected in Scopus, the script downloaded basic data on the number of documents in the “article” categories for the years 2015-2017, as well as

more detailed data on the number of these articles by author's country of origin. The search request in Scopus follows, where xxxx-yyyy stands for searched ISSN:

ISSN(xxxx-yyyy) AND DOCTYPE(ar) AND PUBYEAR > 2014 AND PUBYEAR < 2018

In total, 324 individual journals were identified (see the supporting information files for their full list with ISSNs). 287 are attributed to the list of publishers and 37 to the list of standalone journals. This means that almost 10% of the journals in our database had at least one entry indexed in Scopus. Scopus is not immune to infiltration by predatory journals.

The data on the size of the total research output is necessary to measure the penetration of predatory journals in individual countries. The search was done such that both the number of predatory articles in the country and the size of research output would be maximally comparable. So for all countries, the total number of articles in 2015 – 2017 was downloaded. The list of exact names of countries to download data for was taken from the previous search for predatory journals, supplemented with a few countries for which we did not find any predatory article. The search was performed using the following request:

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

The article, either predatory or not, is linked to the country when affiliation of at least one of his authors is in the country. The number of authors from individual country reported by Scopus does not consider the share of the author on the article. Both the article created by 5 authors from a single country and the one created by 5 authors from 5 different countries is counted as 1 article for each participating country.

The final dataset consists of the cross-section of almost all countries in the world. For each country both the total number of articles reported to Scopus, the number of articles belonging

to Beall's list of standalone journals and the list of publishers. To account for the Frontiers controversy described above and to allow for investigation of the structure of their authors independently, the articles related to the Frontiers publisher were excluded from the list of publishers and treated as independent.

4. Results

For the sake of simplicity, we refer to '*predatory articles*' as to all articles published in Beall's listed journals, i.e., the sum of articles from standalone journals, the journals from the list of publishers and Frontiers.

The analysis considers evidence over the period 2015-2017. As already noted above, using older data bears the risk that some of the journals currently featuring on Beall's lists were not yet predatory at that time. However, we use data from the last three years, rather than only the most recent year, to increase the number of articles available and the robustness of the results.

Out of more than two hundred countries, the data are available for, the dependent territories and countries with fewer than 300 thousand inhabitants were excluded. Among others, this concerns mainly several small island states in the Caribbean and Pacific, which are not the main focus of the analysis. As a result, the sample consists of 174 states, including a large number of developing countries, which altogether cover the overwhelming majority of the world's population and research activity.

In total, 164 073 predatory articles were identified in Scopus in 2015-2017, i.e., 2.8 %¹ of all articles added to Scopus in the respective period. 110 971 come from the list of publishers. Additional 30 867 is related to the publisher Frontiers. The list of standalone journals contains 22 235 articles indexed in Scopus. Hence the list of publishers is the dominant category, while the standalone journals are minor. For some articles, Scopus report the country of origin as *Undefined*. These are neglected in the analysis. However, the overwhelming majority of the articles found were still included in the cross-section analysis as only 1 069 predatory articles is reported as *Undefined* and more than 100 predatory articles are related to the countries excluded from the dataset. More than 99 % of predatory articles are included in the final analysis.

Table 2 presents the average tendency to publish in predatory journals for different country groups. The country figures are calculated as a simple share of articles linked to respective Beall's list on all articles reported to Scopus in the individual country. Countries are grouped by the income and regional categories. The income grouping is derived from the most recent World Bank (2017) classification. Only OECD countries are excluded as an independent group. The same source is also used for regional groups. Only *Europe & Central Asia* and *East Asia & Pacific* are both divided into two independent regions. Because the paper focuses on cross-country heterogeneity, the simple averages are used to calculate country groups. For example, Sri Lanka is weighted equally as India for within-group averages.

¹ 5 820 835 articles was published in Scopus globally in 2015-2017 (Scopus query DOCTYPE(ar) AND PUBYEAR > 2014 AND PUBYEAR < 2018)

Generally, the predatory articles are most often published by authors from middle-income and high-income countries outside the OECD. These include developing countries on the convergence path to technological frontier such as India, Indonesia, Nigeria, or Egypt. Some of these countries have more than 10 % of predatory articles. Frequent contributors to predatory journals often originate also from the high income countries that are not members of OECD. The average for this group is driven up by oil-rich countries like Oman, Brunei, Saudi Arabia, Bahrain, Kuwait, and the United Arab Emirates.

Table 2: The average tendencies to publish in predatory journals (% of all articles)

Country groups*	Average share of predatory articles on all articles in 2015-17				
	<i>Standalone</i>	<i>Frontiers</i>	<i>Other publishers</i>	<i>Total</i>	<i>Total less Frontiers</i>
<i>All countries in dataset</i>	0.46%	0.51%	1.59%	2.56%	2.05%
<i>Income groups*</i>					
OECD countries	0.15%	0.57%	0.67%	1.39%	0.82%
non-OECD countries					
High income	0.38%	0.43%	2.05%	2.86%	2,43%
Upper middle income	0.57%	0.55%	2.04%	3.16%	2,61%
Lower middle income	0.81%	0.45%	2.15%	3.40%	2,96%
Low income	0.16%	0.54%	0.82%	1.51%	0,97%
<i>Regions*</i>					
Middle East & North Africa	1.16%	0.49%	3.67%	5.32%	4.84%
Central Asia	1.00%	0.47%	2.86%	4.33%	3.86%
East Asia	0.62%	0.39%	3.01%	4.02%	3.63%
South Asia	0.62%	0.27%	1.81%	2.69%	2.42%
Sub-Saharan Africa	0.32%	0.57%	1.07%	1.96%	1.39%
Europe	0.34%	0.61%	0.98%	1.93%	1.32%
North America	0.03%	0.56%	0.55%	1.13%	0.57%
Latin America & Caribbean	0.10%	0.48%	0.55%	1.13%	0.65%
Pacific	0.04%	0.47%	0.38%	0.89%	0.42%

* *Within-group cross-border co-authorships are counted multiple times both in Scopus Articles and in Predatory Articles. The resulting share thus should not be systematically biased*

Surprisingly, the opposite end of the spectrum, with the lowest shares of predatory articles is also dominated by developing countries. Chad, North Korea, Afghanistan, or Haiti has less than 0.4 % of predatory articles. In a few cases, there are no authors publishing in predatory journals

whatsoever. In other low-income countries such as Niger, Ethiopia, or Zimbabwe, the tendency to publish in predatory journals is above the world average. However, these are still well below the most severely hit countries, and the size of research output in these countries tends to be rather small. In general, the predatory journals do not belong among major publications platforms in low income countries.

Developed countries are generally less prone to predatory publishing. Except for South Korea and Slovakia, all OECD countries have below average tendency to publish in predatory journals. For example, in the Netherlands, Switzerland, and Denmark, the tendency is below 1.5 %. Moreover, approximately two-thirds of predatory articles reported from these countries are published by publisher Frontiers. As expected, these countries still are not perfectly immune to the predatory phenomenon.

From the regional perspective, the countries severely hit are in the Middle East & North Africa, South Asia and Central Asia, where on average more than 3.5 % of predatory articles are published in predatory journals. Many countries on the convergence path such as India, but also oil-rich countries are in these areas. Sub-Saharan African countries, with few exceptions such as Nigeria or Sudan, do not belong among the most frequent contributors to predatory journals. Similarly, in Latin America and Pacific countries, the propensity to publish predatory articles tends to be low. The relatively low penetration of predatory articles, for example, in Latin America, but also in China or Francophone countries in Sub-Saharan Africa can probably be partly explained by the English bias of Beall's list.

The above described distribution of predatory publishing holds for both the list of standalone journals and the list of publishers. As they account for the majority of all predatory articles, the similar distribution is also apparent when analyzing all predatory articles. However, journals of the Frontiers publisher exhibit a different pattern. The tendency to publish is roughly constant

around 0.5 % across income groups. When neglecting for countries with small research sector (less than 1 000 articles in Scopus in total), the highest propensities to publish in Frontiers journals are in Luxembourg, Chile, Malta, Austria, Switzerland, Netherlands, Germany, Italy, Israel or Belgium. In all these countries, more than 1 % of the country's articles were published in Frontiers journals. In all these countries with the exception of Italy, this is much more than all other predatory journals together. From the geographical perspective, the Frontiers journals is not a typical predatory outlet.

The tendency to publish in predatory journals across countries is also visualized on the world map in figure 1. The Frontiers articles were neglected for the creation of this map as they are atypical. Only results related to the list of publishers, and the list of standalone journals are included. The darker the color, the higher the tendency to publish in predatory journals in the country. Note the scale of the map - almost every fifth article is marked as predatory in the darkest colored countries.

Fig. 1 Share of predatory articles on the total number of articles in 2015-2017



Source: Scopus, own calculation; the articles from the Frontiers publisher were excluded.

The map visually confirms that the most severely hit countries are concentrated in Asia and Northern Africa. 18 out of 21 countries with more than 5 % of predatory articles (without

Frontiers) are in this region. Also, Russia is not very far from the ‘*predatory heartland*.’ In Kazakhstan, Indonesia, Iraq, and Albania out of all articles reported to Scopus more than 10 % of relates to journals on Beall’s list (excluding Frontiers).

The very different picture appears when looking at the predatory articles in absolute terms and ignoring the size of the research sector. Although predatory articles consist only 3.6 % of all Chinese articles, the highest number of articles originate there – 42 434. Second-ranked India reported an additional 28 557 predatory articles. Together with USA (16 677 articles) and South Korea (12 486 articles), these 4 countries account for half of all identified predatory articles. The authors from developed countries are still very important costumers of predatory journals. More than two-fifths of all predatory articles have at least one author affiliated in OECD country. The figure decreases to approximately quarter when excluding Frontiers articles². Still, every fourth predatory article originates from one of the OECD countries. Many of them from countries such the United States, Germany, the United Kingdom, or South Korea.

5. Conclusions

The paper mapped the penetration of scholarly journals suspicious of using predatory practices in Beall’s lists into the citation database Scopus. Using the journal’s ISSNs derived from the Ulrichsweb database, we have identified a list of 3,295 journals with ISSN in the Bealls list.

² It is not possible to get the exact figure due to multiple counting problem when summing articles from different countries. The sum of all predatory article across all OECD countries is 45.1 % all predatory articles and 29.3 % of all articles without Frontiers

Approximately 90 % were indexed in the list of publishers that was rather neglected in the previous estimates.

Scopus search yielded 324 journals with at least one indexed article. Over the period 2015-2017, we identified 164 thousand documents in Scopus that were published in these journals, making up a share of 2.8 % of all Scopus-indexed articles in the respective period. The list of journals is available as supplementary material for the paper.

There is a remarkable heterogeneity in the share of articles published in predatory journals. In a few countries such as Kazakhstan or Indonesia, more than 17 % of the national output indexed by Scopus was marked by Beall as potentially predatory. In Ireland, Denmark, Switzerland or Sweden the same figure is less than 0.4 % (excluding Frontiers publisher). Research systems in the most severely hit countries should start tackling the problem as soon as possible since one of the channels of convergence seems to be significantly hindered.

The results suggest that the most fertile ground for the predators are not the poorest countries. The ideal client for predatory publishers is a researcher from countries developing their scientific infrastructure, but simultaneously struggling with the institutional environment. This might be the case of Indonesia (18 %), India (9 %), Yemen (8 %) or Nigeria (7 % of predatory articles excluding Frontiers). The research evaluation scheme must take the fraudulent behavior of researchers into account seriously when enforcing minimum scientific standards in these countries.

6. References

- Baguess 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
- Beall, J. (2015) Criteria for Determining Predatory Open-Access Publishers. Available from <https://beallist.weebly.com/uploads/3/0/9/5/30958339/criteria-2015.pdf> [cited 19 May 2018].
- Beall, J. (2016) Scholarly Open Access: Critical analysis of scholarly open-access publishing (Beall's blog) Available from <https://scholarlyoa.com>; [shutdown January 2018, archived at <https://archive.org/web/>].
- Berger, M., Cirasella, J. (2015) Beyond Beall's List Better understanding predatory publishers. *College & Research Libraries News*. 76: 132-135.
- Bloudoff-Indelicato, M. (2015) Backlash after Frontiers journals added to list of questionable publishers. *Nature*. 526;613.
- 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.
- 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 News*. Available from <https://www.nature.com/articles/d41586-018-07025-5> [cited 19 May 2018].
- Demir, S. B. (2018a). Predatory journals: Who publishes in them and why? *Journal of Informetrics*, 12(4), 1296–1311. doi:10.1016/j.joi.2018.10.008
- Demir, S. B. (2018b). Scholarly databases under scrutiny. *Journal of Librarianship and Information Science*, 0961000618784159.
- Eriksson, S., Helgesson, G. (2017a) The false academy: predatory publishing in science and bioethics. *Medicine, Health Care and Philosophy*. 2017;20(2): 163-170.
- Eriksson, S., Helgesson, G. (2017b). Time to stop talking about “predatory journals.” *Learned Publishing*, 31(2), 181–183. doi:10.1002/leap.1135
- Good, B., Vermeulen, N., Tiefenthaler, B., Arnold, E. (2015). Counting quality? The Czech performance-based research funding system. *Research Evaluation* 24: 91-105.
- Kurt, S. (2018). Why do authors publish in predatory journals?. *Learned Publishing*, 31(2), 141-147.
- Mongeon, P., Paul-Hus, A. (2016) The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*. 2016; 106(1): 213-228.
- 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.
- Scopus (2018) Scopus Source List (May 2018 version). Current version is available at <https://www.scopus.com/sources> [cited 21 May 2019]

Scopus (2019) Content Policy and Selection; Available from: <https://www.elsevier.com/solutions/scopus/content/content-policy-and-selection>; [cited 19 May 2019].

Scopus (2016). Scimago Country Rank; Available from: <http://www.scimagojr.com/countryrank.php>; [cited 10 October 2016].

Shen C, Björk BC. (2015) 'Predatory' open access: a longitudinal study of article volumes and market characteristics. *BMC Medicine*. 13(230): 1-15.

Shamseer, L., Moher, D., Maduekwe, O., Turner, L., Barbour, V., Burch, R., Clark, J., Galipeau, J., Roberts, J., Shea, B. J. (2017) Potential predatory and legitimate biomedical journals: can you tell the difference? A cross-sectional comparison. *BMC Medicine*. 2017; 15(28)

Silver, A. (2017) Pay-to-view blacklist of predatory journals set to launch. *Nature News*. Available from <https://www.nature.com/news/pay-to-view-blacklist-of-predatory-journals-set-to-launch-1.22090> [cited 19 May 2019].

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).

Straumsheim, C. (2017) No More 'Beall's List'. *Inside Higher Ed*. Available from: <https://www.insidehighered.com/news/2017/01/18/librarians-list-predatory-journals-reportedly-removed-due-threats-and-politics>; [cited 27 March 2017].

Ulrichsweb (2019) Ulrichsweb – Global Serials Directory. 2019; Available from: <http://ulrichsweb.serialssolutions.com/>; [cited 19 May 2019].

World Bank (2016) How does the World Bank classify countries? Available from: <https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries>; [cited 10 October 2016].

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.

Supporting information 1: A list of predatory journals in Scopus – file journals.xlsx

Supporting information 2: Raw country-level data – file data.xlsx

IES Working Paper Series

2019

1. Davit Maskharashvili: *Duopolistic Competition On a Plane*
2. Petr Hanzlík, Petr Teplý: *Key Determinants of Net Interest Margin of EU Banks in the Zero Lower Bound of Interest Rates*
3. Barbora Mátková: *Bank-Sourced Transition Matrices: Are Banks' Internal Credit Risk Estimates Markovian?*
4. Peter Kudela, Tomas Havranek, Dominik Herman, Zuzana Irsova: *Does Daylight Saving Time Save Electricity? Evidence from Slovakia*
5. Dominika Kolcunová, Simona Malovaná: *The Effect of Higher Capital Requirements on Bank Lending: The Capital Surplus Matters*
6. Jaromír Baxa, Tomáš Šestořád: *The Czech Exchange Rate Floor: Depreciation without Inflation?*
7. Karel Janda, Binyi Zhang: *Renewable Energy Financial Modelling: A China Case Study*
8. Anna Alberini, Olha Khymych, Milan Ščasný: *Estimating Energy Price Elasticities When Salience is High: Residential Natural Gas Demand in Ukraine*
9. Anna Alberini, Olha Khymych, Milan Ščasný: *The Elusive Effects of Residential Energy Efficiency Improvements: Evidence from Ukraine*
10. Jozef Baruník, Matěj Nevrla: *Tail Risks, Asset Prices, and Investment Horizons*
11. Barbora Malinska: *Realized Moments and Bond Pricing*
12. Hamza Bennani, Nicolas Fanta, Pavel Gertler, Roman Horvath: *Does Central Bank Communication Signal Future Monetary Policy? The Case of the ECB*
13. Milan Ščasný, Šarlota Smutná: *Estimation of Price and Income Elasticity of Residential Water Demand in the Czech Republic over Three Decades*
14. Mykola Babiak, Olena Chorna, Barbara Pertold-Gebicka: *Minimum Wage Increase and Firm Profitability: Evidence from Poland*
15. Martin Stepanek: *Sectoral Impacts of International Labour Migration and Population Ageing in the Czech Republic*
16. Milan Ščasný, Iva Zvěřinová, Alistair Hunt: *Nature-Based, Structural, or Soft Measures of Adaptation? Preferences for Climate Change Adaptation Measures to Limit Damages from Droughts*
17. Milan Ščasný, Iva Zvěřinová, Vojtěch Máca: *Consumer Preferences for Sustainable and Healthy Lifestyle: Five-Country Discrete Choice Experiments*
18. Jaroslav Pavlíček, Ladislav Kristoufek: *Modeling UK Mortgage Demand Using Online Searches*
19. Josef Bajzik, Tomas Havranek, Zuzana Irsova, Jiri Schwarz: *Estimating the Armington Elasticity: The Importance of Data Choice and Publication Bias*
20. Vít Macháček, Martin Srholec: *Predatory Publications in Scopus: Evidence on Cross-Country Differences*

All papers can be downloaded at: <http://ies.fsv.cuni.cz>.



Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz <http://ies.fsv.cuni.cz>