



Citation patterns between impact-factor and questionable journals

Emanuel Kulczycki¹ · Marek Hołowicki¹ · Zehra Taşkın¹ ·
Franciszek Krawczyk¹

Received: 18 January 2021 / Accepted: 26 July 2021 / Published online: 2 September 2021
© The Author(s) 2021

Abstract

One of the most fundamental issues in academia today is understanding the differences between legitimate and questionable publishing. While decision-makers and managers consider journals indexed in popular citation indexes such as Web of Science or Scopus as legitimate, they use two lists of questionable journals (Beall's and Cabell's), one of which has not been updated for a few years, to identify the so-called predatory journals. The main aim of our study is to reveal the contribution of the journals accepted as legitimate by the authorities to the visibility of questionable journals. For this purpose, 65 questionable journals from social sciences and 2338 Web-of-Science-indexed journals that cited these questionable journals were examined in-depth in terms of index coverages, subject categories, impact factors and self-citation patterns. We have analysed 3234 unique cited papers from questionable journals and 5964 unique citing papers (6750 citations of cited papers) from Web of Science journals. We found that 13% of the questionable papers were cited by WoS journals and 37% of the citations were from impact-factor journals. The findings show that neither the impact factor of citing journals nor the size of cited journals is a good predictor of the number of citations to the questionable journals.

Keywords Questionable journals · Beall's list · Cabell's list · Citation analysis · Predatory journals

Introduction

Journals as a key communication channel in science receive much attention from scholars, editors, policymakers, stakeholders and research-evaluation bodies because these publication channels are used as a proxy of the research quality of the papers published in them. Publishing in top-tier journals is perceived as the mark of the researcher's quality and productivity. In today's science, the best journals that make the research results and papers visible are mostly defined by their Journal Impact Factor (JIF) (Else, 2019), despite the many cases of its abuse. Impact-factor journals are perceived as journals that are legitimized by

✉ Emanuel Kulczycki
emek@amu.edu.pl

¹ Scholarly Communication Research Group, Adam Mickiewicz University, Poznan, Poland

experts in a given field; thus, papers cited by such journals are also valorised more, even though there might be various motivations behind the citations in question (Leydesdorff et al., 2016). The importance given to citation-based metrics used also for evaluation, scholars cite other not only to support, discuss or compare their research to others but also to impact citation numbers of their or other papers.

While the impact factor is often used to identify quality publications in the scientific community, low-quality publications and ethical issues in publishing have recently been discussed. Predatory publishing is one of the most discussed topics regarding journal publishing, which crosses over narrow fields of bibliometrics, scientometrics and academic-publishing studies. This topic related to publishing in the so-called questionable or low-quality journals attracts attention not only in academia but also outside it (Bohannon, 2013; Sorokowski et al., 2017). Predatory journals, accused of damaging science and diminishing the quality of scholarly communication and trust in science, are trying to be classified and listed. In recent years, the most famous attempt to list predatory journals was initiated by Jeffrey Beall, whose list (henceforth: Beall's list) gained attention from scientific fields and media. The second well-known approach is done by the company Cabell's International (henceforth: Cabell's list). Thus, journals listed on Beall's or Cabell's lists are called in this study questionable journals in contrast to reputable journals (listed in reputable international indexes like Scopus, Web of Science Core Collection [WoS] or the European Reference Index for the Humanities and Social Sciences). For many years journals which editorial practices were questioned were called "predatory" journals (Beall, 2012). However, following some recent studies (Eykens et al., 2019; Frandsen, 2019; Nelhans & Bodin, 2020), we will refer to such journals as questionable, in recognition of the fact that it is hard to adequately distinguish bad-faith outlets from low-quality ones.

Beall describes predatory journals as those ready to publish any article for payment and not following basic publishing standards, such as peer reviewing (Beall, 2013). The term 'predatory journals' produces tension because, as Krawczyk and Kulczycki (2020) show, from the beginning it accuses open publishing as being the key source of predatory publishing. Beall has published long lists of criteria used to create his list of predatory journals. However, these criteria were criticised as too subjective, and the whole process of creating of the list was criticised as non-transparent (Olivarez et al., 2018). According to Siler (2020), Cabell's list was created more transparently, and each journal was evaluated by 78 well-defined indices of quality. Different violations can be assigned to each journal, from 'minor' like poor grammar on the journal website to 'severe' like false statements about being indexed in prestigious databases like WoS (Siler, 2020). Moreover, each questionable journal is described alongside its violations of good publishing practices (not the case in Beall's list) (Anderson, 2017). The main disadvantage of Cabell's list is that access to it requires expensive subscriptions, so many scholars are unable to use it.

In this article, we investigate the visibility of questionable journals enabled by citations from journals indexed in WoS. We limited the focus of our study to social sciences journals because they were not so frequently studied by previous research and this study is a part of a larger project of content-based analysis of citations of the questionable papers in which we evaluate the citations in terms of their content (meaning, purpose, shape, array). As we study social sciences journals and we are social scientists ourselves, we have the necessary expertise in the field to understand context of citations (Cano, 1989). In the current study, we analyzed 3234 unique cited papers from 65 questionable journals from social sciences and 5964 unique citing papers (6750 citations of cited papers) from 2338 WoS journals (of which 1047 are impact-factor journals), which allows us to present characteristic of citing and cited journals.

This is the first extensive study looking at citations from WoS-indexed journals to papers published in questionable journals. We focus on WoS-indexed journals because WoS is still widely used in evaluation contexts, though some countries and institutions rely mainly on Scopus or national databases (Sīle et al., 2018). Moreover, in bibliometrics studies, WoS sustains a dominant position as the datasource (Zhu & Liu, 2020). Journals indexed in WoS are treated as reputable, and this credibility extends to all papers published in these journals. In terms of indexing questionable journals, Demir (2018b) pointed out a big difference between the largest citation databases: Scopus indexes 53 predatory journals from Beall's lists, but WoS indexes only three such journals and Somoza-Fernández et al. (2016) reported that this difference is smaller but still visible. This could also suggest a difference in citations to predatory journals in these databases, but previous attempts at analysing such citations were based mostly on Google Scholar data (due to an easier data-acquisition procedure) or Scopus. Björk et al. (2020) analysed Google Scholar citations of 250 articles from predatory journals. Nwagwu and Ojemeni (2015) analysed 32 biomedical journals published by two Nigerian publishers listed on Beall's list. Bagues et al. (2019) investigated how journals listed in Beall's list (in which Italian researchers published their works) are cited in Google Scholar. Oermann et al. (2019) analysed Scopus citations of seven predatory nursing journals. Moussa (2021), using Google Scholar, examined citations of 10 predatory marketing journals. Frandsen (2017) analysed how 124 potential predatory journals are cited in Scopus. Anderson (2019) in a blog post showed how seven predatory journals were cited in WoS, the ScienceDirect database and PLoS One.

The number of citations to the articles in questionable journals varies depending on the methodology used in the study. By using Scopus, Frandsen (2017) found 1295 citations to 125 predatory journals and Oermann et al. (2019) found 814 citations to seven predatory nursing journals that had published at least 100 papers each. When Google Scholar was used to analyse predatory journals, the number of citations to these journals was higher. Björk et al. (2020) found that articles in predatory journals receive on average 2.6 citations per article and 43% of articles are cited. Nwagwu and Ojemeni (2015) reported an average of 2.25 citations per article in a predatory journal. In contrast to other research, Moussa (2021) argues that predatory journals in marketing have a relatively high number of citations in Google Scholar, with an average of 8.8 citations per article.

However, even if we agree with all the aforementioned studies that papers in questionable journals have received much fewer citations than, for example, papers indexed in WoS, then the number of citations in legitimate journals to potentially non-peer-reviewed articles can still be substantial. Shen and Björk (2015) estimated that there were around 8000 active journals listed by Beall or published by publishers listed by him, and those journals published 420,000 articles in 2014. Demir (2018a) found 24,840 papers published in 2017 in 832 journals indexed in Beall's list of potential predatory journals. In February 2021, Cabell's Predatory Report indexes 14,224 journals (Cabells, 2021).

Although Frandsen (2017) argues that authors of articles citing questionable journals are mostly inexperienced researchers (i.e., those who have published less-than-average articles in Scopus) from peripheral countries, Oermann et al. (2019) presented that citing authors were most frequently affiliated, accordingly, to the USA, Australia and Sweden. Oermann et al. (2020) studied further articles in which predatory journals are cited and found that most of the citations are used substantively and placed in the introduction or literature review sections. Moreover, by analysing a small sample of the best-cited articles from predatory journals, Moussa (2021) found that around 10% of citing articles from Google Scholar were published in journals indexed in the Social Sciences Citation Index,

and Oermann et al. (2019) did not find any significant difference in citations in Scopus from impact-factor and non-impact-factor nursing journals.

We do not employ a normative approach, unlike other authors exploring this topic (e.g. Oermann et al., 2020). We analyse citations from indexes that are commonly accepted as ‘legitimate’ to journals perceived as questionable but we do not evaluate the quality or questionable behaviors of the journals. Therefore, we use the term ‘questionable’ and not ‘predatory’ journals. There is no agreed definition or common-sense method of distinguishing predatory journals from non-predatory ones (Cobey et al., 2018), although one can observe some common trends in defining predatory journals (Krawczyk & Kulczycki, 2020). In 2019, Grudniewicz et al. (2019) suggested the consensus definition of predatory journals that highlights self-interest of journals and publishers at the expense of scholarship. However, the idea of predatory publishing can be criticized from the perspective of geopolitics of knowledge production. Stöckelová and Vostal (2017) point out that instead of focussing on publishers, scholars should criticise the whole global system of modern scholarly communication. This is the tension between profit and non-profit journals or the academic North and South that enables operations of predatory journals and not only the ill will of some publishers.

Despite criticism of focusing on publishers and the term “predatory journal” itself, there is no doubt that journals listed as predatory are heavily criticised. Although limitations of Beall’s list were pointed out in the literature (e.g. Olivarez et al., 2018) and it stopped being updated in 2017, it was frequently used to study predatory publishing and to warn scholars against publishing in predatory journals. Krawczyk and Kulczycki (2020) found that during the years 2012–2018 researchers published 48 empirical studies based on Beall’s list. Moreover, the list was popularized by articles published in *Nature* (Beall, 2012) and *Science* (Bohannon, 2013). Thus, our study will focus on the conflict between tools to reveal the illegitimate nature of a journal (Beall’s and Cabell’s lists) and citations in a prestigious citation database (WoS), which is often seen as a source of scholarly legitimisation.

To explore this issue we will answer the following research questions: (1) What share of articles published in questionable journals is cited by WoS journals? (“A.1. Share of cited papers from questionable journals” section of the Results); (2) Which WoS journals cite questionable journals? (“A.2. Web of Science journals citing questionable journals” section); (3) Does Impact Factor correlate with the number of citations to questionable journals? (“A.3. Impact-factor journals citing questionable journals” section); (4) What share of citations to predatory journals are self-citations? (“PART B: self-citation analyses” section)).

Data and methods

Data sources

We have used two lists of questionable journals: Beall’s and Cabell’s lists. Beall created two lists: (1) List of Publishers and (2) List of Standalone Journals. We used only the List of Standalone Journals because after several years of not updating the List of Publishers, it would be extremely hard to reconstruct the original list of journals published by a given publisher included in the List of Publishers.

We have collected journals’ ISSNs from their websites and used the ISSN Portal to find variants of journal titles and ISSNs as well as provide data of the country of publishing.

The data on citations of papers published in selected journals were obtained from the WoS using the Cited Reference Search. We focussed on three main WOS products: Journal Citation Reports (JCR) based on the Science Citation Index Expanded (SCI-E) and the Social Sciences Citation Index, the Arts and Humanities Citation Index (A&HCI), and the Emerging Sources Citation Index (ESCI). We downloaded PDF files of either paper from questionable journals and WoS-indexed journals. Using the questionable journal websites, we collected the data on the number of papers published by each journal.

Selecting questionable journals

We have used two lists of questionable journals, Beall’s list of standalone predatory journals and Cabell’s Blacklist (now Cabell’s Predatory Reports), to select social-sciences journals for our analysis. The first one consisted of 1310 standalone journals, but we found that some of them were duplicates. The list was updated for the last time on 9 January 2017 and then was removed from Beall’s website. We have used the Wayback Machine (<http://web.archive.org>) to obtain this last version. The other one, Cabell’s list, might be perceived as a successor of Beall’s list. In January 2019, Cabell’s list consisted of 10,496 journals. We have decided to include in the analysis only active journals, which have been defined as journals that published at least one paper in each year of the 2012–2018 period, and their websites were active at the moment of the start of this study (May 2019).

The selection of journals was conducted in three steps and summarized in Fig. 1. In the first step, we started from the analysis of Beall’s list, as it was the very first attempt to list questionable journals and the key source for the discussion on predatory publishing. We found that 322 unique journals (24.6%) of 1310 journals from Beall’s list were active at the moment of our analysis. Based on expert decisions, two authors of this study classified journals according to the fields of science across seven groups: (1) Humanities [H], (2) Social Sciences [SS], (3) Hard Science [Hard], (4) Multidisciplinary 1 [scope covers H, SS, Hard], (5) Multidisciplinary 2 [scope covers SS, Hard], (6) Multidisciplinary 3 [scope covers H, SS] and (7) Multidisciplinary 4 [scope covers H, Hard]. We classified journals according to their titles, aims and scopes published on journals’ websites. The points of reference for assigning a particular journal to one of seven categories were the fields of science and technology in the Organisation for Economic Co-operation and Development

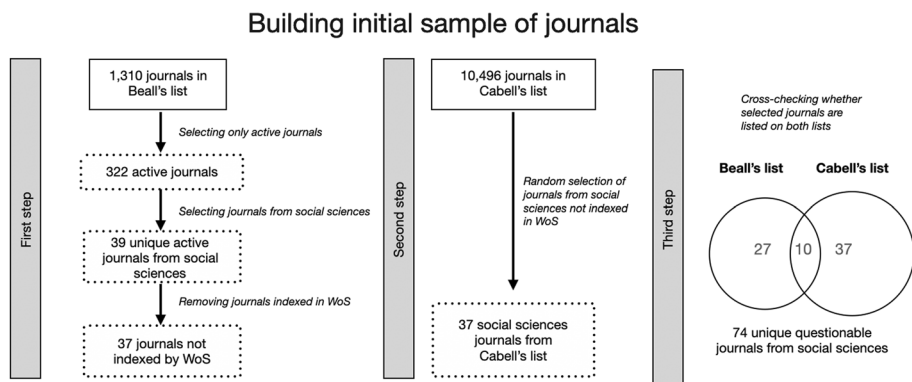


Fig. 1 Process of the selection of journals from Bell’s and Cabell’s lists

(2007). Finally, we decided to merge all multidisciplinary journals into a single category because the boundaries between some of them were blurry.

Therefore, in this study, a multidisciplinary journal is defined as a journal that publishes in at least two of three main fields, i.e., Humanities, Social Sciences and Hard Sciences. The number of active journals from Beall's list ($N=322$) across four fields is as follows: Hard Sciences (203), Humanities (2), Multidisciplinary (78) and Social Sciences (39). We have decided to exclude from the analysis journals that were indexed in WoS even if only for one year in at least one of three selected WoS products during the years 2012–2018 ($N=34$, 10.6%). The reason for this is to focus strictly on journals that were never legitimised by indexing in WoS. Moreover, citations from journals not indexed in WoS could not serve to play with the citation numbers and increase the impact-factor value of other journals. To prepare the final sample, we excluded two journals that were indexed in WoS and included all other active journals from the social sciences ($N=37$).

In the second step of the selection of journals, we decided to include 37 social-sciences journals from Cabell's list. The journals were assigned to the fields of sciences by their titles to provide complementary additions to the initial sample of social-sciences journals. Journals' websites were checked to confirm that every single journal meets the activeness criteria. Journals from Cabell's list were also checked in the ISSN Portal (variants of titles and ISSNs) as well as in the WoS (whether they were indexed in this database). All of those that were selected for the sample were never indexed in the WoS.

In the final step, journals selected from Beall's list were manually searched in Cabell's list. The same procedure was repeated for selected journals from Cabell's list, and we investigated whether they were in Beall's list or not. At this step, we had 74 unique journals in the sample.

Cited and citing papers

We prepared two datasets. The first one consists of the bibliographic data and PDF files of the papers published in the years of 2012–2018 in the social-sciences questionable journals included in this study and were cited by journals indexed in WoS (henceforth: cited papers). The other dataset consists of data on the papers (i.e., all journal publication types) published in journals indexed in WoS in the years 2012–2019 (henceforth: citing papers) that cited papers from the first dataset.

The steps presented in Fig. 2 allowed us to collect the data for both datasets. In the first step, we used the Cited Reference Search to search the cited references of WoS-indexed papers, in which authors referred to papers published in the analysed journals, which are sources not indexed in WoS. We took the titles of analysed journals and prepared 74 search queries with all gathered versions of titles for each journal. We found that papers from three analysed questionable social-science journals have never been cited in the analysed period. Our queries resulted in 4968 bibliographical records linked to the articles published by 71 questionable journals.

Secondly, we identified duplicated records of cited papers and merged them. After the removal of the duplicates, we obtained the dataset consisting of 4615 unique cited papers from 71 questionable journals and 8276 unique citing papers from 3347 WoS-indexed journals. Later, PDF files of cited papers were downloaded. Of the 4615 cited papers, 1204 (26%) were not able to be downloaded or they did not include the information about affiliations. The main reasons for not downloading a file were (1) missing files on the journals' website, (2) no information about a paper on a journal website and

Building final sample: datasets of cited and citing papers

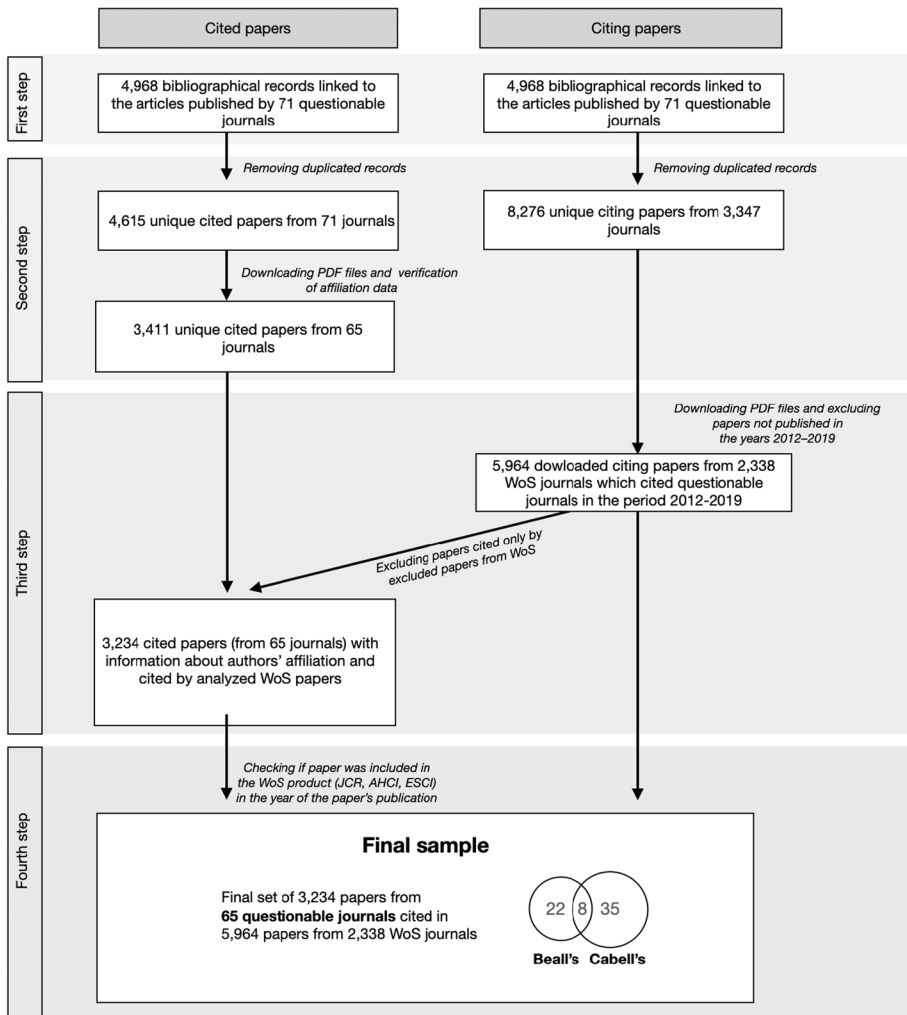


Fig. 2 The process of collecting papers used for citation analysis

(3) an archive on a journal website was inactive. We collected data on the affiliation of the corresponding authors of each cited paper from PDF files, where the corresponding author was explicitly indicated. If the corresponding author was not indicated, the affiliation of the first author was gathered. If the corresponding first author had more than one country of affiliation, then the first affiliation was gathered.

In a further analysis, we included 3411 cited papers from 65 questionable journals that have the PDF files downloaded, were cited at least once in the period 2012–2019 in journals indexed in WoS, and have information about the country affiliation of the corresponding author. To analyse what share of papers from a given journal has been cited

in WoS, we also counted the number of all papers published in a given journal in the analysed period.

Third, we collected the full records and bibliographic data from the WoS of all 8276 citing papers. We collected the affiliation of corresponding authors from the information about 'Reprint Address' included in the WoS data (only the first 'Reprint Address' was counted). Having collected all affiliations, we extracted the names of countries. The names of countries of the cited papers' authors were unified to match the list of countries used in the WoS. We excluded all citing papers that (1) were not published in the period 2012–2019, (2) have a different publication type in WoS than 'Journal' or (3) whose PDF file of the paper we were unable to download using either our institutional-access or open-access repositories. In a further analysis, we included 5964 citing papers. Moreover, 177 cited papers were excluded because they were cited only by excluded papers from WoS, so the final number of cited papers was reduced to 3234.

Fourthly, we collected for each citing paper the information regarding whether a journal of a citing paper was included in the WoS product (JCR, AHCI, ESCI) in the year of the citing paper's publication.

Finally, we analysed 3234 unique cited papers from 65 questionable journals and 5964 unique citing papers (6750 citations of cited papers) from 2338 WoS journals. The list of analysed questionable journals is in "[Appendix](#)" section.

Limitations of the study

Such decisions listed in the Methodology section such as focusing on social sciences, excluding inaccessible article files or journals are the main limitations of our study. However, the size of our last dataset provides sufficient information to reveal general citation patterns between legitimate and questionable journals, and these limitations do not affect the results negatively. Besides, another limitation is using affiliations of corresponding or first authors only. The main purpose of this choice is to collect the data for only the leading affiliations (countries) of each article. In this case, it is possible that some collaborations can be overlooked, as it is difficult to identify countries that have contributed equally. However, since our study does not aim to answer any questions on the collaboration of countries, this limitation can be ignored.

Results

PART A: questionable and Web-of-Science-listed journals

A.1. Share of cited papers from questionable journals

Between 2012 and 2018, 65 analysed journals published 25,146 papers, of which 3234 (13%) were cited by WoS journals.

Figure 3 shows the highest share, 19%, in 2012 and 2013. The mean number of papers published by a single journal in the analysed period was 53.5 (min = 43, max = 2176). On average, 11% of papers published by a journal were cited by WoS journals. The highest shares were found for journals that published 1748 and 259 papers. The shares are 36% (635 papers) and 35% (91 papers), respectively.

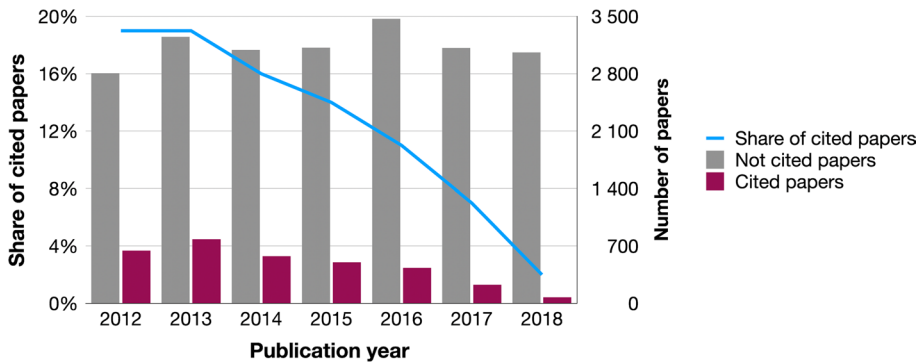


Fig. 3 Share and number of cited papers from questionable journals by WoS Journals

Table 1 shows the number and shares of cited papers according to lists of questionable journals. In the analysed sample, 8327 papers were published in journals listed in Beall’s list (10.3% of them were cited) and 13,910 papers from journals indexed in Cabell’s list (14.7% were cited). 5587 papers were published by journals indexed in both Beall’s and Cabell’s lists. The shares presented in the table may suggest that there is a significant relationship between the number of articles published by a journal and the share of cited papers. However, the Spearman correlation coefficient revealed that there was no significant relationship between the size of the journal and its share ($r_s = 0.019$).

We checked whether 65 analysed journals were covered by Scopus, as this could potentially increase the share of citations. We found that five of 65 were or have been covered: one has been covered by the whole analysed period (24% of papers were cited whereas in the whole sample 13% of papers were cited), one was removed from Scopus before the analysed period (23%), one was covered and removed in the analysed period (36% papers), one was covered in the last year of the analysed period (19%) and one was covered before the analysed period and removed during the period (8%).

A.2. Web of science journals citing questionable journals

We found that 2338 unique WoS journals cited 3234 questionable papers 6750 times. The number of citations per questionable journal from WoS journals is 2.88 (median = 1, minimum = 1, maximum = 218). The mean number of papers from WoS journals that cited analysed questionable papers was 2.88. Half of the citations were from 261 journals. Eighty-nine of 2338 journals cited at least 10 times papers from questionable journals, and four

Table 1 Number and share of cited papers in questionable journals ($N=65$) according to lists of questionable journals

List of questionable journals	Number of journals	Number of all papers	Number of cited papers	Share (%)
Beall’s list	22	8327	860	10.3
Beall & Cabell’s lists	8	5587	323	5.8
Cabell’s list	35	13,910	2051	14.7

WoS journals cited over 100 questionable papers. In the analysed period, one WoS journal published 183 papers, which cited questionable papers from our sample 218 times (all except one published in one questionable journal). One of these questionable papers was cited 36 times by this WoS journal.

We analysed in which WoS product (JCR, ESCI, A&HCI) a journal was indexed when a citing paper was published. We considered the publication year and whether a journal was included in the WoS product in the year in question. We found that 1152 of 2338 journals were indexed in ESCI, 35 in A&HCI, and 1047 in JCR. 104 journals that published papers that produces 366 citations were neither in ESCI, A&HCI nor in JCR, which means they were either in SCIE or SSCI indexes but not yet JCR (e.g., waiting for calculation of their impact factor) or dropped from the indexes because of quality issues or manipulations such as citation stacking or excessive self-citation rates. Figure 4 shows how questionable journals were cited by WoS journals. Of the 6750 citations, 2502 (37.1%) were from JCR journals, 3821 (56.6%) were from ESCI journals and 61 (0.9%) were from A&HCI journals. 366 (5.4%) citations were from journals indexed in SCIE or SSCI indexes but not yet JCR.

The questionable journals were selected only from social sciences; however, citations from all fields in WoS were considered for the evaluation of citations. The distribution of subjects to WoS subjects and broader OECD classifications (Clarivate Analytics, 2012) is shown in Fig. 5. The results are important because they prove the existence of citations from different subjects (such as medical sciences and agriculture) to social-science papers appearing in questionable journals.

Figure 5 shows that 25% of WoS citations were from one field, *Education and Educational Research*. This is followed by *Management* (8.9%) and *Business* (8.7%). In total, 63% of journals were classified in *Social Sciences*. According to the WoS classification of journals, a journal may be classified into two or more different subject fields. Six-hundred and sixty-five papers in 256 journals in our dataset were classified into two or more

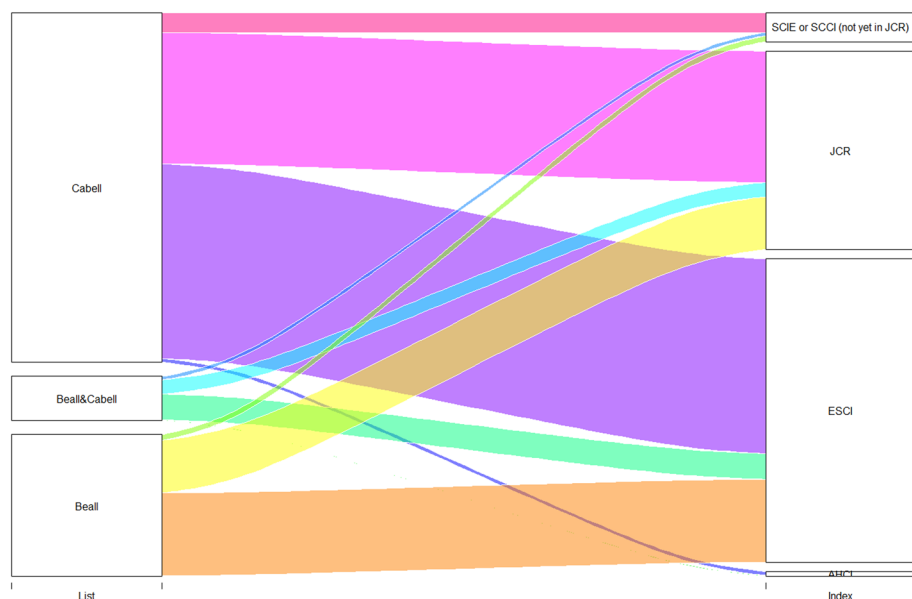


Fig. 4 WoS journals citing questionable journals according to list of questionable journals

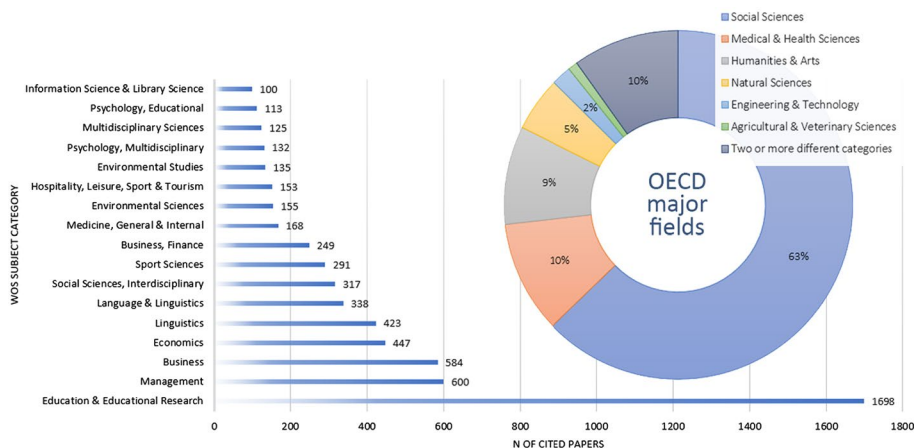


Fig. 5 Distribution of citing journals to WoS Subject categories and OECD major fields

different categories. The Kruskal–Wallis test results show that there are significant differences between the journals' OECD category and the number of citations to questionable journals ($\chi^2=75.641$, $df = 6$, $p<0.001$). When each subject category was compared to each other using the Mann Whitney U Test, the sources of differences were determined for Engineering and Social Sciences ($U=41,585.500$, $Z=-2.924$, $p=0.003$), Humanities and Medical Sciences ($U = 15,038.500$, $Z = -3.276$, $p=0.001$), Humanities and Social Sciences ($U = 92,069.500$, $Z=-2.192$, $p=0.028$), Medical Sciences and Natural Sciences ($U = 17,338.500$, $Z = -2.627$, $p=0.009$), Medical Sciences and Social Sciences ($U = 128,110.000$, $Z=-7.648$, $p<0.001$) and Natural Sciences and Social Sciences ($U = 96,527.000$, $Z = -3.310$, $p=0.001$).

A.3. Impact-factor journals citing questionable journals

The fact that a journal has a valid impact factor or is included in citation indexes is used by policymakers and managers to determine the level of that journal. We analysed the relationship between the impact-factor journals and their citations to questionable journals. To be able to make accurate statistical analyses, the impact factor of all journals cited in questionable journals were gathered with yearly changes. For example, if two articles in the same journal cited the questionable journals in 2018 and 2019, JCR 2017 and JCR 2018 were used. As a result, 1600 impact factors for 1047 IF journals were obtained.

Before presenting the comparisons between impact factors and citations to questionable journals, it is worth mentioning those journals dropped from citation indexes. Twenty impact-factor journals that cited questionable journals 125 times were dropped from JCR or WoS for different reasons. Fifteen of them were dropped from the index without listing any unethical concerns. This means that coverage of the journals did not meet the WoS selection criteria (Clarivate, 2018). *Scientific World Journal* was suppressed from JCR based on citation stacking and four journals (*Business Ethics: A European Review*, *Environmental Engineering and Management Journal*, *Eurasia Journal of Mathematics Science and Technology Education* and *Industria Textila*) were dropped for their excessive self-citation rates. These five journals cited questionable journals 39 times. Furthermore, although they were not indexed in JCR and did not have an impact factor, 15 journals were

excluded from ESCI after being indexed for a couple of years in ESCI. All these findings can be commented as questionable journals in WoS cited questionable journals. However, statistical tests did not confirm this comment: the Spearman's Rho correlation coefficient shows that the correlation between journal impact factors and the number of citations to questionable journals is very low, at a 99% confidence level ($r_s=0.090$, $p<0.001$). Also, according to the Kruskal–Wallis test results, the differences between journal impact-factor quartiles and the number of citations to questionable journals were not significant ($\chi^2=7.785$, $df=3$, $p=0.051$). However, the Mann Whitney U test revealed that the only differences were found between Q1 and Q4 journals' number of citations to questionable journals ($U=72,661.500$, $Z=-2.648$, $p=0.008$).

The impact-factor range of questionable journal citers is from 0 to 27.604 (mean = 1.689, median = 1.378, SD = 1.471, 25% = 0.745, 75% = 2.252), while the minimum impact factor of the whole JCR between 2011 and 2018 is 0 and the maximum is 244.585 (mean = 2.072, median = 1.373, SD = 3.310, 25% = 0.704, 75% = 2.462).

Eighty percent of the journals in JCR cited questionable journals only one time, and there is a significant difference between the impact factors of one-time citers and the others ($U=174,977.000$, $Z=-3.668$, $p<0.001$). However, the surprising result is that the average impact factor of journals that cited questionable journals more than once is 1.896 (median = 1.634), and this is higher than that of one time citers (mean = 1.639, median = 1.318).

Table 2 shows the main features of 30 impact-factor journals that cited questionable journals more than 10 times.

Having open access mega journals such as *SageOpen*, *IEEE Access* and *Plos One* on the list of the most cited IF journals (see Table 2) may suggest that open access mega journals cite questionable journals more frequently. However, the Mann Whitney U test result reveals that there is no significant relationship between open access feature of the journal (gold open or not) and the number of citations to the questionable journals ($U=75,233$, $p=0.988$). Similarly, the Spearman's Rho correlation coefficient shows that there is no correlation between journals' open access rates and the number of citations to questionable journals ($r_s=-0.065$).

All the test results on impact-factor journals prove that it is impossible to evaluate the questionable journals by looking at the impact factors or impact-factor percentiles of the journals because no pattern is identified. The impact factor is neither a descriptor of the quality of a paper nor the quality of citation. For example, *the Journal of Business Ethics*, which is one of the journals with the highest number of citations to questionable journals, is listed among the top 50 financial journals by the Financial Times (Ormans, 2016). On the other hand, *Journal of Cleaner Production*, which is also on the list, has been shown among the problematic journals by Clarivate Analytics due to its self-citation practices (Clarivate Analytics, 2021). Therefore, it reveals the importance of content-based analysis in understanding the purpose of citations to questionable journals.

Part B: self-citation analyses

We have analysed the countries of the corresponding authors of cited and citing papers. Table 3 presents the top 10 countries of cited and citing papers. Corresponding authors of papers published in questionable journals were most often from Turkey (335 papers), whereas citing papers were most often from the USA (555 papers).

Table 2 30 Impact-factor journals that cited questionable journals more than 10 times

JCR abbreviated title	IF 2019	No of articles citing questionable journals	N of cited papers	Publisher country	Journal self-citation rate (2018) (%)	Journal self-citation rate (2019) (%)
ACTA MEDICA MEDITERR	0.249	25	126	Italy	63.7	4.8
SUSTAINABILITY-BASEL	2.576	56	56	Switzerland	38.9	33.6
SAGE OPEN	0.715	16	45	USA	2.8	5.2
SYSTEM	1.979	31	34	England	12.4	16.6
FRONT PSYCHOL	2.067	24	26	Switzerland	15.5	13.7
EGIT BILIM	0.493	19	25	Turkey	14.0	11.6
PLOS ONE	2.740	25	25	USA	5.1	3.3
COMPUT EDUC	5.296	21	22	England	11.6	12.2
COMPUT HUM BEHAV	5.003	20	21	USA	11.4	8.3
INT J BANK MARK	2.800	11	21	England	33.7	26.6
COMPUT ASSIST LANG L	2.642	12	19	England	24.1	17.8
J BAL T SCI EDUC	0.915	16	18	Lithuania	40.8	17.6
EDUC TECHNOL SOC	2.086	14	17	Taiwan	8.9	1.9
BRIT FOOD J	2.102	13	15	England	20.3	26.9
J BUS ECON MANAG	1.640	14	15	Lithuania	5.2	12.7
J CLEAN PROD	7.246	13	13	USA	29.1	23.8
J BUS RES	4.874	10	12	USA	11.8	12.6
REV CERCET INTERV SO	0.736	12	12	Romania	12.0	15.6
AFR DEV REV	1.250	10	11	England	39.8	33.6
INT J EMERG MARK	1.022	10	11	England	68.4	45.8
J BUS ETHICS	4.141	11	11	Netherlands	9.7	9.2
LEARN INDIVID DIFFER	1.916	11	11	Netherlands	8.8	6.2
SOC INDIC RES	1.874	10	11	Netherlands	17.9	17.8
S AFR J EDUC	0.783	9	11	South Africa	27.8	31.9
TOURISM MANAGE	7.432	9	11	England	12.8	11.0

Table 2 (continued)

JCR abbreviated title	IF 2019	No of articles citing questionable journals	N of cited papers	Publisher country	Journal self-citation rate (2018) (%)	Journal self-citation rate (2019) (%)
ECON RES-EKON ISTRAZ	2.229	8	10	Croatia	5.4	11.3
IEEE ACCESS	3.745	9	10	USA	26.4	35.1
INT J ISLAMIC MIDDLE	1.206	9	10	England	6.3	13.4
KYBERNETES	1.754	10	10	England	24.3	14.6
STUD EDUC EVAL	1.983	9	10	Netherlands	14.3	8.5
TEACH TEACH EDUC	2.686	10	10	England	14.9	13.0

Table 3 Top 10 Countries of Cited and Citing Papers

Country of corresponding author of cited papers	Number of papers	Country of corresponding author of citing papers	Number of papers
Turkey	335	USA	555
India	259	Turkey	386
Malaysia	226	Malaysia	368
Iran	174	China	291
Nigeria	160	India	288
China	142	Spain	268
USA	136	Iran	257
Pakistan	120	Indonesia	188
Saudi Arabia	120	South Africa	184
Taiwan	117	Australia	172

We have calculated pairs of the most-often citing countries using the affiliation of corresponding authors. Figure 6 presents the top 50 pairs.

Italian researchers published the highest number of papers in WoS journals, which cited papers authored by Italian researchers in questionable journals. This result is not expected because previous studies have shown that Italian researchers did not publish intensively in questionable journals. Thus, we also conducted an author-level analysis of self-citations, which revealed that Italian researchers cited multiple questionable papers, of which they were co-authors.

In the analysis of self-citation on the author level, we have considered all the authors (not only the corresponding ones). By the analysis of bibliographical data and PDF files of the paper, we have found that 641 (9.5%) of 6750 are self-citations from 369 WoS journals to 53 questionable journals. The highest number of author self-citations from one WoS journal is 65 (all citations to one questionable journal; the corresponding authors of 55 from 65 of those WoS journal papers are affiliated in Italy). The highest number of author self-citations of questionable papers from one journal is 147.

Table 4 presents the top 10 self-citing authors from our sample. In total, we found 641 authors who self-cited their questionable papers.

Discussion and conclusions

The main aim of our study is to reveal the contributions of citation indexes, which are accepted as the authority in research evaluations, to the visibility of questionable journals whose scientific levels are always considered quite low in academia. According to the results, 13% of the questionable articles were cited by WoS journals and 37% of the citations came from the impact-factor journals. If we accept being cited from authority citation indexes as a tool for visibility, it is obvious that the indexes help the questionable journals to be visible regardless of the name of the index, SSCI, A&HCI or ESCI. The question to be asked at this point is: Do citations to the questionable journals make citation indexes questionable, or do these citations require a closer look at articles published in questionable journals? It is easy to accept all the papers published in questionable or questionable



Country affiliation of the author	Number of self-citations	Number of unique questionable papers	Number of unique WoS papers
Italy	41	12	9
Italy	26	6	5
Croatia	21	17	7
Italy	18	6	3
Israel	17	10	2
Italy	16	12	4
Italy	16	7	4
The Netherlands	8	8	1
Italy	8	5	3
Italy	7	5	5
India	7	7	1

journals as low-quality, but without answering the question, it is difficult to draw a boundary for the definition between high and low quality.

The findings show that neither the impact factor of citing journals nor the size of cited journals is a good predictor of the number of citations to the questionable journals. The high-impact-factor journals cite questionable journals, so it is important to understand the reasons behind this. Our analysis reveals that journals which cited questionable journals more than once have a higher average impact factor. Possible reasons might be disciplinary differences and differences in average number of references. In addition, impact-factor is related to journal size (Antonoyiannakis, 2018), thus the multiple citations journals are larger overall, which in turn leads to higher chances of citing questionable journals.

Why do authors cite others? has been the main question of citation-analysis studies in the literature since the beginning, but our question is *Why do authors cite questionable journals?* after revealing the contributions of citation indexes to the visibilities of questionable journals. Some interesting patterns were found, including author, journal or country self-citations; however, they are not enough to understand the whole picture. It is known that authors and journals use self-citations as a visibility strategy because even one self-citation to the work not cited previously play an important role in the citation of the work by increasing visibility (González-Sala et al., 2019). Coercive citations requested by editors or reviewers, excessive self-citations and citation-stacking issues make all the metrics meaningless because it is easy to manipulate them.

To overcome the limitations of such easy-to-game metrics, a new concept has emerged in the information-science field called *content-based citation analysis*, which focusses on the contents of the citations, not the numbers. It is possible to understand and classify the citations in terms of meanings, purposes, arrays and shapes with the help of these analyses (Taşkın & Al, 2018). This also provides the opportunity to understand the citation motivations of authors and the effects of the publish-or-perish culture on their motivations. We will use *content-based citation analyses*, in the separate study which will be the second part of our project. This will enable us to answer why questionable journals are cited.

This is the first study in which a large-scale analysis of citations to predatory journals is conducted using WoS. When compared to the different results present in the citation studies based on Scopus, it is difficult to assess differences in terms of citing predatory journals in these two databases. Taking into account that only 13% of articles in our study are cited, we can be sure that citations to predatory journals are much more frequent in Google Scholar because it was reported that 43% of articles analyzed in were cited (Björk et al., 2020). However, as an academic search engine, Google Scholar indexes all types of academic materials without any selection criteria as distinct from WoS or Scopus. Therefore, the difference in the number of citations is expected.

Since we did not assess the quality of cited papers published in questionable journals, there are two possible interpretations of the main result of our study: (1) up to 13% of worthless articles in predatory journals can still leak to the mainstream literature legitimised by WoS, and (2) up to 13% of papers published in questionable journals are somehow important for developing a scholarly legitimate discussion in social science. Unlike Oermann et al. (2020), we are not so sure that the important conclusion of the studies on predatory journals is to completely stop citing such journals. Further, we prefer to leave the question raised by the result of our study open.

By pointing out that most of the citers are from the USA, our study supports results by Oermann et al. (2019) and differs from the conclusion presented by Frandsen (2017). Both Oermann et al. and Frandsen used Scopus for their analysis, so the difference is most likely caused by the fact that Frandsen was using a version of Beall's list from 2014 and

Oermann et al. (just like us) used the last available version of the list from 2017. However, it is also important to note that the study by Frandsen did not filter journals according to their discipline; the study by Oermann et al. was focussed on nursing journals, and our study focussed on journals from the social sciences.

One could assume that papers in predatory journals by authors from some countries will more likely be cited in WoS than others. In general, our results seem to indicate that such an assumption is false. When it comes to being cited, the affiliation of the author of the article from the predatory journal does not seem to play such a significant role. The countries from which the authors produced most of the cited papers are also the countries that produce most of the predatory journals (Demir, 2018a). It is not so surprising that US authors cited a relatively high share of papers published in predatory journals because the USA has much more publications in WoS than Turkey or India (Schlegel, 2015).

However, it is quite interesting that scholars from the USA cited predatory journals 555 times and scholars from China did so 291 times, even though their yearly output of articles in WoS is similar. This could indicate that when publishing their papers, it is relatively easy for US scholars to distinguish journals deemed as predatory; they are less aware of the predatory nature of journals when it comes to citing. It could also be interpreted that for some US scholars, many predatory journals are less prestigious places to publish, but they can still be sources of legitimate knowledge. However, to better understand this phenomenon, content-based analysis of the citations to predatory journals would be required.

Our results also support a conclusion made by Oermann et al. (2019), who pointed out that there is no significant difference between journals with and without JIF in terms of citing predatory journals. Our results indicate that there is no connection between the value of the JIF of a given journal and this journal's citations to predatory journals. Although the number of such citations is relatively small, this could be another argument against treating JIF as a measure of journals' quality.

Appendix

The list of 65 analysed questionable journals is available here: https://figshare.com/articles/dataset/Appendix_-_List_of_65_blacklisted_journals/13560326.

Acknowledgements The authors would like to thank Ewa A. Rozkosz for her support. Part of the initial result of this study has been submitted to ISSI 2021 conference as a 'Research in progress' paper.

Funding The work was financially supported by the National Science Centre in Poland (Grant Number UMO-2017/26/E/HS2/00019).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anderson, R. (2017). Cabell's new predatory journal blacklist: A review. *Scholarly Kitchen*. https://scholarlykitchen.sspnet.org/2017/07/25/cabells-new-predatory-journal-blacklist-review/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+ScholarlyKitchen+%28The+Scholarly+Kitchen%29.
- Anderson, R. (2019). Citation Contamination: References to Predatory journals in the legitimate scientific literature—The Scholarly Kitchen. *Scholarly Kitchen*. <https://scholarlykitchen.sspnet.org/2019/10/28/citation-contamination-references-to-predatory-journals-in-the-legitimate-scientific-literature/>.
- Antonoyiannakis, M. (2018). Impact factors and the central limit theorem: Why citation averages are scale dependent. *Journal of Informetrics*, 12(4), 1072–1088. <https://doi.org/10.1016/j.joi.2018.08.011>
- 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. (2012). Predatory publishers are corrupting open access. *Nature*, 489(7415), 179–179. <https://doi.org/10.1038/489179a>
- Beall, J. (2013). The open-access movement is not really about open access. *TripleC*, 11(2), 589–597.
- Björk, B.-C., Kanto-Karvonen, S., & Harviainen, J. T. (2020). How frequently are articles in predatory open access journals cited. *Publications*, 8(2), 17. <https://doi.org/10.3390/publications8020017>
- Bohannon, J. (2013). Who's afraid of peer review? *Science*, 342(6154), 60–65. <https://doi.org/10.1126/science.342.6154.60>
- Cabells. (2021). *Get a quote*. <https://www2.cabells.com/get-quote>.
- Cano, V. (1989). Citation behavior: Classification, utility, and location. *Journal of the American Society for Information Science*, 40(4), 284–290.
- Clarivate. (2018). *Web of Science: Editorial statement about dropped journals*. https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-Editorial-statement-about-dropped-journals?language=en_US.
- Clarivate Analytics. (2012). *OECD Category Scheme*. <http://help.prod-incites.com/inCites2Live/filterValuesGroup/researchAreaSchema/oecdCategoryScheme.html>.
- Clarivate Analytics. (2021). Journal Citation Reports editorial expression of concern. *Web of Science Group*. <https://clarivate.com/webofsciencegroup/essays/jcr-editorial-expression-of-concern/>.
- 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.15256.1>
- Demir, S. B. (2018a). 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. (2018b). Scholarly databases under scrutiny. *Journal of Librarianship and Information Science*. <https://doi.org/10.1177/0961000618784159>
- Else, H. (2019). Impact factors are still widely used in academic evaluations. *Nature*. <https://doi.org/10.1038/d41586-019-01151-4>
- Eykens, J., Guns, R., Rahman, A. I. M. J., & Engels, T. C. E. (2019). Identifying publications in questionable journals in the context of performance-based research funding. *PLoS ONE*, 14(11), e0224541. <https://doi.org/10.1371/journal.pone.0224541>
- Ezinwa Nwagwu, W., & Ojemeni, O. (2015). Penetration of Nigerian predatory biomedical open access journals 2007–2012: A bibliometric study. *Learned Publishing*, 28(1), 23–34. <https://doi.org/10.1087/20150105>
- Frandsen, T. F. (2017). Are predatory journals undermining the credibility of science? A bibliometric analysis of citers. *Scientometrics*, 113(3), 1513–1528. <https://doi.org/10.1007/s11192-017-2520-x>
- Frandsen, T. F. (2019). Why do researchers decide to publish in questionable journals? A review of the literature: Why authors publish in questionable journals. *Learned Publishing*, 32(1), 57–62. <https://doi.org/10.1002/leap.1214>
- González-Sala, F., Osca-Lluch, J., & Haba-Osca, J. (2019). Are journal and author self-citations a visibility strategy? *Scientometrics*, 119(3), 1345–1364. <https://doi.org/10.1007/s11192-019-03101-3>
- Grudniewicz, A., Moher, D., Cobey, K. D., Bryson, G. L., Cukier, S., Allen, K., Arden, C., Balcom, L., Barros, T., Berger, M., Ciro, J. B., Cugusi, L., Donaldson, M. R., Egger, M., Graham, I. D., Hodgkinson, M., Khan, K. M., Mabizela, M., Manca, A., ... Lalu, M. M. (2019). Predatory journals: No definition, no defence. *Nature*, 576(7786), 210–212. <https://doi.org/10.1038/d41586-019-03759-y>

- Krawczyk, F., & Kulczycki, E. (2020). How is open access accused of being predatory? The impact of Beall's lists of predatory journals on academic publishing. *The Journal of Academic Librarianship*. <https://doi.org/10.1016/j.acalib.2020.102271>
- Leydesdorff, L., Bornmann, L., Comins, J. A., & Milojević, S. (2016). Citations: Indicators of quality? The impact fallacy. *Frontiers in Research Metrics and Analytics*. <https://doi.org/10.3389/frma.2016.00001>
- Moussa, S. (2021). Citation contagion: A citation analysis of selected predatory marketing journals. *Scientometrics*. <https://doi.org/10.1007/s11192-020-03729-6>
- Nelhans, G., & Bodin, T. (2020). Methodological considerations for identifying questionable publishing in a national context: The case of Swedish Higher Education Institutions. *Quantitative Science Studies*. https://doi.org/10.1162/qss_a_00033
- Oermann, M. H., Nicoll, L. H., Ashton, K. S., Edie, A. H., Amarasekara, S., Chinn, P. L., Carter-Templeton, H., & Ledbetter, L. S. (2020). Analysis of citation patterns and impact of predatory sources in the nursing literature. *Journal of Nursing Scholarship*, 52(3), 311–319. <https://doi.org/10.1111/jnu.12557>
- Oermann, M. H., Nicoll, L. H., Carter-Templeton, H., Woodward, A., Kidayi, P. L., Neal, L. B., Edie, A. H., Ashton, K. S., Chinn, P. L., & Amarasekara, S. (2019). Citations of articles in predatory nursing journals. *Nursing Outlook*, 67(6), 664–670. <https://doi.org/10.1016/j.outlook.2019.05.001>
- Olivarez, J., Bales, S., Sare, L., & vanDuinkerken, W. (2018). Format aside: Applying Beall's criteria to assess the predatory nature of both OA and Non-OA library and information science journals. *College & Research Libraries*, 79(1), 52–67. <https://doi.org/10.5860/crl.79.1.52>
- Organisation for Economic Co-Operation and Development. (2007). *Revised field of science and technology (FOS) classification in the Frascati manual DSTI/EAS/STP/NESTI(2006)19/FINAL*. <http://www.oecd.org/science/inno/38235147.pdf>
- Ormann, L. (2016). 50 Journals used in FT Research Rank. *Financial Times*. <https://www.ft.com/content/3405a512-5cbb-11e1-8f1f-00144feabdc0>
- Schlegel, F. (Ed.). (2015). *UNESCO science report: Towards 2030*. UNESCO Publ.
- Shen, C., & Björk, B.-C. (2015). 'Predatory' open access: A longitudinal study of article volumes and market characteristics. *BMC Medicine*, 13(1), 230–230. <https://doi.org/10.1186/s12916-015-0469-2>
- Sile, L., Pölonen, J., Sivertsen, G., Guns, R., Engels, T. C. E. E., Arefiev, P., Dušková, M., Faurbæk, L., Holl, A., Kulczycki, E., Macan, B., Nelhans, G., Petr, M., Pisk, M., Soós, S., Stojanovski, J., Stone, A., Šušol, J., & Teitelbaum, R. (2018). Comprehensiveness of national bibliographic databases for social sciences and humanities: Findings from a European survey. *Research Evaluation*, 27(4), 310–322. <https://doi.org/10.1093/reseval/rvy016>
- Siler, K. (2020). Demarcating spectrums of predatory publishing: Economic and institutional sources of academic legitimacy. *Journal of the Association for Information Science and Technology*, 71(11), 1386–1401. <https://doi.org/10.1002/asi.24339>
- 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. <https://doi.org/10.3145/epi.2016.sep.03>
- Sorokowski, P., Kulczycki, E., Sorokowska, A., & Pisanski, K. (2017). Predatory journals recruit fake editor. *Nature*, 543(7646), 481–483. <https://doi.org/10.1038/543481a>
- Stöckelová, T., & Vostal, F. (2017). Academic stratospheres-cum-underworlds: When highs and lows of publication cultures meet. *Aslib Journal of Information Management*, 69(5), 516–528. <https://doi.org/10.1108/AJIM-01-2017-0013>
- Taşkın, Z., & Al, U. (2018). A content-based citation analysis study based on text categorization. *Scientometrics*, 114(1), 335–357. <https://doi.org/10.1007/s11192-017-2560-2>
- Zhu, J., & Liu, W. (2020). A tale of two databases: The use of web of science and scopus in academic papers. *Scientometrics*, 123(1), 321–335. <https://doi.org/10.1007/s11192-020-03387-8>