RESEARCH ARTICLE

REVISED Citations and metrics of journals discontinued from Scopus for publication concerns: the GhoS(t)copus Project [version 2; peer review: 2 approved, 1 approved with reservations]

Previously titled: Inflated citations and metrics of journals discontinued from Scopus for publication concerns: the GhoS(t)copus Project

Andrea Cortegiani[1], Mariachiara Ippolito[1], Giulia Ingoglia[1], Andrea Manca[2], Lucia Cugusi[2], Anna Severin[3,4], Michaela Strinzel[3], Vera Panzarella[5], Giuseppina Campisi[5], Lalu Manoj[6], Cesare Gregoretti[1], Sharon Einav[7], David Moher[6], Antonino Giarratano[1]

[1]Department of Surgical, Oncological and Oral Science (Di.Chir.On.S.), University of Palermo, Department of Anesthesia Intensive Care and Emergency, Policlinico Paolo Giaccone, Palermo, Italy
[2]Department of Biomedical Sciences, University of Sassari, Sassari, 07100, Italy
[3]Swiss National Science Foundation, Swiss National Science Foundation, Bern, CH-3001, Switzerland
[4]Institute for Social and Preventive Medicine, Institute for Social and Preventive Medicine, Bern, 3012, Switzerland
[5]Department of Surgical, Oncological and Oral Science (Di.Chir.On.S), Section of Oral Medicine, University of Palermo, Palermo, 90127, Italy
[6]Centre for Journalology, Clinical Epidemiology Program, University of Ottawa, Ottawa, 201B, Canada
[7]Intensive Care Unit of the Shaare Zedek Medical Medical Centre, Hebrew University Faculty of Medicine, Jerusalem, Israel

v2 **First published:** 21 May 2020, **9**:415
https://doi.org/10.12688/f1000research.23847.1
Latest published: 26 Aug 2020, **9**:415
https://doi.org/10.12688/f1000research.23847.2

Abstract

Background: Scopus is a leading bibliometric database. It contains a large part of the articles cited in peer-reviewed publications**.** The journals included in Scopus are periodically re-evaluated to ensure they meet indexing criteria and some journals might be discontinued for 'publication concerns'. Previously published articles may remain indexed and can be cited. Their metrics have yet to be studied. This study aimed to evaluate the main features and metrics of journals discontinued from Scopus for publication concerns, before and after their discontinuation, and to determine the extent of predatory journals among the discontinued journals.
Methods: We surveyed the list of discontinued journals from Scopus (July 2019). Data regarding metrics, citations and indexing were extracted from Scopus or other scientific databases, for the journals discontinued for publication concerns.
Results: A total of 317 journals were evaluated. Ninety-three percent

Open Peer Review

Reviewer Status ✓ ? ✓

	Invited Reviewers		
	1	**2**	**3**
version 2 (revision) 26 Aug 2020			✓ report
version 1 21 May 2020	✓ report	? report	

1. **Johann Mouton**, Stellenbosch University, Stellenbosch, South Africa

of the journals (294/317) declared they published using an Open Access model. The subject areas with the greatest number of discontinued journals were *Medicine* (52/317; 16%), *Agriculture and Biological Science* (34/317; 11%), and *Pharmacology, Toxicology and Pharmaceutics* (31/317; 10%). The mean number of citations per year after discontinuation was significantly higher than before (median of difference 16.89 citations, p<0.0001), and so was the number of citations per document (median of difference 0.42 citations, p<0.0001). Twenty-two percent (72/317) were included in the Cabell's blacklist. The DOAJ currently included only 9 journals while 61 were previously included and discontinued, most for 'suspected editorial misconduct by the publisher'.
Conclusions: Journals discontinued for 'publication concerns' continue to be cited despite discontinuation and predatory behaviour seemed common. These citations may influence scholars' metrics prompting artificial career advancements, bonus systems and promotion. Countermeasures should be taken urgently to ensure the reliability of Scopus metrics for the purpose of scientific assessment of scholarly publishing at both journal- and author-level.

Keywords
predatory, journal, Scopus, metrics, indexing, citation count

This article is included in the Science Policy Research gateway.

2. **Pablo Iriarte**, Library of the University of Geneva, Geneva, Switzerland
Floriane Muller, Library of the University of Geneva, Geneva, Switzerland
Nadia Elia, University of Geneva, Geneva, Switzerland

3. **Arash Afshari**, University of Copenhagen, Copenhagen, Denmark

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Andrea Cortegiani (andrea.cortegiani@unipa.it)

Author roles: Cortegiani A: Conceptualization, Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Ippolito M**: Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Ingoglia G**: Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Manca A**: Data Curation, Writing – Original Draft Preparation, Writing – Review & Editing; **Cugusi L**: Data Curation, Writing – Original Draft Preparation, Writing – Review & Editing; **Severin A**: Data Curation, Writing – Original Draft Preparation, Writing – Review & Editing; **Strinzel M**: Data Curation, Writing – Original Draft Preparation, Writing – Review & Editing; **Panzarella V**: Data Curation, Writing – Review & Editing; **Campisi G**: Supervision, Writing – Review & Editing; **Manoj L**: Supervision, Writing – Review & Editing; **Gregoretti C**: Supervision, Writing – Review & Editing; **Einav S**: Supervision, Writing – Review & Editing; **Moher D**: Supervision, Writing – Review & Editing; **Giarratano A**: Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

How to cite this article: Cortegiani A, Ippolito M, Ingoglia G *et al.* **Citations and metrics of journals discontinued from Scopus for publication concerns: the GhoS(t)copus Project [version 2; peer review: 2 approved, 1 approved with reservations]** F1000Research 2020, **9**:415 https://doi.org/10.12688/f1000research.23847.2

First published: 21 May 2020, **9**:415 https://doi.org/10.12688/f1000research.23847.1

REVISED **Amendments from Version 1**

We are glad to submit a new version of our manuscript, now entitled "Citations and metrics of journals discontinued from Scopus for publication concerns: the GhoS(t)copus Project", and incorporating the insights provided by the reviewers.

In this revised version, more words of caution have been inserted regarding the interpretation of our findings, and the lack of a control group has been listed among the limitations of the study.

Typos were amended and few data were corrected. The findings and conclusions remain consistent with the previous version since no substantial changes were made. English form was revised.

More details have been added in the methods section, in order to improve the reproducibility of the research.

Furthermore, some minor changes were made in the tables, the figure caption and the reference list, following reviewers' suggestions.

Peer review had an important role in improving this manuscript, that now results more balanced. Also, new insights for further research questions have been included.

Any further responses from the reviewers can be found at the end of the article

Introduction

Scopus is a leading bibliometric database launched in 2004 by the publishing and analytics company Elsevier. It was developed by research institutions, researchers and librarians, and contains the largest number of abstracts and articles cited in peer reviewed academic journal articles that cover scientific, technical, medical, and social science fields[1].

Scopus provides bibliometric indicators that many institutions use to rank journals to evaluate the track record of scholars who seek hiring or promotion. These metrics are also used to allocate financial bonuses or to evaluate funding applications[2–4]. Ensuring the quality of the content of the Scopus database is therefore of great importance.

Scopus indexed journals undergo evaluation and periodic review by an independent and international Content Selection and Advisory Board (CSAB), a group of scientists, researchers and librarians, comprised of 17 Subject Chairs, each representing a specific subject field- and by a computerized algorithm[1]. At any time after journal inclusion, concerns regarding its quality may be raised by a formal complaint, thereby flagging the journal for re-evaluation by the CSAB. Should the CSAB panel determine that the journal no longer meets Scopus standards, new articles from that journal are no longer indexed[1]. One of the most common reasons for discontinuation is 'publication concerns', which refers to the quality of editorial practices or other issues that have an impact on its suitability for continued coverage[5]. The list of the discontinued sources is publicly available and is updated approximately every six months[6]. However, articles published in journals that were discontinued and are no longer indexed, are probably not removed from the Scopus database.

It has been claimed that a number of journals discontinued from Scopus for publication concerns might be so-called 'predatory' journals[5]. Predatory journals "prioritize self-interest at the expense of scholarship and are characterized by false or misleading information, deviation from best editorial and publication practices, a lack of transparency, and/or the use of aggressive and indiscriminate solicitation practices"[7]. Since researchers are pressured to publish in indexed journals, predatory journals are constantly trying to be indexed in the Scopus database, thereby boosting their attractiveness to researchers[2,8]. Having articles from predatory journals indexed in Scopus poses a threat to the credibility of science and might cause harm particularly in fields where practitioners rely on empirical evidence in the form of indexed journal articles[8,9].

We hypothesize that, even though Scopus coverage is halted for discontinued journals, they are still cited, as all their documents, that are already indexed, remain available to users. To date, the metrics of those journals discontinued for publication concerns have not been studied yet. Therefore, in the present analysis we set out to (1) evaluate the main scientific features and citation metrics of journals discontinued from Scopus for publication concerns, before and after discontinuation, and (2) determine the extent of predatory journals included in the discontinued journals.

Methods

Search strategy

The freely accessible and regularly updated Elsevier list (see Source data) of journals discontinued from the Scopus database (version July 2019)[10] was accessed on 24th January 2020 (See Underlying data[11]). We restricted our analysis to journals discontinued for "publication concerns". Journals were checked for relevant data (described below), which were then independently collected by four pairs of authors (MI and GI, AM and LC, AS and MS, VP and AC), each pair being assigned one quarter of the data to be collected in duplicate. The data were collected using a standardized data extraction form (Underlying data Table 1). A second check to confirm the data and resolve discrepancies was performed by four additional authors that had not been involved in data collection (LM, CG, SE, AG). Data collection was initiated on 24th January and completed by the end of February 2020. Confirmed data were registered on an Excel datasheet (*Underlying data*, Table 1[11]).

Retrieved data and sources

Data were extracted either from the Scopus database[10] or by searching other sources, such as SCImago Journal & Country Rank (SJCR)[12], Journal Citation Reports[13], Centre for Science and Technology Studies (CWTS) Journal Indicators[14], Beall's updated List[15], Directory of Open Access Journals (DOAJ)[16], PubMed[17] and Web of Science[18]. Open Access policy was checked on journals websites. The standardized data extraction form, independently applied by eight authors (MI, GI, AM, LC, AS, MS, VP, AC), was used to collect the following data: journal title, name and country of the publisher, the number of years of Scopus coverage, year of Scopus discontinuation, subject areas and sub-subject areas, Impact Factor (IF), CiteScore,

SCImago Journal Rank (SJR), Source Normalized Impact *per* Paper (SNIP), best SCImago quartile, the indexing of at least one article in PubMed, Web Of Science (WOS) and DOAJ (for open access journals) indexing, presence in the updated Beall's List, total number of published documents and total number of citations. All the metrics were checked on the year of Scopus discontinuation. In cases of discrepancies between Scopus data and other sources, Scopus data was preferred.

We defined the 'before discontinuation' time frame as the period included within the first year of journal coverage by Scopus and the year of discontinuation, which was not included in our calculations. The 'after discontinuation' time frame, was defined as the period included within the year of Scopus discontinuation and 2020. If the journal had been discontinued more than once, the time frame was based on the last one, according to the date of the last document displayed in the Scopus database. Citations 'before' and 'after' the date of discontinuation were manually counted based on either the Scopus journal overview or the downloadable tables made available by Scopus upon request (see Source data). When evaluating the presence of articles in PubMed (e.g. PubMed Central) and WOS and DOAJ indexing, 2019 was considered the reference year, preventing disadvantages for journals with time gaps for publication.

We calculated the median number of cumulative citations across all discontinued journals per year of coverage and defined it as 'Citations per year'. We also calculated the median number of cumulative citations across all discontinued journals per document ('Citations per document'). We included all documents indexed in Scopus, regardless of type. Finally, one author (AS) checked whether discontinued journals were present in Cabell's whitelist or blacklist[19] or the DOAJ's list of discontinued journals[20]. As some of the journals included in the blacklist lack ISSNs or other unique identifiers, the comparison of the three lists with Scopus's discontinued journals was based on matching the journals' names by similarity using the Jaro-Winkler algorithm in RStudio Desktop 1.2.5033 and RecordLinkage 0.4–11.2 following the approach developed by Strinzel *et al.* (2019)[21,22]. The Jaro-Winkler metric, scaled between 0 (no similarity) and 1 (exact match), was calculated for all possible journals' pairings[23]. We manually inspected all pairs with a Jaro-Winkler metric smaller than one in order to include cases where, due to the orthographical differences between the lists, no exact match was found. For each matched pair, we compared journal publishers and, where possible, ISSNs in order to exclude cases where two journals had the same or a similar name but were edited by different publishers.

Full definitions and descriptions of the sources and metrics are reported in the Extended Data Appendix 1[24].

Statistical analysis

All data management and calculations were performed using Microsoft Excel (version 2013, Microsoft Corporation®, USA) and GraphPad Prism (version 8.3.1, 322, GraphPad software®, San Diego California). Variable distribution was assessed for normality using the D'Agostino-Pearson test. For variables with normal distribution means and standard deviations (SDs) were reported. For non-normally distributed data medians, interquartile ranges (IQRs, 25th–75th) and ranges (minimum value - maximum value) were reported. Categorical data were expressed as proportions and percentages.

The paired sample *t* test or the Wilcoxon matched-pairs signed ranked test were used to compare journal data before and after Scopus discontinuation, as appropriate.

Results

Data could be retrieved regarding 317 of the 348 journals listed as discontinued (91.1%). The remaining journals were not found on the Scopus database using the search tool.

Journals' and publishers' characteristics

Among the 135 publishers identified, those with the largest number of discontinued journals were: *Academic Journals Inc.* (39/317; 12.3%), *Asian Network for Scientific Information* (19/317; 6.0%), and *OMICS Publishing Group* (18/317; 5.7%). Table 1 reports the distribution of journals discontinued from Scopus by publisher. United States (76/317, 24%), India (63/317, 20%) and Pakistan (49/317, 15%) were countries most commonly declared as publisher headquarters (Figure 1 and Table 2).

The subject areas with the greatest number of discontinued journals were *Medicine* (52/317; 16%), *Agriculture and Biological Science* (34/317; 11%), and *Pharmacology, Toxicology and Pharmaceutics* (31/317; 10%) Table 3 and Extended data Table 1[25] report the distribution of discontinued journals by subject area and sub-area in full. Of these journals, 93% (294/317) declared they published using an Open Access model.

First subject area as displayed in Scopus. Note: a journal may have more than one subject area. Table 4 shows the characteristics and metrics of the journals at the time of their discontinuation.

The median time of Scopus coverage prior to discontinuation of the journals was 8 years (IQR 6–10, range 1–54). In total, 299 journals had been assigned to a SCImago quartile (Q); 39 of them (13%) listed in Q1 or Q2, and 260 in Q3 or Q4 (87%). Only ten of the discontinued journals had an Impact Factor at the year of discontinuation, with a median value of 0.84 (IQR 0.37–2.29, range 0.28–4).

Citation metrics

Table 5 shows the total number of documents and citations, the total number of documents per journal and the citations count before and after Scopus discontinuation. The total number of citations received after discontinuation was 607,261, with a median of 713 citations (IQR 254–2,056, range 0–19,468) per journal.

Paired *t*-tests (Wilcoxon matched-pairs signed rank test) revealed that the number of citations per year after discontinuation was significantly higher than before (median of difference 16.89 citations [-13.68-117.5] (-1427-3491), p<0.0001). Likewise, the number of citations per document proved significantly

Table 1. Distribution of journals discontinued from Scopus by publisher.

Publishers (n=135)	% (n)
Academic Journals Inc.	12.3 (39/317)
Asian Network for Scientific Information	6 (19/317)
OMICS Publishing Group	5.7 (18/317)
Medwell Journals	4.1 (13/317)
iMedPub	3.5 (11/317
World Scientific and Engineering Academy and Society	2.8 (9/317)
Science Publications	2.5 (8/317)
Academy Publisher	2.2 (7/317)
Allied Academies	1.9 (6/317)
Canadian Center of Science and Education	1.9 (6/317)
International Digital Organization for Scientific Information (IDOSI)	1.9 (6/317)
Science and Engineering Research Support Society	1.6 (5/317)
Serials Publications (International Science Press)	1.6 (5/317)
AMSE Press	1.3 (4/317)
Eurojournals Inc	1.3 (4/317)
Hikari Ltd	1.3 (4/317)
Research India Publications	1.3 (4/317)
Others	47 (149/317)

Data are reported as percentages and fractions. Publishers with less than four journals discontinued from Scopus were grouped as 'Others'.

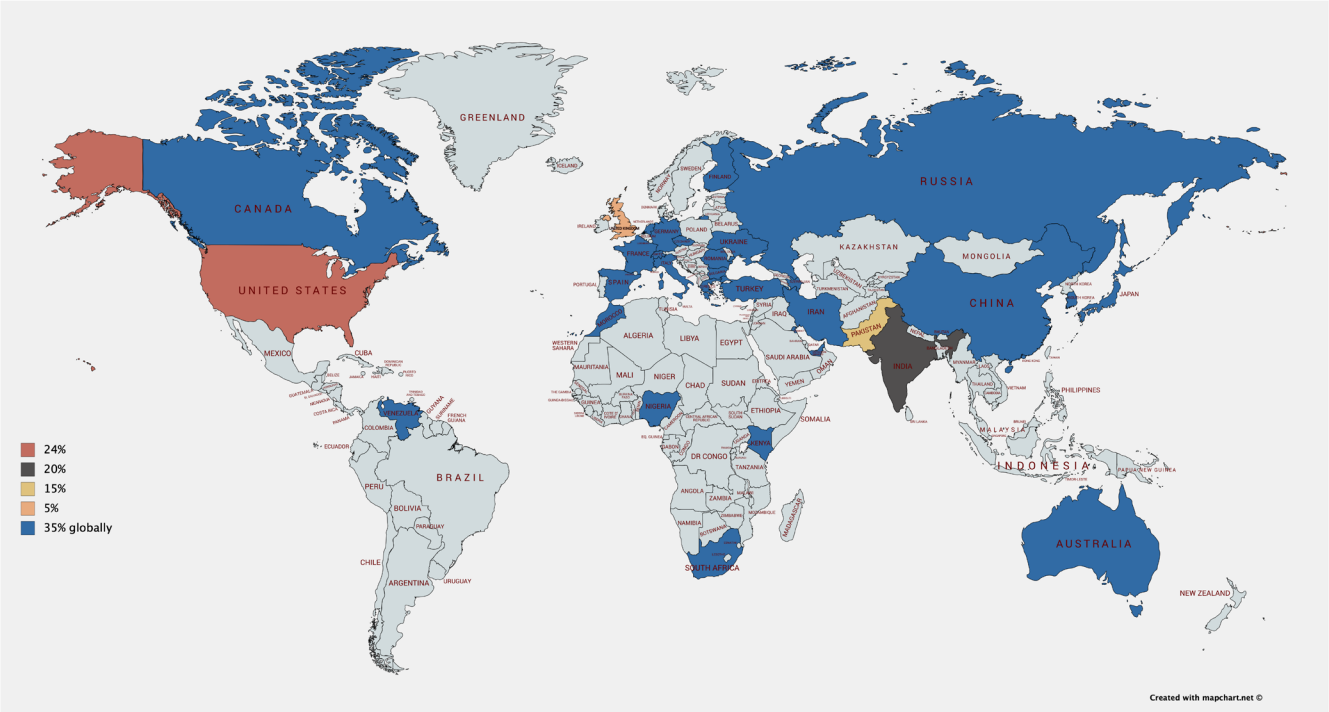

Figure 1. Distribution of journals discontinued from Scopus by country. The map chart shows the different frequencies of distribution by country with different colors.

Table 2. Distribution of journals discontinued from Scopus by country.

Country (n=33)	% (n)
United States of America	24 (76/316)
India	19.9 (63/316)
Pakistan	15.5 (49/316)
United Kingdom	5.4 (17/316)
Turkey	4.1 (13/316)
Greece	3.5 (11/316)
Canada	3.2 (10/316)
Finland	2.5 (8/316)
France	2.2 (7/316)
United Arab Emirates	1.9 (6/316)
Italy	1.6 (5/316)
Romania	1.6 (5/316)
South Korea	1.6 (5/316)
Ukraine	1.6 (5/316)
Australia	1.3 (4/316)
Bulgaria	1.3 (4/316)
Others	8.8 (28/316)

Data were retrieved from Scimago Journal & Country Rank and are reported as percentages and fractions. Countries with less than four Scopus discontinued journals were grouped as 'Others'

higher after discontinuation (median of difference 0.42 citations [-0.32-1.31] (-10.35-79.49), p<0.0001).

Indexing in Cabell's lists, updated Beall's list, DOAJ and scientific databases

Among the discontinued journal, 22% (72/317) were included in the Cabell's blacklist, while 29 (9%) were currently under review for inclusion. Only five journals (2%) were included in Cabell's whitelist. In 243 cases (76.6%), either the journal publisher was included in the updated Beall's list of predatory publishers or the journal was included in the corresponding list of standalone journals (76.6%). The DOAJ currently includes only 9 journals. In total, 61 journals were previously included and discontinued by DOAJ; in 36 cases the reason was 'suspected editorial misconduct by the publisher' in 23 instances it was 'journal not adhering to best practice' and in one case 'no open access or license info'.

Table 6 shows the indexing in Web of Science, updated Beall's list, Cabell's white- and blacklist, and DOAJ (both included and discontinued) and the presence of articles in PubMed.

Table 3. Distribution of journals discontinued from Scopus by subject areas.

Medicine	16.4%	(52/317)
Agricultural and Biological Sciences	10.7%	(34/317)
Pharmacology, Toxicology and Pharmaceutics	9.8%	(31/317)
Engineering	7.9%	(25/317)
Computer Science	7.9%	(25/317)
Biochemistry, Genetics and Molecular Biology	5.4%	(18/317)
Business, Management and Accounting	5.4%	(17/317)
Mathematics	5.4%	(17/317)
Social Sciences	4.7%	(15/317)
Arts and Humanities	3.8%	(12/317)
Multidisciplinary	3.5%	(11/317)
Economics, Econometrics and Finance	2.5%	(8/317)
Environmental Science	2.2%	(7/317)
Immunology and Microbiology	2.2%	(7/317)
Materials Science	2.2%	(7/317)
Veterinary	2.2%	(7/317)
Earth and Planetary Sciences	1.6%	(5/317)
Chemistry	1.3%	(4/317)
Energy	1.3%	(4/317)
Chemical Engineering	0.9%	(3/317)
Physics and Astronomy	0.9%	(3/317)
Nursing	0.6%	(2/317)
Dentistry	0.3%	(1/317)
Health Professions	0.3%	(1/317)
Neuroscience	0.3%	(1/317)

Data were retrieved from Scopus and are reported as percentages and fraction

Table 4. Journal characteristics at the year of Scopus discontinuation.

Scopus coverage (yrs.) *	8 [6-10] (1-54)
Time from Scopus discontinuation (yrs.) *	5 [4-6] (2-12)
Impact Factor†	0.84 [0.37-2.29] (0.28-4)
SjR‡	0.17 [0.13-0.23] (0.1-1.41)
SNIP§	0.4 [0.23-0.65] (0-4.56)
CiteScore °	0.32 [0.17-0.46] (0-10.33)
SCImago Quartile	
Q1 (%, n) **Q2 (%, n)** **Q3 (%, n)** **Q4 (%, n)**	3.3 (10/299) 9.7 (29/299) 40.8 (122/299) 46.1 (138/299)

Data are reported as medians, interquartile ranges [IQRs] and ranges (minimum value – maximum value) or as percentages and fractions.

* No missing data. The analyses were conducted on all the 317 journals discontinued from Scopus.

† Data were available and calculated for 10 journals.

‡ Data were available and calculated for 304 journals.

§ Data were available and calculated for 299 journals.

° Data were available and calculated for 82 journals.

SjR: SCImago Journal & Country Rank; SNIP: Source Normalized Impact *per* Paper; IF: Impact Factor

Table 5. Citations and documents before and after Scopus discontinuation.

Total number of documents	591968	
Total number of citations	1152779	
Documents *per* journal*	429 [159.5–1244] (2–132482)	
	Before Scopus discontintinuation	**After Scopus discontintinuation**
Citations (n)	545518	607261
Citations *per* journal*	415 [120-1580] (0-67529)	713 [254-2056] (0-19468)
Citations *per* year*	51.75 [15.17- 144.3] (0- 2028)	152.9 [49.43-408] (0-4571)
Citations *per* document*	1 [0.39-2.15] (0-17.12)	1.66 [0.93-2.66] (0-80.70)

Data are reported as medians, interquartile ranges [IQRs] and ranges (minimum value – maximum value) unless otherwise specified. * No missing data. Analyses were conducted on all the 317 journals discontinued from Scopus.

Discussion

The present study aimed to scrutinize the main features of journals whose coverage was discontinued by Scopus due to publication concerns. To do so, (a) we counted and compared citation metrics per journal and per document obtained *before* and *after* discontinuation, and (b) we accessed established blacklists and whitelists dealing with the issue of predatory publishing, i.e. Cabell's and updated Beall's list, as well as the DOAJ.

Our main finding was that articles published in these journals before discontinuation remain available to users and continue

Table 6. Discontinued journals' current Open Access policy and the indexing of their articles in major databases.

Open Access journals (%, n)	92.7 (294/317)
PubMed (%, n)*	6.3 (20/317)
Web Of Science (%, n)	9.1 (29/317)
Beall's List (%, n)	76.6 (243/317)
Cabell's Whitelist (%, n)	1.6 (5/317)
DOAJ included (%, n)	2.8 (9/317)
Cabell's Blacklist (%, n)	22.7 (72/317)
DOAJ discontinued (%, n)	19.2 (61/317)

Data are reported as percentages and fractions. DOAJ: Directory of Open Access Journals

* Proportion only of journals with at least one article indexed in PubMed (e.g. PubMed Central).

to be cited after discontinuation, and even more so than before. Moreover, a large number of the discontinued journals are likely to be predatory.

A previous analysis conducted to evaluate the scientific impact of predatory publishing has concluded that "articles published in predatory journals have little scientific impact"[26]. The study evaluated Google Scholar and Scopus citation statistics of 250 randomly sampled articles, that have been published in predatory journals in 2014. The citations were then compared to those of a control group of articles, published in journals included in Scopus database. Our study aimed to evaluate and describe the metrics and citations of all the journals discontinued from Scopus for 'publication concerns'. At a secondary stage, the presence of these journals in the Cabells' and Beall's lists was investigated. The different purposes and designs of the two studies may explain the different findings.

Although Scopus rigorously controls content quality and warns users when a journal is discontinued in its source details, the average user rarely accesses journaldetails, usually focusing on article contents alone. As a result the reader remains unaware that the article they have accessed was issued by a journal discontinued for publication concerns. Therefore, articles issued by journals whose scientific reputation is currently deemed questionable continue to be cited as content from legitimate, up-to-standard journals. Quantification of the effect of discontinuation on the likelihood of citation shows that the articles published by these journals received significantly more citations after discontinuation than before.

Apart from dangerous exposure of scholars, clinicians and even patients to potentially dubious or low quality contents, citations from discontinued journals pose a serious threat to assessment of scientific merit and quality by institutions and academia. These citations contribute to the calculation of author metrics by Scopus. Among these metrics is included the Hirsch index (H-index)[27], a lead descriptor of productivity and scientific impact, upon which career advancements are often determined[2–4]. The fact that discontinued journals contribute to academic promotion is a pertinent issue, and has inspired the vignette depicted in Figure 2: discontinued journals may inflate authors' metrics lifting them unnaturally and effortlessly.

Of greatest concern is our finding that many of the discontinued journals display predatory behaviors in claiming to be open access, without actually being indexed in DOAJ. Exploitation of the open-access publishing model has been shown to go hand in hand with deviation from best editorial and publication practices for self-interest[7]. Predatory journals are not only associated with poor editorial quality, but are also deceptive and misleading by nature, i.e. they prioritize self-interest at the expense of scholars, and lack transparent and independent peer review[7,28]. Young researchers from low- and middle-income countries are probably most susceptible to the false promises and detrimental practices of predatory journals. However, "predatory scholars" also seem to exist, possibly sharing a common interest with deceptive journals and publishers, knowingly using them to achieve their own ends[29,30].

The policy underlying the decision to keep publications prior to discontinuation of indexing is clear. Some of these publications may actually fulfill publishing criteria (e.g. International Committee of Medical Journal Editors, Committee on Publication Ethics). It would be unfair to punish researchers for an eventual deterioration in journal performance; changes in the standards employed by the journal may change over time and the researchers may be unaware of quality issues. On the other hand, as the integrity of the editorial process cannot be vouched for, it is ethically untenable to keep such data available without clearer warnings.

One measure that could be undertaken immediately is, for example, flagging of articles that have been published in discontinued journals with clearly visible information regarding journal discontinuation, its date and its cause. Submitting articles published a certain amount of time before journal discontinuation to post-publication open peer-review is also a possibility. However, as solutions to this problem must balance fairness towards publishing researchers with ensuring the correctness of the metrics and citations deriving from these journals, Scopus may need to to set criteria for deleting discontinued journals from the publicly available database or, in the least, stop tracking their citations. Such measures must only be applied by the CSAB case-by-case, after evaluating the full impact of such action and the severity of the potential misconducts. At the author-level, an alternative may be the provision of two metrics: one with and one without citations from publications in discontinued journals.

This analysis is not free of limitations. First, this study lacks a control group of journals whose coverage had not been discontinued in the Scopus database. Therefore the differences

Figure 2. Discontinued journals may inflate authors' metrics lifting them unnaturally and effortlessly.

we identified in the number of citations before and after discontinuation require further validation. Second, we included the year of discontinuation in the "*after discontinuation*" period, starting from January 1st. This decision may have led to some overestimation in the number of citations received after discontinuation. Third, we included only those journals discontinued from Scopus for "publication concerns" but were not able to retrieve details regarding the specific concern raised. Finally, we did not evaluate the impact of the citations received after discontinuation on author-level metrics.

Conclusions

Journals whose coverage in Scopus has been halted for publication concerns continue to be cited. This paradox may influence scholar metrics, potentially prompting career advancements and promotions. Further studies are needed, also investigating the journals discontinued from Scopus using the criteria "outlier performance – radar", particularly effective in flagging potential predatory journals. Countermeasures should be taken to ensure the validity and reliability of Scopus metrics for both journals and authors due to their importance for scientific assessment of scholarly publishing. Creative thinking is required to resolve this issue without punishing authors who have inadvertently published good quality papers in a failing or predatory discontinued journal.

Data availability

Source data

Discontinued sources from Scopus are available from the following link: https://www.elsevier.com/__data/assets/excel_doc/0005/877523/Discontinued-sources-from-Scopus.xlsx

All the relevant data are freely retrievable from Scopus 'journal overview' or can be requested to Scopus through https://www.scopus.com/sources.

Underlying data

Figshare: Underlying data Table 1.xlsx. https://doi.org/10.6084/m9.figshare.12231083.v3[11]

This project contains the following underlying data:

- Underlying data Table 1.xlsx (Standardized data extraction form with data collected)

Extended data

Figshare: Extended data Appendix 1.
https://doi.org/10.6084/m9.figshare.12231110.v2[24]

This project contains the following extended data:

- Extended data Appendix 1.docx (Definitions of sources and metrics used in the manuscript of the GhoS(t)copus Project)

Figshare: Extended data Table 1. https://doi.org/10.6084/m9.figshare.12233171.v2[25]

This project contains the following extended data:

- Extended data table.docx (Distribution of Scopus discontinued journals by subject sub-areas)

Acknowledgments

We would like to thank Dr. Antonio Corrado ("Korrado 20") for creating and providing the Figure 2.

References

1. Elsevier: **How Scopus works**. [accessed 28 February 2020].
Reference Source
2. Cortegiani A, Manca A, Lalu M, *et al.*: **Inclusion of predatory journals in Scopus is inflating scholars' metrics and advancing careers.** *Int J Public Health.* 2020; **65**: 3–4.
Publisher Full Text
3. Bagues M, Sylos-Labini M, Zinovyeva N: **A walk on the wild side: "Predatory" journals and information asymmetries in scientific evaluations.** *Research Policy.* 2019; **48**(2): 462–477.
Publisher Full Text
4. Hedding DW: **Payouts push professors towards predatory journals.** *Nature.* 2019; **565**(7739): 267.
Publisher Full Text | Free Full Text
5. Holland K, Brimblecombe P, Meester W, *et al.*: **The importance of high-quality content: curation and re-evaluation in Scopus**. [accessed 28 February 2020].
Reference Source
6. Elsevier: **Content Policy and Selection**. [accessed 28 February 2020].
Reference Source
7. Grudniewicz A, Moher D, Cobey KD, *et al.*: **Predatory journals: no definition, no defence.** *Nature.* 2019; **576**(7786): 210–212.
PubMed Abstract | Publisher Full Text
8. Cortegiani A, Longhini F, Sanfilippo F, *et al.*: **Predatory Open-Access Publishing in Anesthesiology.** *Anesth Analg.* 2019; **128**(1): 182–187.
PubMed Abstract | Publisher Full Text
9. Severin A, Low N: **Readers beware! Predatory journals are infiltrating citation databases.** *Int J Public Health.* 2019; **64**(8): 1123–1124.
PubMed Abstract | Publisher Full Text
10. Elsevier: **Scopus®, registered trademark of Elsevier B.V**. [accessed 28 February 2020].
Reference Source
11. Cortegiani A, Ippolito M, Ingoglia G, *et al.*: **Underlying data Table 1: Standardized data extraction form with data collected**. 2020.
http://www.doi.org/10.6084/m9.figshare.12231083.v3
12. **Scimago Lab, Copyright 2007-2020**. [accessed 28 February 2020].
Reference Source
13. **Journal Citation Reports**. Copyright 2020. [accessed 28 Februrary 2020].
Reference Source
14. Centre for Science and Technology Studies: **About CWTS.** Leiden University, The Netherlands. [accessed 28 February 2020].
Reference Source
15. **Beall's list of predatory journals and publishers**. [accessed 28 February 2020].
Reference Source
16. **Directory of Open Access Journals**. Licensed under CC BY-SA. [accessed 28 February 2020].
Reference Source
17. **PubMed Help**. Bethesda (MD): National Center for Biotechnology Information (US); 2005; PubMed Help. [Updated 2019 Jul 25]. [accessed 28 February 2020].
Reference Source
18. **Clarivate Analytics Company**. [accessed 28 February 2020].
Reference Source
19. **Cabell's Scholarly Analytics**. [accessed 28 February 2020].
Reference Source
20. Directory of Open Access Journals: **DOAJ publishes lists of journals removed and added**. Directory of Open Access Journals Blog. [accessed 28 February 2020].
Reference Source
21. Strinzel M, Severin A, Milzow K, *et al.*: **Blacklists and Whitelists To Tackle Predatory Publishing: a Cross-Sectional Comparison and Thematic Analysis.** Wolf JM, editor. *mBio.* 2019; **10**(3): pii: e00411-19.
PubMed Abstract | Publisher Full Text | Free Full Text
22. Winkler WE: **String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage.** ERIC. 1990.
Reference Source
23. Porter EH, Winkler WE: **Approximate string comparison and its effect on an advanced record linkage system.** Citeseer. 1997.
Reference Source
24. Cortegiani A, Ippolito M, Ingoglia G, *et al.*: **Extended data Appendix 1.** *figshare.* Online resource. 2020.
http://www.doi.org/10.6084/m9.figshare.12231110.v2
25. Cortegiani A, Cugusi L, Ippolito M, *et al.*: **Extended data Table 1.** *figshare.* Dataset. 2020.
http://www.doi.org/10.6084/m9.figshare.12233171.v2
26. Bjork BC, Kanto-Karvonen S, Harviainen JT: **How Frequently Are Articles in Predatory Open Access Journals Cited.** *Publications.* 2020; **8**(2): 17.
Publisher Full Text
27. Hirsch JE: **An index to quantify an individual's scientific research output.** *Proc Natl Acad Sci U S A.* 2015; **102**(46): 16569–16572.
PubMed Abstract | Publisher Full Text | Free Full Text
28. Butterworth JF 4th, Vetter TR: **Predatory Journals Undermine Peer Review and Cheapen Scholarship.** *Anesth Analg.* 2019; **128**(1): 11–12.
PubMed Abstract | Publisher Full Text
29. Cobey KD, Grudniewicz A, Lalu MM, *et al.*: **Knowledge and motivations of researchers publishing in presumed predatory journals: a survey.** *BMJ Open.* 2019; **9**(3): e026516.
PubMed Abstract | Publisher Full Text | Free Full Text
30. Pond BB, Brown SD, Stewart DW, *et al.*: **Faculty Applicants' Attempt to Inflate CVs Using Predatory Journals.** *Am J Pharm Educ.* 2019; **83**(1): 7210.
PubMed Abstract | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ? ✔

Version 2

Reviewer Report 23 September 2020

https://doi.org/10.5256/f1000research.28667.r71207

✔ **Arash Afshari**
Department of Pediatric and Obstetric Anesthesia, Rigshospitalet, Juliane Marie Center, University of Copenhagen, Copenhagen, Denmark

I like to thank the authors and the journal for considering me as a peer for this review. I very much enjoyed this revised paper and certainly the indebt revision of the comments from the first round. I only have few general comments on the external validity of this paper:

Exclusion of journals only based on publication concerns may not reflect the complexity facing us of information overload.

The proxy control group, which is absent and is mentioned by the authors is in reality a comparison with journals using an editorial system based on peer review as the gold standard that ensures adequate quality. But considering the issues surrounding the famous hydroxychloroquine paper in Lancet and the change of their editorial policy this week (not published in detail how this will affect the peer review system), the complexity of this issue becomes even more apparent.

I personally am far from convinced that the current editorial system and peer review is at all adequate or up-to-date to detect scientific fraud and ensure high quality papers. What we see is an inflation in the number of journals opting for a payment system. Thus, the financial incentives for publication and the consequent pressure on editors is increasing. This is not really addressed in this paper. Additionally, the journals are in essence cherry picking. Peer reviewers do the job, researchers do their part and the funding comes from public or private sources. The journals end up making the money. And there is often very little transparency about the quality of editorial system and the peer review process in even very well-established papers. For instance, one never knows the number of peers, name, affiliation, conflict of interest and the extent of data scrutiny. Finally, often data are not provided, shared and even to a lesser degree re-analyzed unless very controversial or with a high clinical impact. And there is often very little effort nor incentive to opt for reproducibility.

What we are witnessed to is a tsunami of useless scientific papers. For instance only 3 percent of

systematic reviews published today have adequate quality and address the issue of random error and reproducibility. Thus, a major limitation of this study is that it only reflects the general quality of journals as a proxy indicator for scientific malpractice and retraction per se.

But overall, an enlightening work that adds valuable information to complexity of the issue of predatory journals and their impact.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: EBM, anaesthesia, critical care

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 31 July 2020

https://doi.org/10.5256/f1000research.26314.r64881

?

Pablo Iriarte
Library of the University of Geneva, Geneva, Switzerland

Floriane Muller
Library of the University of Geneva, Geneva, Switzerland

Nadia Elia
Division of Anaesthesiology, Department of Acute Medicine, Geneva University Hospitals, Institute of Global Health, Faculty of Medicine, University of Geneva, Geneva, Switzerland

Thank you for giving us the opportunity to read this article in which the authors describe the characteristics, citations and metrics of journals that have been indexed in the Scopus database, at some point, and have afterwards been "discontinued" within Scopus for different reasons, summarised as "publication concerns".
Since the articles that have been published before the journal's indexation was discontinued remain in the database, and can still be found, they may still be cited.
The authors find this to be particularly problematic since they believe that these journals may be what are often called "predatory journals", and therefore may threaten the credibility of science, by polluting the database with "weak research".
Therefore, the authors aimed to compare the number of citations per year, and per journal, and per document, before and after the journal was delisted. They conclude that the number of citations was actually higher after the journal was discontinued from Scopus.
Although we understand the problem these authors try to highlight, we have some major concerns regarding some aspects of this study (mainly related to baseline assumptions and lack of clear definition) and also some minor ones.

Major concerns:

Baseline assumption: In this article, the authors suggest that if a journal is being indexed, even for a long period of time (half of them have been indexed for 8 to 54 years), and is encountering "publication concern", then all the previously published articles should become suspicious of bad science. We are not sure this should be considered straight forward, for the reason developed under our second major concern.

"Predatory journals": the problem of the lack of a clear definition of what a predatory journal is, remains. The authors use different sources to try to identify journals as "predatory" and we can only realise that the sources do not seem to agree. Although authors auto cite their own "consensus definition" of predatory journals and publishers "(..) entities that prioritize self-interest at the expense of scholarship and are characterized by false or misleading information, deviation from best editorial and publication practices, a lack of transparency, and/or the use of aggressive and indiscriminate solicitation practices." they fail to underline that not everybody agrees with this definition. Also, the recent COVID-19 debacle of very low-quality scientific publications, published in usually highly regarded journals, suggests that bad peer-review and misleading articles may not be a characteristic of any journal.
Also, it remains unclear to us how a journal may be indexed for 8 to 10 years, and all of a sudden become "predatory". Or was it predatory in the first place, but was only uncovered after such a long time? If this is what the authors suggest, then what should we think about "recently" indexed journals? They may all be predatory as well, and will only be uncovered in 5 to 10 years?

"Publication concerns": This term needs to be better defined in order to really understand what lies behind it. It remains unclear why these journals have been excluded from the Scopus

database at some point. Interestingly, half of these journals have been deemed good enough to figure in the database for more than 8 years... that's a lot! And all of a sudden, they are not judged acceptable anymore and are discontinued from Scopus. Ok, why not. It may take some time before someone alerts Scopus of the misbehaviour of a given journal, although more than 10 years for 25% of them seems a lot. Or could it be a problem behind the vague concept of "publication concerns"? Could it be that the publication has stopped? Or the journal has changed its name? Or has merged with another one? Or has changed in quality over time? Illustrating some of the reason for discontinuation would help the reader understand the context.

According to Scopus' document cited in the article (ref. 5), there are 3 causes prompting Scopus to launch a journal re-evaluation: Under performance – metrics ; Outlier performance – radar ; Publication concerns. It might have been interesting to analyse the journals removed using the criteria "outlier performance – radar" as well as, according to Scopus document (ref. 5 cited by the authors) it is "particularly effective in flagging potential predatory journals." Scopus describes it as "an algorithm that flags journals based on approximately 40 outlier predictors, including sudden change in output volume, sudden change in publishing country and/or affiliations, and high journal/author self-citation rates."

Increase in citations: The authors are worried that the citations of these journal have increased after the journal's indexation was discontinued in the Scopus database. The problem here is that they do not seems to consider the fact that this may be the case for all journals (those indexed and those discontinued) which is probably due to the rapid increase in the number of publications over time. Unfortunately, this study lacks a "control group" (journals whose coverage has NOT been discontinued in the Scopus database) which could have help the reader understand whether the increase in citation of these journals was similar, was higher, or was lower than that of "legitimate journals".

Underlying discourse: The term "inflated" used in the title, in Figure 2 and conclusion suggests manipulation or distortion of citations and an artificial advantage for authors of articles published in predatory journals before they are removed from Scopus. This is not demonstrated by the reasoning and data used in the article as a basis for comparison is missing.

Methods and reproducibility: While the authors have provided data alongside the article, we have not been able to reproduce some of their results, such as "citations per year" presented in table 5. Data presented in "underlying data table 1" would benefit from better variable descriptions, such as where exactly was the information collected from, and the date of its collection.
Some variable names and analysis are misleading, such as "Actual Pubmed", described in methods section as "inclusion in PubMed" and in table 6 as "main database indexing". It does not reflect whether the journal is currently indexed in PubMed, but may in some cases only indicate that a single article is present in PubMed or selected citations, due to their deposit in PMC (eg. "Advanced Materials Letters").
Some data seem a bit bizarre... and information provided by the authors like "Citation before and after the date of discontinuation were manually counted based on either the Scopus journal overview or the downloadable tables made available by Scopus upon request (see source data)" (p.3) did not allow us to double check some numbers that were weirdly extreme, and potential typos. Some counts of the number of citations seem erroneous, leading to an aberrant number of citations per document for journals like "Mental Health in Family Medicine" (80 citations per

document before discontinuation) or "Pharmacognosy Reviews" (170 after).
Other example of bizarre data: according to “underlying data table 1” the journal “Advanced material research” has been indexed for 10 years (from 2004 to 2014) and has received during this period only 3 citations. However, after having been delisted from the Scopus database, during a 6 year period (2014 to 2019), it has received 13875 citations. Any thoughts on how/why this could have happened?

Minor concerns:

Abstract:

- Background: “contains the largest number of abstract and articles…” -> "One of the largest" could be better, some databases are bigger than Scopus (Google Scholar and Dimensions, see https://doi.org/10.1016/j.joi.2018.03.006[1] or https://arxiv.org/abs/2004.14329)
- Background: The term “publication concerns” may not be clear to everyone.
- Background: “These journals remain indexed and can be cited.” This sentence is confusing. The articles published before the exclusion remains indexed and their citations continue to be taken into account but new issues of the journals are not indexed any more.
- Methods: The use of the term “discontinued” both for DOAJ (Results) and for journal publication (Background) is confusing. Should we say “excluded” or “delisted”?
- Results: “317 journals were evaluated” but next sentence states ninety-three percent of the journals (294/318)” -> typo for 318?
- Results: “the mean number of citations per year after discontinuation was significantly higher than before, and so was the number of citation per document”. Unclear whether the median difference of 64 is per journal, or cumulative across all “discontinued” journals? What are “documents”? Do you mean “articles”? or are there any other types of publication?
- Conclusions: it’s unclear how the conclusion regarding “predatory journals” is drawn. Also, we don’t think the career advancement are “artificial”, they are real! Although maybe “undue”?

Introduction:

- “publications from no longer indexed journals may not be removed retrospectively … hence articles … could remain part of the database[7]” p.3 -> this conditional statement seems to contradict abstract which categorically states “These journals remains indexed” as well as the author’s conclusion “we propose that CSAB could apply these measure case-by-case”. Reference 7, linked with statement, was not helpful to clarify.

Methods:

- “Independently collected by eight of the authors in pairs”: not very clear: two by two, or checked by two different people independently?
- “the year of our data collection”: more precision maybe?

Results:

- Why were data from 31 journals not retrieved? What was the problem?

- Table 1: Interestingly, none of these journals (discontinued from Scopus for publication concerns) have been published by Elsevier. Could there possibly be a selection bias? An interesting option could be to check if Elsevier's journals that are still in Scopus might have been discontinued from others sources (DOAJ, WOS), and on which grounds?

- Table 1: Maybe the table could be enhanced to provide information about whether other journals by that publisher are still included in Scopus or not? Eg. 39 journals discontinued from Scopus were published by Academic journal Inc. Are there any other journals by this publisher still in Scopus?

- Table 3: don't need 2 decimal precision in %.

- Table 3: Subject area are repetitive in Scopus, and a journal can have more than one, while this table and underlying data mention only 1 per journal, presumably the first appearing in Scopus ? If so, probably worth indicating and better to remove the percentages in table 3 falsely suggesting mutually exclusive categories?

- Table 5 and results (text): unclear where the median difference of 64 comes from? Or the 0.4?

- Table 5: total number of citation does not match number of citation before and after discontinuation. Any thoughts on how/why this could have happened? A note of explanation about that would be useful.

- Citations by year in table 5: The number of years before and after the journal is removed from Scopus is very different, the average is more than 9 years before and 4 years after (median of 8 and 4 respectively) which makes the comparison in Table 5 not relevant. Indeed, the number of citations per year is higher during the 2 or 3 years following the publication of the article and decreases sharply with time (DOI:10.1371/journal.pone.0153730[2]) so that the ratio of citations per year also decreases if a larger number of years is used.

- Distribution of articles: of the 317 journals analysed, 5 contain more than half of the articles concerned by this question. This very inhomogeneous distribution means that the statistical analyses and the percentages per journal do not take this kind of distribution into account.

- Page 5, first paragraph: table 2 should probably read table 5?

- Page 5, 2nd paragraph: In 243 case (243/317)... is useless here. Maybe the authors meant 76.6% (243/217)

- Table 6: maybe a good idea to separate the "positive" facts (being indexed in Pubmed, WoS or Cabell's whitelist) from the negative ones (Beall's list, Cabell's blacklist DOAJ discontinued.)

Discussion:

- p.7 Unclear why the term "ghost journal" suddenly appear and how it is defined.

- p. 7-8 "Of greatest concern is our finding that many of the discontinued journals display

predatory behaviors in claiming to be open access" -> do you mean "displaying... are claiming"? to our understanding the article does not say that open access systematically means predatory. According to ref 9 et 22, the large majority of DOAJ indexed journals were not found in Beall's list or Cabell's Blacklist.

- p. 8: "Such journals" unclear: predatory journals or OA journals?

- The authors highlight that a limitation of their methodology is that they have included the year of discontinuation in the period "after discontinuation", which could have led to overestimations. Then why not present the 2 analyses with the year of discontinuation included in the period BEFORE and in the period AFTER discontinuation, so that the reader can check for him-herself what bias this has induced?

- A mention of or comparison with other databases' practices with regards to removing journals for indexing could be interesting. Do their approaches differ from Scopus'?

Conclusions:

- Proposals are missing to solve the problem addressed and to avoid the stigmatization of the authors of the "suspect" articles. For example, a new open peer-review for articles published within X months before the journal's exclusion would be a possibility.

- Maybe another idea would be to flag published articles that have been published in journals that are not indexed anymore "NB: this article was been published in YEAR, in a journal that has encountered publication concerns in YEAR"

References:

- Reference 5 URL should be https://www.elsevier.com/__data/assets/pdf_file/0004/891058/ACAD_LIB_SC_ART_Importance-of-high-quality-content_WEB.pdf

- Link in reference 12 does not work properly due to a superfluous space in the middle.

- Reference style is not harmonized (cf. 1st § of methods section makes 4 references to the same underlying data table 1 (ref number 12 in the reference list, 2 of whom are not correct and should refer to source data, the rest having various citation style).

- There might be some mix up in references: ex. §4 on page 3 lists ref 6-8. Is it possible it should read 5, 7-8 instead?

- The literature review would benefit from additional references to complete or contrast with the author's findings: ex. doi:10.3390/publications8020017[3] that concludes that "articles published in predatory journals have little scientific impact."

- Auto-citations: There are different ways to increase a researchers' number of citations or H-index. Publishing in a "predatory journal" may be one of them, but auto-citation is also one. Of the 30 references cited at the end of this paper, 11 (37%) are auto-citations (citation of a reference including at least on author of the present paper), 7 (23%) are articles from others, and the remaining 12 were websites.

Typo/language

- In Underlying data table 1: last column title - DOAJ instead of DOAH ?

- p.3 §3 “journal are no longer be indexed” -> “journal are no longer indexed”
- p.3 §5 “still they can get” -> “they can still get”

References

1. Thelwall M: Dimensions: A competitor to Scopus and the Web of Science?. *Journal of Informetrics*. 2018; **12** (2): 430-435 Publisher Full Text
2. Heneberg P: From Excessive Journal Self-Cites to Citation Stacking: Analysis of Journal Self-Citation Kinetics in Search for Journals, Which Boost Their Scientometric Indicators.*PLoS One*. 2016; **11** (4): e0153730 PubMed Abstract | Publisher Full Text
3. Björk B, Kanto-Karvonen S, Harviainen J: How Frequently Are Articles in Predatory Open Access Journals Cited. *Publications*. 2020; **8** (2). Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 13 Aug 2020

Andrea Cortegiani, University of Palermo, Palermo, Italy

Dear reviewers,
We are glad to submit a revised version of our manuscript, previously entitled ‘**Inflated citations and metrics of journals discontinued from Scopus for publication concerns: the GhoS(t)copus Project**’. The comments you provided were more than helpful in revising and improving the manuscript.

We hereby provide a point-by-point reply to the comments.

Best regards,

Andrea Cortegiani, MD
On behalf of co-authors

Response to Reviewers team 2 (Pablo Iriarte, Library of the University of Geneva, Geneva, Switzerland; Floriane Muller, Library of the University of Geneva, Geneva, Switzerland; Nadia Elia, Division of Anaesthesiology, Department of Acute Medicine, Geneva University Hospitals, Institute of Global Health, Faculty of Medicine, University of Geneva, Geneva, Switzerland)

Comment 1: *Thank you for giving us the opportunity to read this article in which the authors describe the characteristics, citations and metrics of journals that have been indexed in the Scopus database, at some point, and have afterwards been "discontinued" within Scopus for different reasons, summarised as "publication concerns".*
Since the articles that have been published before the journal's indexation was discontinued remain in the database, and can still be found, they may still be cited.
The authors find this to be particularly problematic since they believe that these journals may be what are often called "predatory journals", and therefore may threaten the credibility of science, by polluting the database with "weak research".
Therefore, the authors aimed to compare the number of citations per year, and per journal, and per document, before and after the journal was delisted. They conclude that the number of citations was actually higher after the journal was discontinued from Scopus.
Although we understand the problem these authors try to highlight, we have some major concerns regarding some aspects of this study (mainly related to baseline assumptions and lack of clear definition) and also some minor ones.

Reply: We are very grateful for the insights of the reviewers which have led us to improve our manuscript. We have now submitted a revised version of the manuscript and herein is our point-by-point reply to the comments. English form was also revised.

Major comments:

Comment 2: *Baseline assumption: In this article, the authors suggest that if a journal is being indexed, even for a long period of time (half of them have been indexed for 8 to 54 years), and is encountering "publication concern", then all the previously published articles should become suspicious of bad science. We are not sure this should be considered straight forward, for the reason developed under our second major concern.*

Reply: Thank you for the opportunity to clarify this. We made no such claim in the paper and, in fact, our key message was different. We clearly stated in the discussion that "It

would be unfair to punish researchers for an eventual deterioration in journal performance; changes in the standards employed by the journal may change over time and the researchers may be unaware of quality issues". We aimed to provide an analysis and describe the main scientific features and citation metrics of journals discontinued from Scopus for publication concerns as we strongly believe that this phenomenon merits discussion. We fully agree with the reviewer that further evaluation is required before the best solution for all aspects of this complex issue is determined. In fact, this study is the first to provide some of the information required to answer this question, albeit not all. We also agree with the reviewer regardless of the solution that is decided upon in the future, it should ensure that researchers are not unfairly punished. This is clearly stated. However, as we also point out, this issue can no longer be ignored; it involves a large number of journals and published documents.

Comment 3: *"Predatory journals": the problem of the lack of a clear definition of what a predatory journal is, remains. The authors use different sources to try to identify journals as "predatory" and we can only realise that the sources do not seem to agree. Although authors auto cite their own "consensus definition" of predatory journals and publishers "(..) entities that prioritize self-interest at the expense of scholarship and are characterized by false or misleading information, deviation from best editorial and publication practices, a lack of transparency, and/or the use of aggressive and indiscriminate solicitation practices." they fail to underline that not everybody agrees with this definition. Also, the recent COVID-19 debacle of very low-quality scientific publications, published in usually highly regarded journals, suggests that bad peer-review and misleading articles may not be a characteristic of any journal.*

Reply: First, we would like to highlight that the definition we reported for "predatory" is not the authors' own. It was taken from an international collaboration of 35 authors who extensively studied the topic. Although we agree that no definition is perfect, this is most certainly not something we decided on ourselves and a consensus process was involved in its determination. If the reviewer wishes to argue with the definition provided, this should ideally be taken up with those involved in the consensus process. We surveyed recognized lists (i.e. Cabell, updated Beall, DOAJ) to evaluate the extent of predatory journals among the discontinued journals.
With regards to the comment regarding the quality of COVID-19 research: Very true. We too have been following this topic with great interest. However, two wrongs do not make a right. In fact, this precise issue makes the discussion of journal metrics and our responsibilities towards them even more pertinent. Our research highlights some of the issues that arose with monitoring of the publication process from a different angle. It also promotes the need to continue to increase awareness within the scientific community itself regarding the damage that could potentially be caused by low-quality papers.

Comment 3: *Also, it remains unclear to us how a journal may be indexed for 8 to 10 years, and all of a sudden become "predatory". Or was it predatory in the first place, but was only uncovered after such a long time? If this is what the authors suggest, then what should we think about "recently" indexed journals? They may all be predatory as well, and will only be uncovered in 5 to 10 years?*
Reply: It is our impression that the process may occur in two manners: (1) Some of the more recently indexed journals may indeed turn out to be predatory. So indeed perhaps

newly indexed journals need to undergo more rigorous monitoring than well established journals. Whether our impression is correct and, if so, how this should be done, are questions far beyond the scope of our research; (2) Some of the discontinued older journals probably did deteriorate slowly. Our impression was that this process is typically a "slippery slope" and does not have an abrupt cutoff. As our analysis was not intended to study this question, we prefer not to speculate on the ideal timing for journal discontinuation. More data and expert input is needed on how to identify this process in the future.

Comment 4: *"Publication concerns": This term needs to be better defined in order to really understand what lies behind it. It remains unclear why these journals have been excluded from the Scopus database at some point. Interestingly, half of these journals have been deemed good enough to figure in the database for more than 8 years... that's a lot! And all of a sudden, they are not judged acceptable anymore and are discontinued from Scopus. Ok, why not. It may take some time before someone alerts Scopus of the misbehaviour of a given journal, although more than 10 years for 25% of them seems a lot. Or could it be a problem behind the vague concept of "publication concerns"? Could it be that the publication has stopped? Or the journal has changed its name? Or has merged with another one? Or has changed in quality over time? Illustrating some of the reason for discontinuation would help the reader understand the context.*
Reply: The term 'publication concerns' is not one which spontaneously decided upon. It is the term defined and used by Scopus. Indeed, we report in the manuscript all the available definitions and details provided by Scopus. Unfortunately, no additional details are publicly available regarding the criteria used to discontinue a journal because of 'publication concerns'. We too would be delighted to receive more details as they may be important. Having said this, we honestly doubt that merging with another paper or changing a journal name is cause for publication concern.
With regards to the reviewers' rumination on the time gap for discontinuation: As noted above, it is indeed possible that some journals have changed quality over time or that they were evaluated only several years after indexing. This information would most certainly be interesting if it were publicly available, but it is not. Furthermore, as also stated above, this is not within the scope of our project.

Comment 5: *According to Scopus' document cited in the article (ref. 5), there are 3 causes prompting Scopus to launch a journal re-evaluation: Under performance – metrics ; Outlier performance – radar ; Publication concerns. It might have been interesting to analyse the journals removed using the criteria "outlier performance – radar" as well as, according to Scopus document (ref. 5 cited by the authors) it is "particularly effective in flagging potential predatory journals." Scopus describes it as "an algorithm that flags journals based on approximately 40 outlier predictors, including sudden change in output volume, sudden change in publishing country and/or affiliations, and high journal/author self-citation rates.*
Reply: We thank the reviewer for this insightful comment which suggests some new research directions for the future. We have added this suggestion to the concluding paragraph of our paper.

Comment 6: *Increase in citations: The authors are worried that the citations of these journal have increased after the journal's indexation was discontinued in the Scopus database. The problem here is that they do not seems to consider the fact that this may be the case for all journals (those indexed and those discontinued) which is probably due to the rapid increase in*

the number of publications over time. Unfortunately, this study lacks a "control group" (journals whose coverage has NOT been discontinued in the Scopus database) which could have help the reader understand whether the increase in citation of these journals was similar, was higher, or was lower than that of "legitimate journals".
Reply: Please see below our response to this and the next comment together.

Comment 7: *Underlying discourse: The term "inflated" used in the title, in Figure 2 and conclusion suggests manipulation or distortion of citations and an artificial advantage for authors of articles published in predatory journals before they are removed from Scopus. This is not demonstrated by the reasoning and data used in the article as a basis for comparison is missing.*
Reply: Indeed, the lack of a control group is a study limitation. We now point this out in the discussion section (see page 13). However, the authors have no interest vested in presenting an "underlying discourse" we have taken this comment very seriously. We have now removed the term "inflated" from both the title and the conclusions. We also modified the caption of Figure 2, substituting 'can' with 'may'.

Comment 8.1: *Methods and reproducibility: While the authors have provided data alongside the article, we have not been able to reproduce some of their results, such as "citations per year" presented in table 5. Data presented in "underlying data table 1" would benefit from better variable descriptions, such as where exactly was the information collected from, and the date of its collection.*
Reply: We salute the reviewer for being so thorough as to attempt to reproduce our results.
Our decision to submit the full database for publication and to select an Open Research publishing platform stems from precisely this reason – we would be delighted if this study was repeated and expanded on in the future.
We calculated "citations per year" as the ratio between the total number of citations (before discontinuation plus after discontinuation) and the number of Scopus years. In the revised version of underlying data table 1 we have now added a box with a more detailed description to enable the readers to repeat our analysis.
However, we must point out that online data changes daily. Therefore, in order to reproduce the data to perfection, one would need to know for which one of the 317 journals that we studied - on which day through the duration of the study period we downloaded the data. The overall process took about a month as described in the paper. This issue may render the data not reproducible to the dot. However, at any time of examination, the overall trends should remain the same.

Comment 8.2: *Some variable names and analysis are misleading, such as "Actual Pubmed", described in methods section as "inclusion in PubMed" and in table 6 as "main database indexing". It does not reflect whether the journal is currently indexed in PubMed, but may in some cases only indicate that a single article is present in PubMed or selected citations, due to their deposit in PMC (eg. "Advanced Materials Letters").*
Reply: Thank you for pointing out this omission. We have now revised both the manuscript and underlying data table 1 to specify that we collected data on the inclusion of articles in PubMed. We also changed the title of Table 6 as follows: "Table 6. Discontinued journals' current Open Access policy and the indexing of their articles in major databases".

Comment 8.3: *Some data seem a bit bizarre... and information provided by the authors like "Citation before and after the date of discontinuation were manually counted based on either the Scopus journal overview or the downloadable tables made available by Scopus upon request (see source data)" (p.3) did not allow us to double check some numbers that were weirdly extreme, and potential typos. Some counts of the number of citations seem erroneous, leading to an aberrant number of citations per document for journals like "Mental Health in Family Medicine" (80 citations per document before discontinuation) or "Pharmacognosy Reviews" (170 after). Other example of bizarre data: according to "underlying data table 1" the journal "Advanced material research" has been indexed for 10 years (from 2004 to 2014) and has received during this period only 3 citations. However, after having been delisted from the Scopus database, during a 6 year period (2014 to 2019), it has received 13875 citations. Any thoughts on how/why this could have happened?"*

Reply: Bizarre or not, the data is what it is. Here we describe how the data was collected in detail:

1. We searched the name of each journal using the Scopus database "Sources" page (https://www.scopus.com/sources.uri?zone=TopNavBar&origin=searchbasic), then we opened the page of the journal and selected "view all documents".
2. At the following page, we checked the box "All" to select all the documents and then clicked "View citation overview".
3. At this stage, two options may appear: a) a page with a line chart reporting the number of citations per year, or b) a flag inviting a request for a citation overview download due to the large size of the overview which does not allow on-site display as a line chart.
4. In the a) situation, we manually summed the citations displayed. In the b) situation, we inserted in the request form an email address and Scopus sent us back a .csv file for each request, containing the requested data.

As this detailed description add little to the main text of the paper we have not added it to the description of the methods.

We wish to reiterate the point made in our previous response – that data collection was not completed in a single day. The data for this study required accessing several sources for each paper, documentation of different variables from each source and performance of individual calculations. For this reason, we report the time frame during which the data was collected (24th January - end of February 2020).

Citations are continuously updated as they grow and new journals are indexed in the database, so at any time other than the exact time we downloaded the data, the number of citations would be slightly different than ours.

Specifically, with regards to the examples given for "bizarre" data, we have re-checked these items and our replies are as follows:

1. Our table 1 did not and does not state "*80 citations per document before discontinuation*" for Mental Health in Family Medicine. We report a total of 175 documents, with 213 citations before discontinuation and 14123 after discontinuation. Thus, 1.22 citations per document before discontinuation and 80 after discontinuation.
2. Regarding Pharmacognosy Reviews - We are grateful to the reviewer for picking up on this typo - In the table, the number of citations before discontinuation was

erroneously written as '43451' rather than '4345'. This led to the number of 170 citations per year for the period before discontinuation. We have corrected the resultant calculations. We have also re-checked the database for additional typos (none were found). The main findings of the manuscript did not change after this correction. Nonetheless thank you for pointing out the mistake.

3. Regarding Advanced Material Research - we have re-checked the data and confirm that it is correct. We do not have an explanation for the huge difference between the period preceding and succeeding discontinuation.

Minor comments:

Comment 1: *Abstract:*
Background: "contains the largest number of abstract and articles..." -> "One of the largest" could be better, some databases are bigger than Scopus (Google Scholar and Dimensions, see https://doi.org/10.1016/j.joi.2018.03.0061 or https://arxiv.org/abs/2004.14329)
Reply: Modified as suggested.

Comment 2: *Background: The term "publication concerns" may not be clear to everyone.*
Reply: Thank you for this comment. As this is a Scopus term, we cannot modify it. We have added quotation marks to make this clear to the readers.

Comment 3: *Background: "These journals remain indexed and can be cited." This sentence is confusing. The articles published before the exclusion remains indexed and their citations continue to be taken into account but new issues of the journals are not indexed any more.*
Reply: Thank you for pointing out this misnomer. We have substituted 'journals' for 'previously published articles'.

Comment 4: *Methods: The use of the term "discontinued" both for DOAJ (Results) and for journal publication (Background) is confusing. Should we say "excluded" or "delisted"?*
Reply: The term 'discontinued' is that used in the Scopus database. The label 'coverage discontinued in Scopus' is also displayed on the discontinued journals' page. The downloadable list of journals whose coverage has been discontinued is also named by Scopus as 'Discontinued sources from Scopus'.
As it is important that the labels used in the manuscript remain consistent with official labels and definitions, we felt we could not change the term 'discontinued'.

Comment 5: *Results: "317 journals were evaluated" but next sentence states ninety-three percent of the journals (294/318)" -> typo for 318?*
Reply: Thank you. It was indeed a typo which has been corrected. Once again, thank you for the scrupulous review which has improved our paper.

Comment 6: *Results: "the mean number of citations per year after discontinuation was significantly higher than before, and so was the number of citation per document". Unclear whether the median difference of 64 is per journal, or cumulative across all "discontinued" journals? What are "documents"? Do you mean "articles"? or are there any other types of publication?*
Reply: We calculated the median number of cumulative citations across all discontinued

journals per year of coverage and defined it as 'Citations per year'. We calculated the median number of cumulative citations across all discontinued journals per document ('Citations per document'). We did not give a subjective definition of 'document' but included all the indexed documents provided by the Scopus database. We have added this detailed description in the methods section of the paper, and we also specified the calculations performed in underlying table 1.

Comment 7: *Conclusions: it's unclear how the conclusion regarding "predatory journals" is drawn. Also, we don't think the career advancement are "artificial", they are real! Although maybe "undue"?*
Reply: As a result of this comment we have modified the conclusions to state as follows:
"Journals whose coverage in Scopus has been halted for publication concerns continue to be cited. This paradox may influence scholar metrics, potentially prompting career advancements and promotions. Further studies are needed, also investigating the journals discontinued from Scopus using the criteria "outlier performance – radar", particularly effective in flagging potential predatory journals. Countermeasures should be taken to ensure the validity and reliability of Scopus metrics for both journals and authors due to their importance for scientific assessment of scholarly publishing. Creative thinking is required to resolve this issue without punishing authors who have inadvertently published good quality papers in a failing or predatory discontinued journal."

Comment 8: *Introduction:*
"publications from no longer indexed journals may not be removed retrospectively … hence articles … could remain part of the database7" p.3 -> this conditional statement seems to contradict abstract which categorically states "These journals remains indexed" as well as the author's conclusion "we propose that CSAB could apply these measure case-by-case". Reference 7, linked with statement, was not helpful to clarify.
Reply: Thank you for pointing out that the language in this sentence requires improvement.
We have revised this to read more succinctly:
"The list of the discontinued sources is publicly available and is updated approximately every six months [6]. However, articles published in journals that were discontinued and are no longer indexed, are probably not removed from the Scopus database.
It has been claimed…"

Comment 9: *Methods:*
"Independently collected by eight of the authors in pairs": not very clear: two by two, or checked by two different people independently?
Reply: Agree. We specified that four pairs of authors independently collected the data (i.e. two people independently collected the same quarter of the data. The entire database is the result of eight people collecting the data).

Comment 10: *"the year of our data collection": more precision maybe?*
Reply: Agree. We changed **"**the year of our data collection" with "2020".

Comment 11: *Results:*
Why were data from 31 journals not retrieved? What was the problem?

Reply: The journals were not found on Scopus database using the search tool. This is now also stated in the paper.

Comment 12: *Table 1: Interestingly, none of these journals (discontinued from Scopus for publication concerns) have been published by Elsevier. Could there possibly be a selection bias? An interesting option could be to check if Elsevier's journals that are still in Scopus might have been discontinued from others sources (DOAJ, WOS), and on which grounds?*
Reply: We agree that this would be an interesting finding. In this study, we focused on journals discontinued by Scopus for 'publication concerns', and we afterwards checked whether they had also been delisted from DOAJ. Should the data mentioned by the reviewers regarding WOS be available online, it may worth further investigation. Thank you for this important insight.

Comment 13: *Table 1: Maybe the table could be enhanced to provide information about whether other journals by that publisher are still included in Scopus or not? Eg. 39 journals discontinued from Scopus were published by Academic journal Inc. Are there any other journals by this publisher still in Scopus?*
Reply: The relation between the publisher and de-indexing of articles in Scopus after discontinuation is an important question that should be addressed in further research. We aimed to provide a snapshot of the effect of ongoing article availability, rather than explore publisher and/or Scopus policies associated with journal discontinuation.

Comment 14: *Table 3: don't need 2 decimal precision in %.*
Reply: This has been changed in accordance with the reviewer's request.

Comment 15: *Table 3: Subject area are repetitive in Scopus, and a journal can have more than one, while this table and underlying data mention only 1 per journal, presumably the first appearing in Scopus? If so, probably worth indicating and better to remove the percentages in table 3 falsely suggesting mutually exclusive categories?*
Reply: Indeed the reviewer is correct. The table caption now reads "First subject area as displayed in Scopus. Note: a journal may have more than one subject area".

Comment 16: *Table 5 and results (text): unclear where the median difference of 64 comes from? Or the 0.4?*
Reply: Thank you for asking. These are the results of the Wilcoxon matched-pairs signed rank test. We removed the reference to the table to avoid misunderstandings. The results differ slightly after corrections made following previous reviewers' comments. The raw data are available in underlying data table 1 and the analyses can be easily recreated by the readers.

Comment 17: *Table 5: total number of citations does not match number of citation before and after discontinuation. Any thoughts on how/why this could have happened? A note of explanation about that would be useful.*
Reply: Again- we are grateful for the reviewers' sharp eye. We rechecked the data and found a typo: The number of citations after discontinuation was reported as 607621 when it should have been 607261 (this can be seen in our underlying data table 1 and in the main text). We have corrected this in the new version of the manuscript.

Comment 18: *Citations by year in table 5: The number of years before and after the journal is removed from Scopus is very different, the average is more than 9 years before and 4 years after (median of 8 and 4 respectively) which makes the comparison in Table 5 not relevant. Indeed, the number of citations per year is higher during the 2 or 3 years following the publication of the article and decreases sharply with time (DOI:10.1371/journal.pone.01537302) so that the ratio of citations per year also decreases if a larger number of years is used.*
Reply: This is probably true. Although some papers may undergo resurgence this is probably not common. However, we found no better way of coping with the issue of the different number of Scopus coverage years across journals. And furthermore, this "decay" is likely to be fairly consistent across all journals both before and after discontinuation. As we also added to the study limitations the lack of a control group of non-discontinued journals, we have also inserted a word of caution regarding our results.

Comment 19: *Distribution of articles: of the 317 journals analysed, 5 contain more than half of the articles concerned by this question. This very inhomogeneous distribution means that the statistical analyses and the percentages per journal do not take this kind of distribution into account.*
Reply: We used non-parametric tests to reduce the effect of non normal distribution of data on our findings precisely for this reason. We also present IQRs and ranges to be more informative.

Comment 20: *Page 5, first paragraph: table 2 should probably read table 5?*
Reply: We have removed this reference to table 2 in order to accommodate a previous reviewer's comment (Comment 16).

Comment 21: *Page 5, 2nd paragraph: In 243 case (243/317)... is useless here. Maybe the authors meant 76.6% (243/217)*
Reply: This has been amended. Thank you.

Comment 22: *Table 6: maybe a good idea to separate the "positive" facts (being indexed in Pubmed, WoS or Cabell's whitelist) from the negative ones (Beall's list, Cabell's blacklist DOAJ discontinued.)*
Reply: This was a good idea. We have changed the order of the items in Table 6 accordingly.

Comment 23:
Discussion:
p.7 Unclear why the term "ghost journal" suddenly appear and how it is defined.
Reply: We modified the phrase; the text now reads "The fact that discontinued journals might help to move up in academia is a relevant issue, and has inspired the vignette depicted in Figure 2: "Discontinued journals may inflate authors' metrics lifting them unnaturally and effortlessly."
We decide not to remove the term 'inflate' from Figure 2 caption as the figure is an allegorical and so intentionally exaggerated representation of the phenomenon.

Comment 24: *p. 7-8 "Of greatest concern is our finding that many of the discontinued journals display predatory behaviors in claiming to be open access" -> do you mean "displaying... are*

claiming"? to our understanding the article does not say that open access systematically means predatory. According to ref 9 et 22, the large majority of DOAJ indexed journals were not found in Beall's list or Cabell's Blacklist.
Reply: We have changed the phrase "Of greatest concern is our finding that many of the discontinued journals display predatory behaviours in claiming to be open access" to "Of greatest concern is our finding that many of the discontinued journals display predatory behaviours in claiming to be open access without actually being indexed in DOAJ."

Comment 25: *p. 8: "Such journals" unclear: predatory journals or OA journals?*
Reply: 'Predatory'. This is now written.

Comment 26: *The authors highlight that a limitation of their methodology is that they have included the year of discontinuation in the period "after discontinuation", which could have led to overestimations. Then why not present the 2 analyses with the year of discontinuation included in the period BEFORE and in the period AFTER discontinuation, so that the reader can check for him-herself what bias this has induced?*
Reply: We mentioned the possibility of overestimation in order to be entirely honest. However, our impression, just from eyeballing the data during collection, was that this would not lead to much of a change. More importantly, in the early stage, we planned no such analysis and therefore did not collect the data that would be required to do this analysis. At this stage performing such an analysis practically requires that the data be recollected in near entirety again which is no simple task.

Comment 27: *A mention of or comparison with other databases' practices with regards to removing journals for indexing could be interesting. Do their approaches differ from Scopus'?*
Reply: This question again is one of policy and therefore not in the scope of our study. The reviewers' comments indeed present much food for thought in terms of future research.

Comment 28: *Conclusions:*
Proposals are missing to solve the problem addressed and to avoid the stigmatization of the authors of the "suspect" articles. For example, a new open peer-review for articles published within X months before the journal's exclusion would be a possibility.
Maybe another idea would be to flag published articles that have been published in journals that are not indexed anymore "NB: this article was been published in YEAR, in a journal that has encountered publication concerns in YEAR"
Reply: We appreciate the reviewers' proposals and have now added them to the discussion. Solutions should also address metrics and citations deriving from these journals. We have therefore added that while this may be an immediately implementable temporising measure, additional thought should be dedicated to address of these aspects as well while maintaining fairness.

Comment 29: *References:*
Reference 5 URL should be
https://www.elsevier.com/__data/assets/pdf_file/0004/891058/ACAD_LIB_SC_ART_Importance-of-high-quality-content_WEB.pdf
Reply: Corrected.

Comment 30: *Link in reference 12 does not work properly due to a superfluous space in the middle.*
Reply: Corrected.

Comment 31: *Reference style is not harmonized (cf. 1st § of methods section makes 4 references to the same underlying data table 1 (ref number 12 in the reference list, 2 of whom are not correct and should refer to source data, the rest having various citation style).*
Reply: Now harmonised.

Comment 32: There might be some mix up in references: ex. §4 on page 3 lists ref 6-8. Is it possible it should read 5, 7-8 instead?
Reply: We again salute the work done by the reviewer. An additional critical eye is always helpful. We have re-reviewed the references to ensure that no additional issues exist, but are of course willing to make any additional corrections if such are identified.

Comment 33: The literature review would benefit from additional references to complete or contrast with the author's findings: ex. doi:10.3390/publications80200173 that concludes that "articles published in predatory journals have little scientific impact."
Reply: Thank you. We have now added this reference and discussed it.

Comment 34: *Auto-citations: There are different ways to increase a researchers' number of citations or H-index. Publishing in a "predatory journal" may be one of them, but auto-citation is also one. Of the 30 references cited at the end of this paper, 11 (37%) are auto-citations (citation of a reference including at least on author of the present paper), 7 (23%) are articles from others, and the remaining 12 were websites.*
Reply: Auto-citations are indeed an issue in the literature. However, in this case, we wish to highlight several points that make this comment moot:

1. The reviewer included in their count three references: Underlying data Table 1: Standardized data extraction form with data collected [previously Ref. n° 12], Extended data Appendix 1 [previously Ref. n° 25] and Extended data Table 1. [previously Ref. n°26], that are study materials. F1000research policy explicitly requires adding supplementary study materials and all the underlying data (e.g. databases) in the reference list. Consequently, the real percentage of self quotations is actually lower.
2. This manuscript has a long list of authors, many of which have published (1) together (2) on this topic. Hence the probability of citing previous research by these authors is high. Is the reviewer suggesting that researchers never base their newest work on their previous work? Experts that consistently research any field of research are likely to work together and to cite their own work - each step supports the next.

Despite the nature of this comment and our concern that it may detract from the quality of our paper, we have removed some of the references. The authors of this paper have no need of self citations – for promotion or any other purpose.

Comment 35: *Typo/language*
In Underlying data table 1: last column title - DOAJ instead of DOAH ?
Reply: Corrected.

Comment 36: *p.3 §3 "journal are no longer be indexed" -> "journal are no longer indexed"*
Reply: Corrected.

Comment 37: *p.3 §5 "still they can get" -> "they can still get"*
Reply: Corrected.

Competing Interests: None

Reviewer Report 16 June 2020

https://doi.org/10.5256/f1000research.26314.r63789

Johann Mouton
Centre for Research on Evaluation, Science and Technology (CREST), DST/NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University, Stellenbosch, South Africa

This is a very relevant study in the growing scholarship around predatory publishing. It is one of the first studies that look at how predatory or at least questionable journals that have been delisted from citation database continue to have a presence in academia. More specifically, the paper asks the very important question why databases like Scopus (and others) continue to track the citations of journals that have been removed. This creates a distortion at many levels, including at the individual publication profile level.

I am happy to recommend indexing of this paper as it is (some minor grammatical editing is required).

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bibliometrics; scholarly publishing; science policy; sociology of science

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 13 Aug 2020

Andrea Cortegiani, University of Palermo, Palermo, Italy

Dear reviewer,
We are glad to submit a revised version of our manuscript, previously entitled '**Inflated citations and metrics of journals discontinued from Scopus for publication concerns: the GhoS(t)copus Project'**. The comment you provided was helpful in revising and improving the manuscript.

We hereby provide our reply.

Best regards,

Andrea Cortegiani, MD
On behalf of co-authors

Response to Reviewer 1 (Johann Mouton, Centre for Research on Evaluation, Science and Technology (CREST), DST/NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University, Stellenbosch, South Africa)

Comment: *This is a very relevant study in the growing scholarship around predatory publishing. It is one of the first studies that look at how predatory or at least questionable journals that have been delisted from citation database continue to have a presence in academia. More specifically, the paper asks the very important question why databases like Scopus (and others) continue to track the citations of journals that have been removed. This creates a distortion at many levels, including at the individual publication profile level.*
I am happy to recommend indexing of this paper as it is (some minor grammatical editing is required).

Reply: We thank the reviewer for the comment. We have now edited the manuscript and corrected errors and typos.

Competing Interests: None