Author's response to reviews

Title:"Predatory" Open Access - A longitudinal study of article volumes and market characteristics

Authors:

   Cenyu Shen (cenyu.shen@hanken.fi)
   Bo Christer Björk (Bo-Christer.Bjork@hanken.fi)

Version:5 Date:18 June 2015

Author's response to reviews: see over
Overall comments
Cenyu Shen and Bo-Christer Björk

The reviews were very helpful and constructive. We have thoroughly revised the whole text and many of the figures. The text is now about 15% shorter, and the overall tone is more neutral. Many of the changes are due to the reviews below, but some also came about for other reasons. Some of the results have changed a bit since we increased the number of sampled journals in three of the strata in order to increase the reliability of the results.

Hopefully the manuscript is now acceptable for publication. Due to the extensive revisions we have not yet used external help with the language, but are willing to do so, if required.

Review 1
Reviewer's report: Title:"Predatory" Open Access - An overview
Reviewer:Marion Broome

Reviewer's report: The question posed in this paper are interesting and important to the continuing discussion of open access journal publishing. The approach taken to collecting data to answer the questions was clearly time consuming and thorough given the parameters the authors’ provided.

However, the organization of the paper is often disjointed and rambles in sections (ie introduction and discussion), while in others intermixes content across sections (e.g. findings and discussion).

Authors' comments: We have made substantial cuts in the text, the introduction and discussion have been shortened. The comparative elements in the results section have been moved to the discussion.

There are too many unnecessary derogatory statements and many statements (ie. methods) left unsourced.

Authors' comments: Some statements, which possibly might be interpreted as derogatory, have been removed or changed.
I hope my comments below are helpful to the authors’ in any revision of their paper.

Major compulsory revisions:

Introduction
The introduction to this paper is 6 pages long and the reader gets lost in much of the rambling philosophical discussion about definitions of predatory journals with the authors’ proclaiming their objections to the term without really convincing the reader of their logic.

Authors’ comments: The introduction has been substantially shortened. Our discussion about the term “predatory” has been moved to the discussion and revised.

Six pages without a concise purpose statement is too long for the average reader and I would suggest that much of this content, especially that of authors” opinions, be moved to the discussion section if appropriate.

Authors’ comments: Moves to discussion done. A more concise purpose statement has been put at the end of the introduction.

The first page of the introduction and section on earlier research is useful and should be retained. The style also needs revision as it is very informal. Research questions should follow the purpose statement. However, the 7 RQs could be reduced to broader conceptual categories. It is not clear why some RQ compare data on these journals over time, while others seem to only describe current state.

Authors’ comments: We prefer to have a short purpose at the end of the introduction and the more detailed research questions just before the methods section. That we studied only article volumes longitudinally over time has to do with a number of reasons. The evolution of time of predatory publishing volumes was assumed to have been considerable, as the results showed. Also the development of average volumes per journals was deemed important. Both are easy to gather data about retrospectively. APCs would be difficult to study other than at the time of data collection, and probably they would not have changed over time. To get more longitudinal data about of author countries and publishing speed would have required sampling a lot articles, i.e. would have been very labor consuming. We also suspect there would not have been big changes in these over the short time period.

Methods:
The first paragraph includes a confusing discussion of illness and symptoms. A more straightforward presentation of Beall’s list of the journals should be presented.
Authors’ comments: We revised the description of Beall’s criteria and skipped the use of the terms illness and symptoms in favor of a more neutral tone

The authors’ also intermix findings in this section describing method-another confusing tactic. Findings should be grouped together.

Authors’ comments: Hopefully the manuscript is now better organized.

The sampling description is highly technical and no citation is provided for the original source for this method.

Authors’ comments: Stratified sampling is a very well known technique in statistical analysis and we don’t think a reference to a general text book is needed. On the other hand the use of stratified sampling for studying the characteristics of scientific journal publishing is to our understanding quite novel and we haven’t found any prior reports.

The authors’ allude to some problems with this sampling plan but do not elaborate not justify their choice.

Authors’ comments: We reorganized the sampling subsection to present and discuss our choice of using stratified sampling method. A new subsection called “Limitation” is now added to discuss the limitation of this methods in our view.

Data collection
Grammatical issues throughout eg, “data was”. Needs subheads or better organization for the reader to understand the various fields of data collected and how it was organized.

Authors’ comments: We have double-checked the grammar used in the texts. We have restructured the data collection subsection to make it more concise and to the point.

Findings
Under journals the authors’ estimate that 7% of predatory journals are indexed in Bealls-by what criteria was this estimate made? What did they mean by ‘overlap’ Field of science.

Authors’ comments: We have the data about whether a sampled journal is registered in DOAJ or not and estimate the proportion of existing predatory journals on Beall’ list indexed in DOAJ based on generalizing this.
The authors’ suggest some additional analyses should have been made and reported about the papers classified under general category. Presumably they have that data and should include it here.

**Authors’ comments:** Journals in this category have titles such as “International journal of sciences”. In order to further classify the articles published in them into the more specialised categories, we would need to make quite extensive samples of articles in these journals which would have been classified by inspection. This would add considerably to the research effort, and we don’t feel the possible results would warrant the effort.

Articles published
In this section there are several inconsistencies. The first is that in the text the authors’ state that Table 1 includes data without taking empty placeholders in account and yet Table 1 title is “Annual average article per stratum excluding empty placeholders” suggesting they did take them into account by excluding them. They also state there was a linear increase (grew steadily) but this was only true (according to data in the table) for two of the journal categories.

**Authors’ comments:** We have tried to better explain how we have dealt with “empty placeholders. Since we increased the sample size, the overall average number of articles per journal for the whole population has changed a bit to show a linear growth from 2010 until 2012, but after that we haven’t seen much growth. Within the different strata, only the two smallest strata grew steadily. So we don’t think there exists an inconsistency.

Author’s origin, APC charges and publishing speed
In these sections the authors’ intersperse comparison of their data with that of previous research which is better left for the discussion section instead of distracting the reader with this discussion in one part of the findings.

**Authors’ comments:** We have now followed this advice and moved such comparisons to the discussion section.

Although the table content should not be repeated in the text for each of these sections a more detailed presentation of the data summary could be added and the comparison with other literature saved for discussion.

**Authors’ comments:** We have revised these subsections to present our main results that have been summarized.

Discussion
The opening paragraph is inappropriate. A summary of the overall findings should be added and ‘overzealous’ comments restricted. After reading the first paragraph, which discounted their method and left one wondering why they took such careless approach left me wondering why I would recommend this paper for publication if the authors’ themselves can’t provide a reasoned argument to support their own method.

Authors’ comments: The opening paragraph has been moved later and removed. We justified using the stratified sampling method in the sampling subsection and we added a limitation subsection.

The rest of the discussion seesaws between criticism of J. Beall and cogent discussion of more constructive approaches taken by others dealing with Open Access movement. There is no conclusion section and the authors’ do not offer any insight into what is next in terms of policy or even future research.

Authors’ comments: We do make conclusions about the overall development in the last section. We also discuss the policies of organizations such as DOAJ and OASPA.

Quality of written English: Not suitable for publication unless extensively edited
Statistical review: Yes, but I do not feel adequately qualified to assess the statistics. Declaration of competing interests: I declare I have no competing interests

Review 2
Reviewer: Jean-Claude GUÉDON
Reviewer's report: A. Major compulsory revisions.

1. Language The first important revision is to have the writing of this paper revised in depth, preferably by a native English speaker, or someone with near-native ability in English. Sometimes, the writing is downright opaque. Let me give a couple of examples: a. "It is, however, very difficult to generalize from these figures to all predatory journals since more than half of the included journals were published in India, which might not be the case if all journals from publishers with multiple journals are included as well." This sentence makes little sense. b. "Also generalized to all predatory articles, the overall average APC is only half as high (135 USD) as the average calculated over journals, indicating a clear author preference for lower priced journals, leading to higher publication volumes in these." This sentence is quite opaque. c. "The share coming out of the authors’ own pocket would probably be much higher for predatory journals, since the threshold for asking for institutional or external funding for the APCs in highly questionable journals would seem quite high." What does this mean?
Conclusion: cleaning up the writing of this paper will begin to improve it greatly. It is a sine qua non condition.
Authors’ comments: We are aware that the English of the first version we submitted was not as good at it should have been. We have now make major revisions to the manuscript and would prefer to resubmit it as is, having done our best to improve the language. Getting the help of an external language editor always takes time, and would, given that the holiday season is just now starting in Finland, delay the resubmission, and we hope the language is good enough to judge the substance of the paper. If the editor and the reviewers approve the content changes made, we are quite willing to make a last language check after conditional acceptance.

2. Important questions: a. I am no statistician, but the sampling of journals appears to have been done with some care. However, the determination of the sampling weight of each journal appears as potentially problematic in the following passage: "This method may from a statistical viewpoint have some problems, but the quite diversified results that we obtained seem to warrant our choice." A brief discussion of the validity of the method, and its limitations, would be useful. otherwise, we are faced with a certain degree of apparent statistical sophistication that does not seem sufficient to alleviate doubts about the sampling technique. How are we to trust the rest of the analysis?

Authors’ comments: To better present our sampling method, we reorganised the sampling subsection to better justify about our choice of using a stratified sampling method, in the particular way we did. In addition to it, we added a subsection to discuss the limitations of this method.

b. All the results about article distribution, etc. should be compared to results taken from the Web of Science and from DOAJ. There is a clear lack of control metrics in this paper.

Authors’ comments: We have added comparisons to DOAJ and WoS concerning the overall journal and article volumes

c. The 7% titles that belong both in Beall's list and DOAJ should have been scrutinized further: they might lead to interesting results about what might be called "border-line" cases.

Authors’ comments: This could certainly be worthwhile to study, but the article is quite long as it is.

B. Minor essential revisions

a. If it is difficult to see whether a publisher really is based in the USA, why should we trust any other location, short of saying that everyone wants to appear as if he/she works in the US. This might well be an artefact of Beall's behaviour as he seems to believe that publishers outside the North Atlantic are not to be trusted, almost by default. I remember his taking Hindawi to task, in part because it was based in Cairo.
Authors’ comments: If a publisher states that they are based in say India (and also advertise APCs in the local currency) we don’t think there is any reason not to trust that information. The problem is mainly confined to publishers claiming an address in the US. We don’t think the problem is in predatory publishers awareness of Beall’s list but rather in the academic evaluation practices in a few countries where publishing in “international” journals is overemphasized.

b. The introductory section aiming at clarifying what “predatory journals” are could be shortened.

Authors’ comments: The introduction section has been shortened.

c. On the other hand, Beall's criteria for predatory journals could have been analyzed, scrutinized and criticized. This could have supported an analysis of inherent bias alluded to in the text, toward the end.

Authors’ comments: Beall’s list of criteria would be a good topic for a detailed study, but the list has almost 50 items and an analysis of these would constitute an article in itself. Such an article would be better suited for a specialist journal in information science. This could well be combined with the more detailed study of predatory journals also indexed in DOAJ called for by this reviewer (see above)

NOTE: ADD THIS TO A DISCUSSION OF FURTHER RESEARCH AT THE END OF PAPER

d. The explanation for the small presence of predatory journals in Latin America is very superficial. It should be noted that Scielo is not the only player in town: Redalyc also plays an important role in scholarly publishing in Latin America. Moreover, many journals are published within universities and appear in the web sites of these institutions. To some extent, this institutional branding will limit rogue behaviour because of concern for institutional reputations. Latindex provides a list of several thousand journals that have been vetted according to more than 30 criteria. Many countries in the region use quality lists derived from Latindex (e.g. Qualis in Brazil). Although not perfect, these national lists offer an added barrier to rogue behaviour. On the other hand, not all Scielo journals are in the WoS.

Authors’ comments: We still note the surprisingly low share of Latin America in the text, but don’t attempt to explain it. Rather we point out that this could be an interesting topic to study.

e. The control on articles attempted through "Scholar One" may introduce another form of bias that is not even acknowledged

Authors’ comments: We removed the section reporting acceptance rates for different countries in journals using the Scholar One IT infrastructure.
f. The variations in APC's are not adequately interpreted. It is quite difficult to draw any general lesson from what is said in the paper.

*Authors’ comments: We prefer not to try to interpret the variations, mainly to report them. The only analysis we make is that journals with lower APCs seem to attract bigger submissions numbers (scattergram).*

g. Waivers for APC’s are generally applied by countries and according to GDP/person statistics. On such a scale, India is quite low, but India is sometimes excluded from such waivers on account of its overall size and the existence of a sizeable middle class. This point is not discussed, despite the importance granted to India in this study.

*Authors’ comments: This might possibly be the case, but would require very detailed knowledge of policies of individual OA publishers, policies which might be regarded as almost trade secrets.*

h. The number of authors’ in predatory journals compared to the number of authors’ in WoS-indexed articles does not tell the whole story, given the well-established difficulties of "peripheral" scientists to publish in "core" journals. How about comparing that number with DOAJ articles, particularly with journals that are gratis for authors’.

*Authors’ comments: We are not aware of any studies of the country affiliation distribution of authors’ in DOAJ journals (not to mention gratis ones) to be compared with. If such data were readily available it would certainly be a meaningful comparison, but doing it now would again add a lot of work to the study. The only indexes for which such data is readily available are Web of Science and Scopus and we feel the comparison we make is useful background information. We have added more discussion about the problems of "peripheral" scientists to publish in "core" journals in order to give a more nuanced picture.*

I, p 1: Non-commercial journals are not always run by volunteers (which, I assume, is mentioned to "explain" how such journals can be without fees. Many journals are subsidized, including subscription-based journals, and including journals in rich countries. The case is frequent, for example, in Canada where large programs exist to subsidize journals and monographs. There is no reason to intimate that journals that are gratis for authors’ are somehow "amateurish" or "unsustaintable" because dependent only on "volunteer"

*Authors’ comments: We rephrased the passage to the neutral “Open Access scholarly publishing also includes OA journals without publishing fees”.*
Quality of written English: Not suitable for publication unless extensively edited

Author comments: We promise to do a proper language check, provided that the manuscript is otherwise acceptable for publishing.

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics. Declaration of competing interests: I declare that I have no competing interests

Review 3
Reviewer: Leslie Chan

Reviewer’s report:
Please note that all my comments below could be grouped under “Discretionary Revisions”

1. Is the question posed by the authors’ new and well defined?
I. The so-called “predatory” OA publisher has been a topic of much heated debates in recent years, though as the authors’ pointed out, little empirical research has been done on the phenomenon. So the present study to uncover the magnitude of the problem is timely and necessary, as it fills a significant void in the current literature. While the study aims to discover “the extent of the problem, its causes and possible remedies”, the research question and sub-questions focus primarily on the volume, growth and geographic extent of both the “predatory” journals and those who publish in them. While having a picture of the locations and distribution of the publishers is useful, it would be good if the authors’ provide an explicit rationale as to why this is important to investigate.

Authors’ comments: We think understanding both the regional distribution of publishers and authors’ is an important issue. Some readers might find it interesting, others not, but since we have the results we prefer to report them.

II. The study is also interested in finding out the negative impact of the proliferation of “predatory” publishers, but some explicit measures of negative impact and how they could be measured would be helpful. (More on this point in regard to question 7)

Authors’ comments: We have toned down the discussion about the negative impact of predatory publishing in the revised manuscript, since we haven’t done any research on that issue in this study. Again this could be a topic for further research.

2. Are the methods appropriate and well described, and are sufficient details provided to replicate the work?

I. Given the large number of journals and the varying number of journals by different publishers, the stratified multistage sampling method seems appropriate. However, I am not familiar with this technique and I am also not competent enough to comment on the statistical aspect of the sampling process. Fig. 2 of the sampling process is very helpful indeed. In fact it was very
difficult to understand how the total number of journals “11,873” were arrived at without the diagram. Perhaps the description of the sampling process could be reworded to improve readability.

Authors’ comments: We have described in the beginning of methods section that the first step we had was to review each publisher's website to count the number of journals they have, but to avoid confusion, following that description, we presented that we found 11,873 journals.

II. Another minor language issue: the last sentence in the Identifying Predatory Publishers section is somewhat confusing: “Another small pilot test in terms of articles volumes published in 2014 by 5 each journal randomly selected from big and small publishers indicated that the journals from big publishers often publish fewer articles than those of small publishers.” How many publishers were selected in the pilot test overall? Are there some missing words?

Authors’ comments: In this pilot test, 20 journals were selected from 5 big publishers and 5 small publishers (10 each). We have modified and merged it into the sampling subsection.

3. Are the data sound and well controlled?
I. The Calculation on the average APC charged by sampled journals is reasonable. In the section on Number of Journals, it is estimated that “7% of predatory journals are indexed in DOAJ.” How is this estimate arrived at? Would this need to be revised in view of DOAJ’s recent inclusion criteria?

Authors’ comments: We have the data about whether a sampled journal is registered in DOAJ or not and estimate the proportion of existing predatory journals on Beall’ list indexed in DOAJ based on generalizing this. Some predatory journal have dropped out of DOAJ, but this is a process that takes time. We suspect that the changes in the overlap % have been very minor since we checked this.

With regard to the Field of Science, “Quite noticeable is the large share of Engineering journals.” Equally noticeable is the complete lack of publications in the area of “Physics and Astronomy” Could the authors’ provide an explanation for this?

Authors’ comments: We wanted to single out the large share of articles in Engineering journals, but without trying to explain it. If we started to try to explain the shares of all the discipline categories that would again be a new study, and would require a couple of pages in an already quite long article.

“ Further to the Field of Science, the category “General” needs to be further refined, especially given the very high number in this category. Along this line, is it possible that authors’ publish in
these journals because they are not confined to traditional disciplinary line? If that were the case, finding appropriate reviewers could indeed be a real problem.

Authors’ comments: We already commented on that in answering reviewer 1.

Country of publishers – why is this of interest? Does it matter and why? India’s dominance – what does that imply? (India and Nigeria are post-colonial state…legacy of

Authors’ comments: We already commented on this above. The aim of our study has explicitly been to provide empirical robust information about the quantitative aspects of this phenomenon. We feel obliged to report all the analysis we have done. It will be the job of others to try to interpret the data and to follow up with other studies.

4. Do the figures appear to be genuine, i.e. without evidence of manipulation?
Yes, the figures look genuine without manipulation.

I. Not sure if Fig. 1 of the spam email is necessary.

Authors’ comments: Point taken, we have removed it.

II. Fig. 8 could be improved by lengthening the Y axis.

Authors’ comments: Done

5. Does the manuscript adhere to the relevant standards for reporting and data deposition? The manuscript follows a standard format for reporting. Not sure about data deposition.

6. Are the discussion and conclusions well balanced and adequately supported by the data?
The data generated by the study indeed support the claim that “the problems caused by predatory journals are rather limited and regional, and that the publishing volumes in such journals will cease growing in the near future.” However, some areas of the discussion could be better contextualized to provide better balance.

Authors’ comments: We have made a substantial revision of the discussion section

I. The authors’ acknowledge that it is problematic to rely on Beall’s list, but they do so for practical reasons, and they acknowledge that the list probably contains titles that are not “predatory”. Given the low volume of publications of many single title publishers and the low APC they charge, could it be the case that some of these titles are genuine attempt to start up a journal while experimenting with the APC model? In many low and middle income countries, it is
common place for university departments to start up their own journals, often without financial sustainability. Could this partly explain the large number of single journal publishers from India?

Authors’ comments: Certainly there may be publishers and journals on Beall’s list that genuinely try to work with a proper peer review. But finding such cases among the several hundred journals we sample, not to mention the 11000 + journals we counted, would be a very time-consuming task. Also even if we found that 5% of journals should not be there that would not much affect the total numbers. Note also that the average number of articles per journal of single journal publishers was on the contrary very high, even if some journals may have low numbers.

II. The weakest area of discussion has to do with the potential harm of “predatory” publishers. This may have to do with the lack of a clear definition and ways of measuring harm, as indicated in my comment earlier for question

Authors’ comments: We have tried to tone down this discussion.

1. In the introduction, the authors’ suggest that authors’ who publish in “predatory” journals are largely willing participants and so are their university administrations and so they are not “real victim” of this phenomenon. Later in the discussion, the authors’ suggests that “the major direct effect has probably been that some mediocre academics in a number of countries have met the research productivity metrics imposed by their universities and governments, and have passed evaluations and secured appointments that they would otherwise perhaps not have managed.” And since such developments are highly regional, mainly confined to places like Nigeria and India, it is of no great cause for concerns. Earlier in the introduction, they claim that: “The real victim in this case is respectable open access publishing, the development of which may have been delayed due to the bad image predators have given to OA.” Even though the study did not provide evidence to support this claim. I found this characterization of harm done by “predatory” publishers to be highly imbalance and somewhat insensitive. The authors’ did cite some other studies that suggest that researchers in many developing countries, and particularly former colonies, are forced to seek publications in “International” journals because of demand by their administrations. These studies also point to long term “academic dependency” created by the global North as one of the key reasons why institutions in developing countries are unable to break free from the increasingly globalized and homogenized view of academic excellence based on “where” and how often one publish, instead of “what” is published and whether the results are relevant to local needs. In that sense, these authors’ and their institutions are part of a highly structurally injustice global system that exclude them from publishing in “high quality” journals on the one hand and confine them to publish in dubious journals on the other. They are therefore surely victims of a larger problem of misaligned incentives that is plaguing so much of science and scholarly publishing today. Just because authors’
“choose” to publish in these dubious journals doesn’t make them less of a victim. This is much akin to those living in poverty who are often forced to consume highly packaged and high calorie food that are bad for their health, not because they prefer these food, but because they are too busy to make ends meet and have no time for proper food preparation. I would also add that honest and hard working researchers from low income countries as well as legitimate and struggling journals from these regions are also victims, since they are often unfairly branded because they come from the same regions. Such harms are very real and serious indeed.

Authors’ comments: We have read this comment carefully and the referee makes good points. As a result we have included a few sentences almost verbatim into the manuscript, hope this is OK.

III. “An interesting finding is the very low share of South America both among publisher’s (0.5 %) and corresponding authors’ (2.2 %).” The authors’ suggest that the availability of the SciELO platform supporting Open Access for established local journals could have accounted for the low number of authors’ from the region publishing in “predatory” journals. SciELO was originally a state funded initiative in Brazil and it’s original intent was to provide an alternative system and platform to the northern system (ironically WoS) that have traditionally excluded publications from Brazil. The platform and methodology has since diffused to many Latin American countries as well as South Africa. This distinct regional history and differences with other regions point to the importance of understanding local contexts and the caution needed in making generalizations about the impact of “predatory” publishers. In addition, the difference between Brazil (and other Latin American countries) and countries like Nigeria and India could be highly instructive for comparative studies. It suggests that national policies (like those supporting OA and SciELO in Brazil) could have a decisive influence on the health of science and science dissemination within a country. Likewise, lack of coherent policies could allow the opportunistic development of dubious journals that cause further harm to local science.

Authors’ comments: We have kept the mention of South America but without speculating on the causes (see also comments from reviewer 2)

7. Do the title and abstract accurately convey what has been found? The abstract is comprehensive and accurately convey the findings. The title however, is a bit generic. Given the authors’ reservation about the “misleading” nature of the term “predatory”, could the scholarly community collectively come up with a better term, instead of legitimizing an admittedly dubious term coined by an anti-OA individual?

Authors’ comments: We have changed the title a bit, but have kept the Predatory term. The reason for sticking with it, despite the fact that we do not particularly like is now better motivated in the text.

8. Is the writing acceptable? Yes, with minor exceptions pointed in the comments above.
Quality of written English: Acceptable
Statistical review: Yes, but I do not feel adequately qualified to assess the statistics. Declaration of competing interests: I declare that I have no competing interests.