

Editorial committee BMJ,

Please find enclosed the revised version of our manuscript “Unidentified retracted articles on publishers’ websites, bibliographic databases, academic social networks, and the Sci-Hub black open access website: a problem that should no longer be ignored” which we would like to resubmit for publication in BMJ.

We would like to thank the Editorial committee and the reviewers for their helpful criticism and recommendations that have helped us to improve this paper. As asked by the Editorial committee, the manuscript has been re-written following their comments and recommendations. We take into account all of the reviewers’ comments. We believe that we have addressed the Editorial committee and reviewers’ comments at length and that our revised version will satisfy their recommendations.

We plan to replace the title “Unidentified retracted articles on publishers’ websites, bibliographic databases, academic social networks, and the Sci-Hub black open access website: a problem that should no longer be ignored” by this new one: “Why do retracted papers survive?” Do you agree with this change? Maybe you have another suggestion?

Please note that the wordcount in initial version was 5098 (including references, Table...), and the wordcount in the re-written version is now 4922.

Sincerely,

C. Boudry

---

Dear Dr. Boudry,

Many thanks for sending us your article and giving us the chance to consider your work. We sent your paper out for external peer review and discussed it at a recent Analysis committee meeting (present: Dr Jennifer Rasanathan, Dr Juan Franco, Dr Rachael Hinton, Dr Huseyin Naci, Dr Alison Tonks and myself).

- Thank you so much for your comments and recommendations.

This paper covers an important topic that would be of interest to our readers, and we are interested in publishing on this topic. However, as it is currently written, this paper does not meet the aims of the Analysis section.

Based on comments from the editorial committee and peer reviewers, we believe this manuscript requires substantial revision - if not a total rewrite - for further consideration as a BMJ Analysis paper. If you are willing to undertake this scale of revision in line with our comments, we would welcome a new submission from you on the same topic.

- Following your recommendations, as you can see by comparing it with our initial version, our article has been almost completely redesigned and re-written to meet the aims of the Analysis section.

The reviewers' reports are available at the end of this letter. The editors comments are listed below:

An Analysis paper is deceptively hard to write: it is neither a narrative review or an editorial, but a persuasive paper that draws on evidence to convince the reader of a particular view on an issue where there is some debate (also noting - and/or refuting! - evidence that exists to support alternate or opposing views).

Papers build toward a central thesis, providing evidence to clearly explain why readers should agree with you rather than counter-views or alternatives.

1. The analysis of citations could be more robust. Some concrete examples can help. You mention a paper published in Cell in 2010, which was later retracted, but still appeared to be cited several years later. Do we know if these citations are not simply to say that this article was retracted? An analysis of whether citations were because they were retracted, beyond case reports would be helpful.

- This has been done. A paragraph named "Unidentified retracted articles continue to be cited without reference to the retraction" has been added. Examples are given to show that citations of retracted articles are most often without reference to the retraction. For the paper published in Cell, it has been specified that "Of the 33 post retraction citations, 31 (94%) were positive without reference to the retraction."

2. A better description of RWDB, how exhaustive is its breadth?

- A paragraph has been added to describe RWDB:

"RWDB has been specifically developed to report retracted articles (« Retraction Watch Database » s. d.). It has identified 24140 retracted articles in all disciplines. It is the most comprehensive and largest database of retracted articles and its content is regularly updated (Brainard 2018)."

3. The conclusion and solutions should be clear and concise to send the reader away satisfied and educated.

- The paragraph "What solutions should be considered and possibly implemented to prevent citation of unidentified retracted articles?" is now less detailed and more concise. A Box "Top 3 changes needed and recommendations to prevent citation of unidentified retracted articles" has been added.

4. The mini study is difficult to follow, especially the methods, particularly the contribution of references. The argument rests on this data and so needs to be more clear with references.

- Table 1 has been simplified and a line by reference to PubMed has been added. Furthermore, 3 references were added by reference to the methodology we used that was already used in other studies.

5. The omission of PubMed/MEDLINE needs clarifying.

- PubMed was used as a baseline for our mini study. We added this paragraph in Table 1:

“As done in previous studies (Bhatt 2021; Wager et Williams 2011), PubMed was used as a baseline because it shows the best performance in adhering to procedures for documenting and updating retracted publications and, is considered by the International Committee of Medical Journal Editors (ICMJE) as the authoritative source for information about retractions (« ICMJE | Recommendations | Preparing a Manuscript for Submission to a Medical Journal » s. d.).”

Furthermore, PubMed has been noted to be a potential solution for sites hosting references and/or full-text articles to identify retracted articles in the following sentence:

“Considering partnership with the Retraction Watch DataBase (RWDB) and/or refereeing to PubMed to identify retracted articles could improve this situation. RWDB”

Finally, as asked by reviewer 2, we added PubMed as a solution to prevent researchers from retrieving unidentified retracted articles. The following paragraph has been added:

“To prevent themselves from retrieving unidentified retracted articles, researchers should favor the PubMed database as a source for discovering new articles. Authors can identify retracted articles by searching PubMed for “Retracted publication [pt]”, where the term “pt” in square brackets stands for publication type, or by going directly to the PubMed list of retracted publications (<https://pubmed.ncbi.nlm.nih.gov/?term=retracted+publication+%5Bpt%5D>) (« ICMJE | Recommendations | Preparing a Manuscript for Submission to a Medical Journal » s. d.).”

6. We recommend adding more context to the paper. Can you take a step back and tell the readers about the drivers of this issue (as one reviewer highlights, the publish or perish culture, time pressures during Covid-19)?

- This has been done in the first part of the article which is now named “Why are article retractions so important?” and has been fully rewritten.

Is there any evidence to suggest that retracted articles play a role when developing new research ideas?

- We did not find any articles that support this.

To what extent is this resulting in research waste?

- A specific paragraph has been added in the first part of the article named “Why are article retractions so important?”:

Retractions “represent wasted resources incurring significant financial costs” (Stern et al. 2014). As an example, Stern et al. found that retracted articles due to misconduct

accounted for approximately \$58 million in direct funding by the NIH between 1992 and 2012 (Stern et al. 2014). However, the greatest costs of retraction are preventable illnesses or the loss of human life due to misinformation in the medical literature (Stern et al. 2014).“

Do retracted articles contribute to meta analyses and what is the potential impact.

- In the part named “Why are article retractions so important?”, we added the following sentence:

... a study was done including 229 meta-analyses published between 2013 and 2016 that cited a retracted study. It concluded that “meta-analyses have a problematically high probability of citing retracted articles and of including them in their pooled summaries” (Fanelli, Wong, et Moher 2021)

7. Is there an inequality angle here? You mention that Sci-Hub is more widely used by researchers in low and middle income settings due to limited institutional access to published material. Does this mean that researchers in those settings have the double burden of not having good access to publications and also seeing a potentially biased subset of research that doesn't account for retractions? What are the implications of this?

- You are right. Indeed, researchers in those settings have a double burden. So, in the part named “Publishers’ websites and sites hosting references and/or full-text articles insufficiently identify retracted articles” we added the following paragraph:

“This site is widely used, especially in low income countries having less efficient institutional access to full text “paywalled articles” (Bohannon 2016). It has been shown that Sci-Hub allow clinicians in these countries to obtain essential information and respond appropriately to patient care needs (Boudry et al. 2019). Researchers in those settings thus have the double burden of not having good access to full text articles, and also using a potentially biased subset of articles that does not account for retractions. Improving access to full text articles by enhancing open access to publications should be a priority to overcome this worrying and unfair situation.”

8. We are told on Line 103 that there are specific recommendations from COPE 104 related to article retraction as described since 2011 - this should be set up earlier - what are these recommendations? And how are publishing practices insufficient? And for those readers who don't know COPE, it would be helpful to explain. What is current good practice / approaches that are working? It currently reads as quite one-sided.

- This has been done. The following paragraph has been added in the first part of the article named “Why are article retractions so important?”:

“In order to document retractions, the Committee on Publication Ethics (COPE) (« Retraction Guidelines » s. d.) has published specific recommendations related to article retraction. These recommendations explain in detail which publications should be retracted (e.g. redundant publications, unreliable findings), what form a retraction should take (e.g. clearly identifying retracted articles in all online sources), who should issue the retraction, who is retracting the article (e.g. one author, the publisher) and

the reason for retraction (e.g. fraud, error). These recommendations are meant to help publishers facilitate and standardize their management of the retraction process. “

Insufficient publishing practices are discussed in the part named “Survival of retracted articles: preprint servers, publishers’ websites, bibliographic databases, academic social networks or, illegal black open access websites such as Sci-Hub insufficiently identify retracted articles”

9. Perhaps a top 3 changes needed to address the problem, to keep it concise.

- The last part of the article named “What solutions should be considered and possibly implemented to prevent citation of unidentified retracted articles?” has been simplified and is more concise now after re-writing (e.g. all the parts dealing with Crossref and CrossMark have been removed)

Furthermore, we added a box named “Top 3 changes needed and recommendations to prevent the citation of unidentified retracted articles “

10. The argument needs clarification and should come early, e.g there are gaps / issues with identifying retracted articles which has an impact on xx and xxx. We argue that xx approach is needed to identify retracted articles.

- This has been done in the second part named “Survival of retracted articles: preprint servers, publishers’ websites, bibliographic databases, academic social networks or, illegal black open access websites such as Sci-Hub insufficiently identify retracted articles”

A possible outline would be;

Define what retraction is and why it's important

Highlight current approaches to retraction that are working /guidance e.g. COPE's recommendations for retraction

Highlight the gaps in identifying retracted articles e.g. Evidence suggests guidance / COPE's recommendations aren't being followed, (or there is a lack of consistency in retracting articles) and reasons for this (more than guessing)

Discuss the impact of this e.g. what happens when retracted articles aren't identified

What needs to change and how - based on addressing the gaps

- We have completely reorganized the article into 4 parts to fit with your proposition.

- **Why are article retractions so important?**

-> Define what retraction is and why it's important

-> Highlight current approaches to retraction that are working /guidance e.g. COPE's recommendations for retraction

**- Survival of retracted articles: preprint servers, publishers' websites, bibliographic databases, academic social networks or, illegal black open access websites such as Sci-Hub insufficiently identify retracted articles**

-> Highlight the gaps in identifying retracted articles e.g. Evidence suggests guidance / COPE's recommendations aren't being followed, (or there is a lack of consistency in retracting articles) and reasons for this (more than guessing)

**- Unidentified retracted articles continue to be cited without reference to the retraction**

-> Discuss the impact of this e.g. what happens when retracted articles aren't identified

**- What solutions should be considered and possibly implemented to prevent citation of unidentified retracted articles?**

From the standpoint of a publisher

From the standpoint of sites hosting references and/or full-text articles

From the standpoint of researchers

-> What needs to change and how - based on addressing the gaps

We hope these changes will fit with your expectations.

So, our topline comment to you is to reflect on and revise the central thesis you wish to get across

Please let us know in the next two weeks if you intend to submit a revised version of this manuscript; otherwise we will consider other papers we receive on the same general topic.

If you have any questions about these comments or if you would like further clarification, I am happy to have a call with you to discuss this decision. We appreciate your understanding that at this stage, we cannot guarantee eventual acceptance of a revised paper.

-----  
**Reviewer: 1**

Recommendation:

Comments:

1. When building the rationale for the present analysis, the authors may consider recalling the role played by the 'Publish-or-perish' research culture, and how pressure to publish leads researchers to cut corners, which is pretty alarming, indeed.

- This has been done in the first part of the article which is now named "Why are article retractions so important"? The following sentence has been added:

“It has been suggested that pressure to ‘publish or perish’, and its ensuing increased competition may contribute to an atmosphere in which some individuals could be tempted to selectively report results, or worse, commit outright fraud, both of which may lead to article retraction” (Casadevall et Fang 2012)”

2. I find it awkward that the Authors never mention the magnificent job made by the NLM for MEDLINE and PubMed when it comes to identify retractions. There’s only a small note almost hidden in Table 1 caption. This should be clearly described in the running text. As an author, I am not particularly happy with how PubMed and PMC are dealing with predatory journals, still what they are doing with the retraction policies is quite an impressive action.

- PubMed was used as a baseline for the identification of retracted articles in our mini study. You are right, mentioning its performances concerning the identification of retracted articles is very important and lacking in the first version of our manuscript. So we added several mentions of PubMed in the running text.

The following paragraph has been added in Table 1:

“As done in previous studies (Bhatt 2021; Wager et Williams 2011), PubMed was used as a baseline because it shows the best performance in adhering to procedures for documenting and updating retracted publications and, is considered by the International Committee of Medical Journal Editors (ICMJE) as the authoritative source for information about retractions (« ICMJE | Recommendations | Preparing a Manuscript for Submission to a Medical Journal » s. d.).”

Furthermore, PubMed has been noted to be a potential solution for sites hosting references and/or full-text articles to identify retracted articles in the following sentence:

“Considering partnership with the Retraction Watch DataBase (RWDB) and/or refereeing to PubMed to identify retracted articles could improve this situation.”

Finally, we added PubMed as a solution to prevent researchers from retrieving unidentified retracted articles. The following paragraph has been added:

“To prevent themselves from retrieving unidentified retracted articles, researchers should favor the PubMed database as a source for discovering new articles. Authors can easily identify retracted articles by searching PubMed for “Retracted publication [pt]”, where the term “pt” in square brackets stands for publication type, or by going directly to the PubMed list of retracted publications (<https://pubmed.ncbi.nlm.nih.gov/?term=retracted+publication+%5Bpt%5D>) (« ICMJE | Recommendations | Preparing a Manuscript for Submission to a Medical Journal » s. d.)

3. The Authors make their case by referring to resources such as RG, Scholar, Sci-Hub, that are not the primary source of information for scholars and authors. Adding to my comment #2, the most frequently consulted online scientific medical resource in the world is MEDLINE®/PubMed, however the Authors chose not to mention this database, which is also publicly available.

- As requested, we have mentioned PubMed. Please see above for our response.

4. Line 129: Google Scholar lacks quality control and clear indexing guidelines, this should be conveyed to the reader.

- This has been done in the following sentence:

“Google Scholar is considered as the most comprehensive source of scientific information (Halevi, Moed, et Bar-Ilan 2017), but is also well known to lack quality control and clear indexing guidelines (Falagas et al. 2008).”

5. Line 150: Sci-Hub: I find it somewhat non-sense reading that Sci-Hub has not implemented any monitoring mechanisms to alert readers to retractions. Isn't it an openly illegal initiative? I do understand that this is becoming a vital source for scholars from low-income countries, and that much more should be done to advance creative commons and open access policies (and their true impact); still I see no relevance in appraising ethical issues in such a controversial initiative. Please, provide more justification for including Sci-Hub in the analyses.

You are right. Sci-Hub is an openly illegal initiative, and as such has no obligation whatsoever, to implement monitoring mechanisms to alert readers to retractions. Nevertheless, as noted, this site is widely used in low income countries having less efficient institutional access to full text “paywalled articles”. In this situation, the fact that Sci-hub has not implemented monitoring mechanisms to alert readers to retractions has important negative consequences for researchers situated in these countries. They have the double burden of not having good access to full text articles and also using a potentially biased subset of articles that doesn't account for retractions. In our opinion, this must be noted and researchers in this situation should be alerted. This justifies including Sci-Hub in this analysis.

We have added the following paragraph:

“This site is widely used, especially in low income countries having less efficient institutional access to full text “paywalled articles” (Bohannon 2016). It has been shown that Sci-Hub allow clinicians in these countries to obtain essential information and respond appropriately to patient care needs (Boudry et al. 2019). Researchers in those settings thus have the double burden of not having good access to full text articles, and also using a potentially biased subset of articles that does not account for retractions. Improving access to full text articles by enhancing open access to publications should be a priority to overcome this worrying and unfair situation.”

6. Line 173 and foll. Among the solutions to the issue that the authors put forward, adequate attention should be devoted to highlight how PubMed could serve as the reference source for verifying whether a retraction notice exists.

- As mentioned above, PubMed has been noted to be a potential solution for sites hosting references and/or full-text articles to identify retracted articles in the following sentence:

“Considering partnership with the Retraction Watch DataBase (RWDB) and/or refereeing to PubMed to identify retracted articles could improve this situation. RWDB”



PubMed has been noted as a solution to prevent researchers from retrieving unidentified retracted articles. The following paragraph has been added:

“To prevent themselves from retrieving unidentified retracted articles, researchers should favor the PubMed database as a source for discovering new articles. Authors can easily identify retracted articles by searching PubMed for “Retracted publication [pt]”, where the term “pt” in square brackets stands for publication type, or by going directly to the PubMed list of retracted publications (<https://pubmed.ncbi.nlm.nih.gov/?term=retracted+publication+%5Bpt%5D>) (« ICMJE | Recommendations | Preparing a Manuscript for Submission to a Medical Journal » s. d.)

7. Are the Authors aware of the tools that are available in the PubMed query to spot retractions and have a comprehensive list of them? For instance:

Retracted and republished in:	hasretractedandrepublishedin
Retracted and republished from:	hasretractedandrepublishedfrom
Retraction in:	hasretractionin
Retraction of:	hasretractionof

By typing, for instance, the appropriate PubMed Format tags (e.g., ROF, which stands for retraction of) AND author’s name with the TAG [Au], one can verify.see: <https://pubmed.ncbi.nlm.nih.gov/help/>

- Yes, we are aware of these search strategies (C.B. provides training on PubMed to colleagues in Caen university). However, from our point of view, it is very complicated to introduce such tools in our article because this will lead us into complicated explanations concerning the querying of the PubMed database, knowing that this is not necessarily the objective of our article, nor that of this type of 'analysis' article.

Nevertheless, as mentioned above, we have added the following sentence explaining how to identify retracted articles in PubMed:

Authors can identify retracted articles by searching PubMed for “Retracted publication [pt]”, where the term “pt” in square brackets stands for publication type, or by going directly to the PubMed list of retracted publications (<https://pubmed.ncbi.nlm.nih.gov/?term=retracted+publication+%5Bpt%5D>) (« ICMJE | Recommendations | Preparing a Manuscript for Submission to a Medical Journal » s. d.)

8. In this reviewer’s opinion adding these strategies and tools would significantly enhance the educational weight of the present submission.

- Thank you so much for your suggestions and recommendations

---

**Reviewer: 2**

Recommendation:

Comments:

The authors are dealing with a topic of growing importance: the retraction of scientific articles. Their work has been well documented and they have adopted an objective stance toward all issues they have discussed, including the most challenging one due to its illegal nature, such as the use of Sci-Hub. By doing this, they have presented several perspectives that will be very useful for readers and they have included an analysis that has been directed in general at the research ecosystem as well as at the various stakeholders, with recommendations for actions. I do not have any significant objections with the contents of the manuscript or with its format, except for the following comments and recommendations, which authors can take into consideration for submitting a revision:

In page 1, line 28, correct the fragment "identified this retracted article and FM was able to identify it" (avoid using identify twice)

- We replace identify by recognize in the sentence:

"Zotero had identified this retracted article and FM was able to recognize it".

In page 5, line 137, where it says "mainly indexed by researchers in their profiles". I am not sure that is the most precise way of saying it. Indeed they are indexed in Researchgate's database, but more precisely articles are published by their authors.

- As the article has been largely re-written as asked by the Editorial committee, this sentence no longer appears in the revised version.

In page 8, line 174, correct the fragment "Data shown in this analysis show" (avoid using show twice)

- As above, this sentence no longer appears in the revised version.

Page 8, line 180, you can just say "COPE retraction guidelines"

- This has been done.

Page 9, after reading the first paragraph I wondered, how many articles have been retracted? Is there an approximate number? Perhaps the authors wish to express at least the number that Retraction Watch has about this. I think it would be useful for readers to better grasp the issue and to see that it is very serious. The authors have noted the increase in retractions, but perhaps it can be useful to give an approximate number. Or is this approximate number 5,000 as stated in line 222? If so, I did not find it so clear.

-This information has been added with the presentation of the RWDB:

"RWDB has been specifically developed to report retracted articles (« Retraction Watch Database » s. d.). It has identified 24140 retracted articles in all disciplines. It is the most comprehensive and largest database of retracted articles and its content is regularly updated (Brainard 2018)."

Page 9, after reading the second paragraph I wondered about SciHub. The way I believe it still works (unless this has changed) is that it downloads a copy of a given

article on demand (when a user requests it) and it adds this copy to its repository for future requests of the same article. If the article is requested before being retracted, then there are no retraction marks (obviously), but this is the copy that remains in SciHub's repository.

- Yes it is the case.

Hence, I believe that SciHub has to find a more elaborate workaround, given the way it works. However, the same solutions you are proposing may be valid, just run a check against CrossMark and just add a watermark or something...

- You're right, solutions proposed may also be efficient for Sci-Hub.

Page 10, line 248, in the fragment "accept to implement in their daily practice". I think it needs an "it" or "this" "accept to implement this in their daily practice".

- As the article has been largely re-written as asked by the Editorial committee, this sentence no longer appears in the revised version.

Thank you so much.