

Editorial committee, BMJ

It is an honor to have been reviewed by BMJ again. 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 all comments and recommendations. We take into account all of the reviewer' comments. We believe that we have addressed the Editorial committee and reviewer' comments at length and that our revised version will satisfy their recommendations.

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

Our BMJ analysis article was reviewed by 2 independent evaluators (Andrea Manca and Juan D. Machin-Mastromatteo) in June 2022, and we responded to their criticisms in August 2022. According discussions with Dr. Rasanathan we added other changes in early 2023. We then resubmitted the article in February 2023. This version was then reviewed by another independent reviewer, Caitlin Bakker. She has issued some minor criticisms to improve our article and we want to thank her for that. However, she also issued a major criticism which seems to us to be unjustified regarding the evolution of our article. Indeed, she criticizes for the lack of explanations concerning the methodology we used to obtain the results presented in our article. For example, "readers are given no information regarding how the articles were randomly selected, what information was gathered from the individual articles and how that information was gathered, how data quality was assured, and so forth" and "The identification of retracted status is simply dichotomized as either identified or unidentified, but we have no information regarding what constitutes being identified. If a retracted publication has its publication type changed to "retracted" but neither its title nor abstract is annotated to indicate this, does this qualify as being identified? If the retraction notice were to simultaneously appear when searching for this article, would that constitute being identified as retracted?"

We would like to point out that the detail of our methodology was integrated in the legend of Table 1 into the 1st version that we submitted in April 2022 as shown below:

A search for retracted articles was done on PubMed on April 8th 2021. A total of 8559 journal articles were retrieved from which 500 were randomly selected. For this, a random sequence of integers between 1 and 8559 was generated using Random.org and attributed to the 8559 articles. An ascending classification was performed based on these random integers, and only the first 500 articles were considered for analysis. A check for the presence of any information allowing identification of retraction in references and full-text articles was performed for each of these 500 articles and on each of the sites or services studied. Digital Object Identifiers (DOI) or titles of articles were used as queries. All searches were conducted between April and July 2021. Full-text paywalled articles on publishers' websites were searched using the institutional access of our university. Searches using Sci-Hub were performed in France using non-university internet access. No university or institution affiliated with the authors of this article were therefore involved in downloading articles via Sci-Hub.

According our discussions with Dr Rasanathan, the whole "methodology" part was removed to better fit to BMJ analysis articles type, and therefore logically no longer appeared in the version evaluated by Caitlin Bakker.

Thus, to follow the recommendation of Caitlin Bakker we have therefore added, as requested, this methodology part in the legend of Table 1.

Please note that the wordcount in this final version is 2329 (excluding the references and words in boxes and tables).

Sincerely,

C. Boudry

F Mouriaux

K. Howard

## Editors comments

1. Please lay out more clearly the impact and implications for patients of a paper being retracted but not identified as such. How have patients been put at risk? What are the potential consequences in medicine for patient health? Could this lead to a rise in scientific misconduct? Why should the reader be invested? The connection between retracted articles and their implications needs to be strengthened so that readers understand the importance of the issue.

*We have replaced the following paragraph line 84:*

*“Moreover, article retractions participate in scientific error propagation and may put patients at risk especially when clinical papers are based on findings from retracted papers (Marcus 2018; Frampton, Woods, and Scott 2021, Steen 2011), or when retracted articles continue to be cited without reference to the retraction (da Silva and Bornemann-Cimenti 2017; Bornemann-Cimenti, Szilagyi, and Sandner-Kiesling 2016; Wright and McDaid 2011; da Silva and Dobranszki 2017; Bar-Ilan and Halevi 2018; 2017), also in clinical trials reports (Schneider et al. 2020) and meta analyses (Fanelli, Wong, and Moher 2021).”*

*By:*

*“Moreover, article retractions may put patients at risk (Marcus 2018; Frampton, Woods, and Scott 2021). Stern et al. also pointed out that “some studies indicate that hundreds of thousands of patients have been placed at risk of improper medical care due to enrolment in fraudulent studies or the administration of treatment based on fraudulent studies” that were retracted (Stern et al. 2014). Article retractions also participate in scientific error propagation when retracted articles continue to be cited without reference to the retraction (da Silva and Bornemann-Cimenti 2017; Bornemann-Cimenti, Szilagyi, and Sandner-Kiesling 2016; Wright and McDaid 2011; da Silva and Dobranszki 2017; Bar-Ilan and Halevi 2018; 2017), also in clinical trials reports (Schneider et al. 2020) and meta analyses (Fanelli, Wong, and Moher 2021).”*

2. We also had concerns about the accuracy of some of your conclusions regarding the impact of the pandemic on article retractions. Given that it takes time for reliable retraction rates to be known, it's probably more appropriate for you to discuss retractions during the pandemic over time: early concern, but latest data shows no overall change from baseline. Please see our recent feature on this here (for which the author interviewed the Director of Retraction Watch).

*This has been done. Line 65, the paragraph:*

*“The COVID-19 crisis has been the source of a significant number of article retractions (Frampton, Woods, and Scott 2021; Cortegiani et al. 2021; Anderson, Nugent, and Peterson 2021), with high retraction rates compared to some related research fields for example up to four times higher compared to other infectious diseases such as HIV, H1N1 or Ebola.*

*Has been completed:*

*“The COVID-19 crisis has been the source of a significant number of article retractions (Frampton, Woods, and Scott 2021; Cortegiani et al. 2021; Anderson, Nugent, and Peterson 2021), with high retraction rates in the earlier and acute phase of COVID compared to some related research fields for example up to four times higher compared to other infectious diseases such as HIV, H1N1 or Ebola (Yeo-Teh and Tang 2021). Three years after the outbreak of the pandemic, however, it is now estimated that the retraction rate of COVID-19 articles is “consistent with the expected overall rate of retraction” (Clark 2023).”*

*Line 130 (in Box 1):*

*We added “the acute phase of” in the sentence:*

*“For journals’ part, accelerated publication of research, as it was the case during the acute phase of the COVID-19 crisis (Anderson, Nugent, and Peterson 2021; El-Menyar et al. 2021; Watson 2022) may be associated with less rigorous peer-review, (Teixeira da Silva, Bornemann-Ciment, and Tsigaris 2021) further increasing the risk of retraction.”*

Relatedly, we were concerned that some of your text did not accurately reflect findings in the papers you cited. Please carefully cross-check all references with your text to ensure you are accurately sharing data with readers.

*This has been done. All references have been cross-checked, and modified if necessary:*

*Line 63, the sentence:*

*“Although retractions are still rare, with around 5 retractions per 10,000 articles published (Brainard 2018; Bhatt 2021; Gaudino et al. 2021), numerous papers have highlighted an increasing rate of retractions over time— in fact, it has doubled in the last ten years (Brainard 2018; Bhatt 2021; Van Noorden 2011; Fanelli 2013; Grieneisen and Zhang 2012) (see Box 1)”*

*Has been replaced by:*

*“Although retractions are still rare, with around 5 retractions per 10,000 articles published (Brainard 2018; Bhatt 2021; Gaudino et al. 2021), numerous papers have highlighted an increasing rate of retractions over time (Brainard 2018; Bhatt 2021; Van Noorden 2011; Fanelli 2013; Grieneisen and Zhang 2012) (see Box 1)”*

*Because the references (Brainard 2018; Bhatt 2021; Van Noorden 2011; Fanelli 2013; Grieneisen and Zhang 2012) show an increasing rate of retraction over time without necessarily mentioning a doubling of the rate over the last 10 years.*

*Line 69:*

*The reference “Yeo-Teh NSL, Tang BL. An alarming retraction rate for scientific publications on Coronavirus Disease 2019 (COVID-19). Account Res 2021;28:47–53. doi:10.1080/08989621.2020.1782203”*

*has been added at the end of the sentence “... compared to some related research fields for example up to four times higher compared to other infectious diseases such as HIV, H1N1 or Ebola.”*

*This reference has been accidentally removed during the previous revisions. Please apologize for this mistake.*

*Line 71:*

*The reference “Clark J. How covid-19 bolstered an already perverse publishing system. BMJ 2023;380:p689. doi:10.1136/bmj.p689”*

*has been added with the sentence “Three years after the outbreak of the pandemic, however it is “estimated that just 0.07% of Covid papers have been retracted—consistent with an expected overall rate of retraction””*

*Line 80:*

*The reference “Brainard J. Rethinking retractions. Science 2018;362:390–3. doi:10.1126/science.362.6413.390)”*

*has been replaced by “Marcus A. A scientist’s fraudulent studies put patients at risk. Science 2018;362:394–394. doi:10.1126/science.362.6413.394-a).” at the end of the sentence : “Moreover, article retractions may put patients at risk (Frampton, Woods, and Scott 2021; Marcus 2018; Steen 2011b) ”*

*These 2 articles are part of a series of articles on retractions published in Science, and we pointed to the main article of this series (Brainard), while that of Marcus is more appropriate.*

3. Some of your claims may need greater justification, e.g. Is there any evidence or a particular reason why COPE guidelines are not being followed?

*This has been done line 105. The paragraph:*

*“While the Committee on Publication Ethics (COPE) (“Retraction Guidelines” n.d.) has published specific recommendations to help publishers facilitate and standardize their management of the retraction process, publishers may not correctly identify retracted articles as such on their website (Wright and McDaid 2011; Decullier et al. 2013; Bakker and Riegelman 2018; Dal-Ré and Ayuso 2021).”*

*Has been replaced by:*

*“The Committee on Publication Ethics (COPE) (“Retraction Guidelines” n.d.) has published specific recommendations to help publishers facilitate and standardize their management of the retraction process. Several studies have shown that these recommendations are inconstantly applied: e.g. publishers may not correctly identify retracted articles as such on their website (Wright and McDaid 2011; Decullier et al. 2013; Bakker and Riegelman 2018; Dal-Ré and Ayuso 2021), although this is clearly stated in COPE guidelines. This could reflect “an inadequate awareness of the existence of the COPE guidelines” (Frampton,*

*Woods, and Scott 2021) which could explain why they are not well respected (Decullier et al. 2013)."*

4. Similarly, the methods used for gathering information on retracted articles have raised some questions on validity and reliability and this needs to be addressed (see Bakker comments).

*This has been done. Please see the introduction of our rebuttal letter and our response to the reviewer.*

5. Finally, we would ask you to provide greater nuance and balance by acknowledging that there are other reasons for retracted articles not being publicised (e.g. reviewer Bakker).

*This has been done. Please see our response to the reviewer.*

Minor comments :

- Line 76 - what do issues with journal or publisher mean?

*The sentence:*

*"In 2020, the most common reason for retraction of biological and medical research articles was scientific misconduct (62.2%), followed by error(s) in the manuscript (37.4%) and issues with the journal or publisher (19.4%)".*

*Has been replaced by:*

*"In 2020, the most common reason for retraction of biological and medical research articles was scientific misconduct (62.2%), followed by error(s) in the manuscript (37.4%) and issues with the journal or publisher such as accidental duplicate publication, publication in wrong journal or preliminary version accidentally published without final author correction (19.4%)".*

- The point on 'wasted resources' in the introduction is interesting in terms of the impact, and could be expanded.

*Line 81, the following sentence has been added:*

*"These authors also note the cost of unproductive research by researchers which articles are retracted or based their work on retracted articles and, those related to*

*publishing process of retracted articles (e.g. time spent for reviewing and managing the articles)”*

- The section on 'online survival of unidentified retracted papers' repeats similar points to the example from PLoS and Cell in the introduction; this could be pared down to save you words.

*The examples from PLoS and Cell in the introduction deal with post-retraction citations, which is not covered in the 'online survival of unidentified retracted' section. Unfortunately, we couldn't remove anything to save words in this section.*

- Sci-fi hub is a site for the illegal download of full texts and as such is not a database, however has been included as one. Please correct.

*We are sorry but, after careful examination, we did not find a place in the article where Sci-Hub is likened to a database. We found only the terms "Sites hosting references and/or full-text articles", "illegal black open access website", "Black open access (illegal)” to characterize it.*

## Reviewer: 1

Recommendation:

Comments:

While this is an interesting topic, there are several significant issues with this manuscript which lead me to conclude that it is not yet suitable for publication.

Perhaps most crucially, no methods are offered to describe how it was that the authors arrived at their findings. The authors state that "[w]e checked 500 randomly selected retracted articles (from a total of 8559 retracted articles listed in PubMed in April 2021) for the presence of any information allowing identification of retraction in references and full-text articles." Readers are given no information regarding how the articles were randomly selected, what information was gathered from the individual articles and how that information was gathered, how data quality was assured, and so forth.



They also don't speak to the limitations of using PubMed as a data source, given that PubMed has a limited scope in terms of disciplines of interest and does not have perfect indexing of retracted publications. It is impossible to determine the validity of the methods or their findings with this little information.

The table of results is also somewhat perplexing. The identification of retracted status is simply dichotomized as either identified or unidentified, but we have no information regarding what constitutes being identified. If a retracted publication has its publication type changed to "retracted" but neither its title nor abstract is annotated to indicate this, does this qualify as being identified? If the retraction notice were to simultaneously appear when searching for this article, would that constitute being identified as retracted? There are multiple guidelines and existing best practices on this topic, and it's not clear if or how these were utilized. Because of the very limited methods provided, the table as it stands does not feel particularly meaningful. It is also unclear how the authors accounted for duplicate entries for retracted publications. They note the issue of multiple versions in Google Scholar in an earlier paragraph, but don't provide details of how they account for multiple versions in their calculations, and particularly where some versions may be indicated as retracted and others may not.

*The detail of our methodology was included in the legend of Table 1 into the 1st version that we submitted in April 2022 as shown below:*

*A search for retracted articles was done on PubMed on April 8th 2021. A total of 8559 journal articles were retrieved from which 500 were randomly selected. For this, a random sequence of integers between 1 and 8559 was generated using Random.org and attributed to the 8559 articles. An ascending classification was performed based on these random integers, and only the first 500 articles were considered for analysis. A check for the presence of any information allowing identification of retraction in references and full-text articles was performed for each of these 500 articles and on each of the sites or services studied. Digital Object Identifiers (DOI) or titles of articles were used as queries. All searches were conducted between April and July 2021. Full-text paywalled articles on publishers' websites were searched using the institutional access of our university. Searches using Sci-Hub were performed in France using non-university internet access. No university or institution affiliated with the authors of this article were therefore involved in downloading articles via Sci-Hub.*

*This whole “methodology” part was removed following exchanges we have had with the 2 first reviewers and Dr Rasanathan (BMJ clinical editor) to better fit to BMJ Analysis articles type, and therefore logically no longer appeared in the version you evaluated*

*To respond to your request, we have added this methodology part (in the legend of Table 1). We hope this will suit you.*

*- Regarding your comment on Pubmed, we modified the previous version of article at the request from Andrea Manca (reviewer) who asked us to highlight “the magnificent job made by the NLM for MEDLINE and PubMed when it comes to identify retractions”, and “...what they are doing with the retraction policies is quite an impressive action.” We have therefore modified the article and integrated the following sentence to respond to this request:*

*“PubMed best adheres to procedures for documenting and updating retracted publications and is considered “the authoritative source for information about retractions” (« ICMJE | Recommendations | Preparing a Manuscript for Submission to a Medical Journal » s. d.)”*

*Furthermore, studies which have used PubMed to study article retractions in biological and medical science are numerous. Please see as examples “Wager E, Williams P. Why and how do journals retract articles? An analysis of Medline retractions 1988-2008. Journal of Medical Ethics 2011;37:567–70.*

*doi:10.1136/jme.2010.040964 or 1 Bhatt B. A Multi-perspective Analysis of Retractions in Life Sciences. bioRxiv 2020;:2020.04.29.063016.*

*doi:10.1101/2020.04.29.063016.*

*To our point of view, PubMed, as pointed by the ICMJE, is a relevant source about retractions.*

*Line 206, to respond to your comment “PubMed has a limited scope in terms of disciplines” we added “The bibliographic database specialized in biology and medicine” at the beginning of the sentence “PubMed best adheres to procedures for documenting and updating retracted publications and is considered “the authoritative source for information about retractions””*

*- Regarding your comment "They note the issue of multiple versions in Google Scholar in an earlier paragraph, but don't provide details of how they account for multiple versions in their calculations, and particularly where some versions may be indicated as retracted and others may not": Digital Object Identifiers (DOI) or titles of articles (in case of absence of DOI) were used as queries to search Google Scholar for the 500 articles studied. In this case, Google Scholar offers only one version of the article (the different versions of an article are visible only when you click on "All X versions"). It is therefore this version that researchers have at their disposal when they perform a search on Google Scholar. Indeed, it is on this unique version that we checked for the presence of any information allowing identification of retraction.*

I also note that the authors seem to oversimplify retractions and the communication of retractions. They present two reasons why publishers may not consistently indicate that publications have been retracted: "is it fear for their reputation or ignorance of the seriousness of the problem of retractions?" This is an unfair assertion. It assumes that the inconsistency with which retractions are noted must either be the product of ignorance or self-interest and neglect. It does not account for the fact that international standards regarding the transmission and display of metadata for retracted publications are still under development, and oversimplifies the transmission of information between publishers, aggregators and search engines. There is a similar condemnation of aggregators, as the authors state that "they either never... or imperfectly... update their databases." The authors do not provide evidence that any resource never updates its database, and do not consider the myriad of reasons why resources are imperfectly updated. There is an implicit assumption that this is due to willful neglect or ignorance rather than the possibility that aggregators are also grappling with how to receive, process and display metadata from the broad range of publishers with whom they are working. While I do not argue against the idea that all retracted publications should be indicated as such through all platforms, I do push back against the stated belief that the only reason this is not happening is due to such malicious motives.

*You're right we need to provide greater nuance and balance regarding the reasons why publishers and sites hosting references and/or full texts articles insufficiently identify retracted articles.*

*The paragraph:*

*“Reasons why publishers fail to alert readers to retractions (Frampton, Woods, and Scott 2021) (Frampton, Woods, and Scott 2021) remain unexplained: is it fear for their reputation or ignorance of the seriousness of the problem of retractions? Improvement in adherence to COPE guidelines could improve this situation (“Retraction Guidelines” n.d.; Decullier et al. 2013), but this step alone would of course not solve the problem. (Bakker and Riegelman 2018). Sites hosting references and/or full-text articles do not identify retracted articles as retracted because they either never (e.g. Google Scholar or Sci-Hub) or imperfectly (e.g. Scopus or the WoS) update their databases by themselves. The fact that PubMed can and does update its references to reflect article retractions suggests that this problem has not been recognized or considered a priority by these other sites.”*

*has been modified to fit with your expectations:*

*“As pointed out by Frampton et al., publishers fail to alert readers to retractions because some are not sufficiently “aware of the problem” (Frampton, Woods, and Scott 2021). It is important therefore that the limitations in the retraction process are widely communicated and that publishers share their best practices (Frampton, Woods, and Scott 2021). Improvement in adherence to COPE guidelines could improve this situation (“Retraction Guidelines” n.d.; Decullier et al. 2013), but this step alone would of course not solve all problems. Indeed, there are potential limits to interoperability and the workflows that update articles because international standards regarding the transmission and display of metadata for retracted publications are still under development (Bakker and Riegelman 2018). These difficulties alone do not explain the poor performance of Google Scholar or Sci-Hub in correctly identifying retracted articles. The fact that PubMed can and does update its references to reflect article retractions suggests that article retractions has not been recognized or considered a priority by these other sites.”*

There are other, smaller issues with the manuscript. However, given the significance of the methodological issues, further discussion of minor concerns is unwarranted without significant revision of the manuscript.