

Response to reviewers

Dear editors and reviewers,

We wholeheartedly appreciate the reviewer's careful comments, which have helped us very much to improve the paper. We have revised our paper carefully according to the comments and suggestions. The revised content is highlighted within the document by using the track changes mode in MS Word. We submit here the revised manuscript and a list of changes. Major revisions of original manuscript responses to the comments are listed below.

We hope this edition address most of the comments and suggestions.

Reviewer# 1

This study investigates the effect of retraction on the counts and sentiments of the citation received by the retracted articles. It provides evidence that largely confirmed people's intuition about what happens after the retraction. I think the topic is interesting and that the study indeed helps the readers gain knowledge about the effect of a retraction. However, there is large room to improve before the manuscript is capable of being published in this journal. Below list some of my own points of view and suggestions.

The introduction and literature review need to be reframed. I advise the authors to raise the research questions in the introduction section and highlight its significance.

The literature review should focus on what previous researchers found rather than what they did. Based on a summary of extant knowledge, the authors then identify the research gap. Simply piling up references does not serve the purpose.

Moreover, the hypotheses, especially the latter ones, seem to be drawn from weak logic and literature basis.

-Thanks for your comments on academic writing. We have rewritten these two sections. In introduction, the research questions are raised, and the logic laid out was integrated. In related studies, we rewrote all the references focused on the conclusion in previous studies. Based on the knowledge, we identified the research gap in social science, research methods, and detailed research questions step by step. The logical basis of each hypothesis was enhanced. Related literature published recently was also included to support our hypotheses.

Personally speaking, I suggest the authors avoid using the expression "PSM-DID" or "PSM-DID model" because PSM and DID are two separate procedures which are only sometimes combined to solve some types of research questions. There are also a handful of alternative matching methods other than PSM that can be combined with DID. So, using such expressions can largely add to the confusion of subsequent researchers who just learnt about these methods.

-Thanks for your rigorous comments. All expressions about research methods were revised. PSM and DID were introduced separately and explained in detail in data and methods.

The authors can reconsider the appropriateness of Hypothesis 2. To answer the second research question, I more intuitive value of interest may be the proportion of negative citations rather than

the number.

-Thanks for your insightful comments. Indeed, the count of negative citations can not reflect the change of citation sentiment. Instead, we used the proportion of negative citations to testify the effect of retractions. The expression in the hypothesis was revised. The calculation and results later in this manuscript were also modified.

In DATA AND METHODS, the authors wrote:

“Specifically, we selected articles that were not retracted but had the same citation counts per year before the retraction year, the same source journal, and the same publication year with the retracted articles as the control group.”

From the description, I cannot recognize the matching method as a PSM. It seems like an EM (Exact Matching) where only if all the covariates take on the same value can the cases be matched. In an EM, there is no such thing as propensity score at all. I advise the authors to reexamine their matching method to see whether there is propensity score or not.

-Thanks for your professional comments. We made a comparison among the common matching methods and decided to use PSM in this study. We also did the matching over to make sure all the details were convincing. In PSM model, the *publication year*, *journal*, and *volume* were matched exactly, while the *pre-retraction citation per year* and the *mean value of pre-retraction citation per year* were matched with “nearest neighbor”. Before we performed the matching, *journal* was treated as a string variable. The *publication year* and *volume* were treated as categorical variables. These variables entered the model as a vector to conduct the exact matching. The statement about matching methods in data and methods was rewritten. The related results (regression models) were also revised with the latest version of matching.

Figure 1 should be either reframed or discarded because currently it is not well-organized or helpful.

-Thanks for your comments. Figure 1 was removed.

I advise the authors to read some high-qualified papers in economics that used regression analysis and learn from them how to use subscripts and report the descriptive statistics.

-Thanks for your comments. I changed the improper variable names and reframed Table 2 according to my understanding of papers published in economics area. If there is still anything unsuitable, I am glad to revise them over.

The balance check of the matching outcome should not be done using Wilcoxon test. The authors can refer to doi:10.1111/j.1467-985X.2007.00527.x for detailed reasons.

I do not suggest the authors to test for the parallel trend in order to show the effect of the matching because in Figure 5 there are only data for two years before the retraction. No trend can be observed from only two years at all. If the authors used PSM, it is typical to check the balance of the covariates after matching. If EM is used, there is no need to evaluate the effect of the matching at all.

-Thanks for your professional comments. The basic logic of hypothesis testing is to infer the overall characteristics from the sample characteristics, which is not suitable for the matching sample. To eliminate the potential bias, following your comments, we checked the balance of the covariates after matching. The result is presented in Table 3.

Moreover, it seems that the manuscript was not written with great care. The authors should check for the issues such as inconsistent variable names (e.g., Cit_senti), wrong terminologies (e.g., selection problem), redundant sentences, grammatical errors, etc.

-Thanks for your comments. We have gone through the manuscript. The inconsistent variable names (e.g., NegCitation), wrong terminologies (e.g., selection bias), and grammatical errors are revised.

Another minor issue is that it is still unclear the motivation of selecting psychology as the focal field. I personally think that the reasons currently provided by the authors can suit the context of any scientific field.

-Thanks for your comments. The reason why we chose psychology as the focal field is stated in the second paragraph of related studies. First, the misconducts and retraction practices in psychology took place frequently recently and received attention from the public and academia. Second, retractions from social science are seldom studied by scholars. Misconduct in psychology is representative of other subfields that borrow research methods from hard science, such as “p-hacking”, data fraud, etc. Deepening the understanding of retraction practices in Psychology sheds light on other cognate social science.

Reviewer# 2

Comments to the Author

This manuscript investigates the retraction practice by analyzing the retracted papers in Psychology. The data and method are solid, and the results and findings are robust. Thus, I recommend it for publication after the following minor revisions.

- P5. Line 54 "After the transformation to panel data, we obtained 2148 article-year observations for our research". Please detail the method for the data transformation here.

-Thanks for your comments. We have added details of transforming the cross-section data to panel data in the 4th footnote.

- P11. Figure 3(b). Since the axis Y represent the share of negative citations, please use "%" in the axis.

-Thanks for your comments. We have added “%” in the axis of Figure 3(b) (Figure 2(b) for now).

- Since the sample size is still small, please acknowledge the limitation of this study in the end.

-Thanks for your comments. We have illuminated the limitation of the sample size in the last paragraph of discussion.