Journal of Informetrics 17 (2023) 101423

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

The effectiveness of peer review in identifying issues leading to retractions

Xiang Zheng [a], Jiajing Chen [b], Alison Tollas [a], Chaoqun Ni [a,*]

[a] *Information School, University of Wisconsin-Madison, Madison, WI 53706, USA*
[b] *Department of Computer Science, New York University, New York 10012, USA*

ARTICLE INFO

Keywords:
Retraction
Peer review
Scientific publishing
Reviewer assignment
Reviewer

ABSTRACT

Retractions are necessary to remove flawed research from citable literature but cannot offset the negative impact those publications have on science advances and public trust. The editorial peer-review process is intended to prevent flawed research from being published. However, there is limited empirical evidence of its effectiveness in identifying issues that lead to retractions. This study analyzed the peer-review comments (provided by Clarivate Analytics) for a sample of retracted publications (provided by Retraction Watch) to investigate how the peer-review process effectively detects the areas where the retraction causes lie and whether reviewer characteristics are related to the effectiveness. We found that a small proportion of peer reviews suggested rejections during the peer review stage, while about half suggested acceptance or minor revision for those later retracted papers. The peer-review process was more effective in identifying retraction causes related to data, methods, and results than those related to text plagiarism and references. Additionally, factors such as reviewer seniority and the level of match between reviewers' expertise and the submission were significant in determining the possibility of peer reviews identifying suspicious areas in submissions. We discussed potential insights from these findings and called for collective efforts to prevent retractions.

1. Introduction

Retraction is a self-correcting mechanism for science to remove seriously flawed and published research from the citable literature (Hsiao & Schneider, 2021; Steen et al., 2013). Generally, retraction cannot thoroughly delete or hide the retracted publication from public databases. It is intended to alert readers that the published paper contains seriously flawed or erroneous contents or data that undermine its reliability (COPE Council, 2019). Journals also increasingly use retractions to self-police their content and maintain their scientific integrity and quality (Brainard, 2018; Fang & Casadevall, 2011; Steen, 2011). According to Retraction Watch (RW) (The Center For Scientific Integrity, 2018), the number of recorded retracted research articles spiked from 29 in 2000 to nearly 3000 in 2021. Although retracted papers only account for a small proportion (Fang et al., 2012) of published literature, the potential damage could not be omitted.

Nonetheless, post-publication retractions cannot offset the negative impact on the science advance and public trust (Fang & Casadevall, 2011). Retracted papers may be diffused on social media even after retractions and spread misinformation (Serghiou et al., 2021; Shamsi et al., 2022). Researchers may also circulate and cite these retracted papers before and even after they are identified as flawed, misleading their future research (Bar-Ilan & Halevi, 2018; Bolland et al., 2022; Hsiao & Schneider, 2021; Kühberger et al., 2022; Van Noorden, 2011). Retractions also waste invested human effort, time, and research resources, stigmatize the researchers'

* Corresponding author.
E-mail address: chaoqun.ni@wisc.edu (C. Ni).

https://doi.org/10.1016/j.joi.2023.101423
Received 1 February 2023; Received in revised form 6 April 2023; Accepted 16 May 2023

reputation, and impede their career development (Azoulay et al., 2017). As a more far-reaching impact, the problematic methods and conclusions from retracted papers may have already been put into practice before getting retracted, placing the public at risk (Steen, 2011; Teixeira da Silva et al., 2021). In biomedical fields, because biomedical research often informs life-or-death health decisions, flawed research in this domain also risks public health. Retractions may stigmatize the authors, journals, and associated affiliations, impede the usage of correct knowledge in other papers, and damage public trust in the scientific community("The Science of Retraction," 2002; Byrne, 2019; Lu et al., 2013; Xu & Hu, 2022). These negative impacts are impossible to eliminate by only retracting problematic publications.

Multiple reasons can lead to retractions of scientific publications, including misconduct, scientific mistakes, and administrative errors. Misconduct, such as fraud or suspected fraud, duplicate publication, plagiarism, improper authorship, and failure to follow ethical procedures, is a common reason for retractions (Fang et al., 2012; Marcovitch, 2007). Fang et al. (2012) estimated that 67.4% of the retractions were attributed to misconduct rather than honest error, and the percentage of retractions due to fraud had increased ten times since 1975. Additionally, Campos-Varela and Ruano-Raviña (2019) found that confirmed misconduct made up 65.3% of total retractions. Scientific mistakes, such as honest errors or naive mistakes resulting in unreliable results or data, can also lead to retractions (COPE Council, 2019). Steen (2011) identified an overall rise in retractions and increasing levels of both fraud and scientific mistakes in papers indexed by PubMed between 2000 and 2010. Data, methods, and results are also common areas that cause retractions. Issues such as data fabrication or manipulation and concerns about errors in data or methods are major reasons for retractions (Brown et al., 2022; Nair et al., 2020). An increasing proportion of meta-analysis publications have been retracted due to methodological errors and flaws (Chen et al., 2021). Beyond the above, administrative errors, such as wrong publishing issues, may also lead to retractions (Bar-Ilan & Halevi, 2018). It is noted that the reasons for retractions may extend beyond the explicit retraction statements and involve multifaceted underlying factors, as reflected by some recent controversial retractions (e.g., Abramo et al., 2023). While this issue lies outside the scope of this study, future studies may continue to investigate possibly implicated factors other than the stated retraction reasons.

To prevent publishing problematic publications before retractions in the first place, the scientific community should increase the effectiveness of peer review to detect non-administrative errors (Azoulay et al., 2017; Bar-Ilan & Halevi, 2018; Horbach & Halffman, 2019). Peer review is a pre-publication quality control system run by professional same-field experts to distinguish the accurate, significant, and novel research works for publishing and provide constructive feedback for the authors (Adewoyin & Vassileva, 2014; Gerwing et al., 2020; Ortega, 2017). The peer review process should recommend rejection if a paper contains significant errors or suspicious misconduct (Bornmann et al., 2008). Some studies endorsed the peer review system's effectiveness in rejecting flawed manuscripts and improving research integrity and quality (Casnici et al., 2017; Siler et al., 2015). The surging preprints, which are usually not peer-reviewed before they are shared on preprint servers, also appear to have a higher retraction rate than peer-reviewed journal papers in the COVID-19 research (Kharasch et al., 2021).

In contrast with its broad impact on the scientific enterprise, peer review has also been criticized for its low effectiveness in identifying suspicious research. Some researchers argue that peer review fails to catch all instances of significant errors and misconduct that lead to retractions in submitted manuscripts (Bornmann et al., 2008; Fox, 1994; Horbach & Halffman, 2019; Resnik & Elmore, 2016; Schroter et al., 2008). Although the peer review process is widely regarded as the frontline in identifying flawed research, researchers doubt if it can do so (Mulligan et al., 2013). Evidence shows that peer review rarely changes the core content of a manuscript but tends to focus on narrower, more technical details (Siler et al., 2015). Unprofessional and biased reviewers may write inadequate, inconsistent, and biased reviews that omit the problems of papers (Gerwing et al., 2020; Lee et al., 2013; Rennie, 2016; Resnik et al., 2008). Gerwing et al. (2020) found that 12% of 1491 sets of reviewer comments from "Ecology and Evolution" and "Behavioral Medicine" included at least one unprofessional comment towards the author or their work, and 41% contained incomplete, inaccurate, or unsubstantiated critiques. Only 2% of the comments included an accusation of questionable research practices.

However, there is a knowledge gap about whether the peer-review process should be held accountable for failing to detect and report problematic areas in submitted manuscripts that lead to retractions. We aim to identify the effectiveness of the peer review process in identifying suspicious areas in submissions that later lead to retraction. We cannot rule out the possibility that the editors ignored these risk factors even though peer reviewers have suggested the issues and given unfavorable recommendations. Understanding the factors in the peer-review process that are related to the successful identification of malicious components in the submissions is critical for evaluating the effectiveness of the peer-review process in preventing retractions.

This study utilizes a unique combination of peer-review and paper retraction data to investigate the effectiveness of the peer-review process in identifying issues that lead to retractions. Generally, journals do not provide detailed information about their peer-review process, and reviewers' identities and comments are typically kept confidential, despite the recent trend of open peer review (Wolfram et al., 2020). This has limited the ability to conduct large-scale investigations into the effectiveness of peer review. This study aims to fill this gap by examining the relationships between reviewer characteristics and the effectiveness of peer review in identifying issues leading to retractions. The findings of this study will provide insight into the role of peer review in preventing scientific retractions and offer suggestions for journals on how to improve the effectiveness of the peer-review process.

2. Materials and methods

2.1. Data sources

This study relies on two data sources: the retracted publication list by RW (The Center For Scientific Integrity, 2018) and peer review comments from Publons by Clarivate Analytics (2012). RW keeps track of retracted scientific publications by documenting

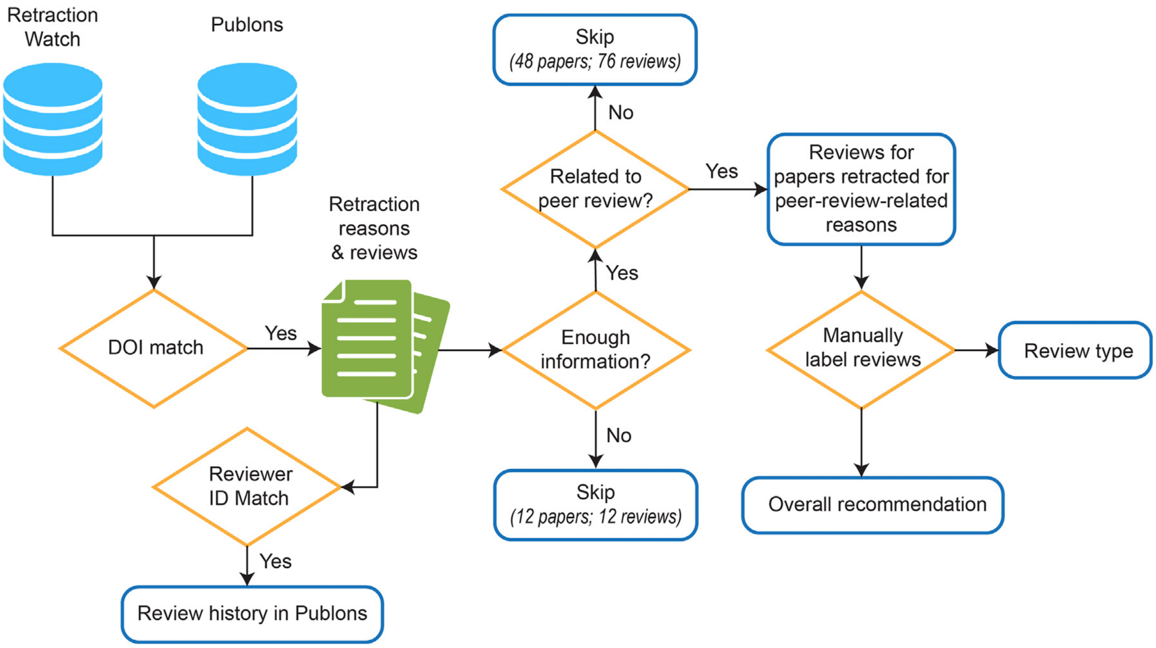

Fig. 1. Data preparation pipeline.

the critical metadata information of retracted publications, including author names, publication titles, source titles, digital object identifiers (DOI), disciplinary fields, and specific reasons for retraction. Publons documents scientists' invisible peer review contributions by tracking their peer review records. Each record in Publons is a peer review comment for a manuscript, including the unique identification for the reviewed manuscript, the source title, the unique identifier for the reviewer, the review comment text, and the DOI of the reviewed manuscript if published. The peer review data underwent anonymization and deidentification process by Publons before they were used in this study.

We obtained peer review comments for retracted publications by matching the DOIs from the RW database and Publons. Peer review comments associated with those DOIs of retracted publications were then extracted for further analysis. We found 348 first-round peer reviews for 211 retracted papers. Among the reviews, 12 reviews (associated with 12 retracted papers) have insufficient information for analysis (e.g., "no comments," "see my comments attached," "see the report," and "see above") and were thus excluded for subsequent analysis. This leaves 206 retracted papers and 336 reviews for further analysis (See Fig. 1).

2.2. Data sources

The RW database records the reason(s) for each retraction. Each retracted paper is associated with one or more reasons from the 102 reasons listed by RW (see Appendix A, Table A.1). We excluded 76 reviews concerning 48 retracted papers that were retracted due to administrative reasons that are unrelated to the peer review process, such as "Copyright Claims," "Objections by Third Party," and "Error by Journal/Publisher." This leaves 32 retraction reasons, covering 160 retracted papers and 260 peer-review comments in our dataset. Because submitting peer review records to Publons is a voluntary practice for reviewers, Publons does not necessarily have all the reviews conducted for each manuscript, which we admit is a limitation of our study. As shown in Fig. 2, the dataset only has one review record for about 102 retracted papers and two review records for 38 retracted papers.

To increase the interpretability of the results, we aggregated the remaining 32 reasons for retraction provided by RW into seven categories: *plagiarism, data, methods & analysis, result, reference, author*, and *other* (COPE Council, 2019; Marcovitch, 2007; Nair et al., 2020). The number of retracted papers and the corresponding review comments by each retraction reason category are shown in Table 1 (see Appendix A, Table A.1 for the breakdown by the more specific reasons as documented by RW). It should be noted that some papers could be retracted for multiple reasons and could also have multiple peer-review comments.

2.3. Coding and labeling review comments

To understand the gatekeeping role of the peer review process in identifying issues leading to retractions, we read and manually coded the peer review comments for each retracted paper. We first coded each peer review by the type of recommendation it implied in the comments. Two independent coders coded each peer review comment into one of the four recommendation categories using the following criteria:

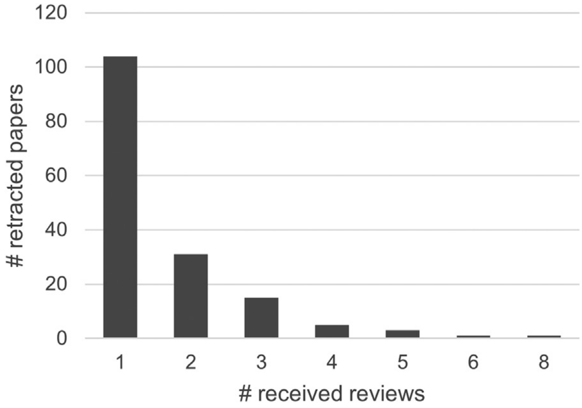

Fig. 2. The distribution of retracted papers by the number of available reviews.

Table 1
Aggregated retraction causes by the number of retracted papers and reviews.

	Paper		Corresponding review	
Aggregated retraction causes	Number	Percentage (%)	Number	Percentage (%)
Plagiarism	58	36.25	103	39.62
Data	69	43.13	95	36.54
Method/Analysis	33	20.63	55	21.15
Result	92	57.50	135	51.92
Reference	7	4.38	8	3.08
Author	3	1.88	3	1.15
Other	2	1.25	2	0.77
Total	160	100.00	260	100.00

Table 2
Number of reviews (and associated papers) by coded recommendation.

	Accept	Minor revision	Major revision	Reject	Total
Number of reviews	55	73	111	21	260
Percentage of reviews	21.15%	28.08%	42.69%	8.08%	100.00%
Number of related papers	46	58	89	19	160

- Reject: The comment implies a rejection of the manuscript to be published in the journal by indicating that the reviewer did not see a path to publication in this journal.
- Major revision: The comment implies the possibility that the reviewer could recommend publication after significant changes regarding various aspects of the manuscript.
- Minor revision: The comment implies the reviewer is likely to recommend publication after minor changes (or conditional acceptance), such as changes regarding language edits, formating, and other trivial aspects.
- Accept: The comment advocates accepting the manuscript of publication without changes.

The two coders reached an agreement on about 88.46% of all the peer review comments, with a Cohen's Kappa value of 0.83 (See Appendix A, Table A.2 for the inter-coder agreement regarding the overall recommendation). This is considered a "good" coding agreement according to Krippendorff's Alpha interpretation scales of Kappa (Krippendorff, 2004). A third coder labeled the disagreed reviews between the two coders. Table 2 lists the number of reviews (and associated papers) by coded recommendation after the third coder coded the disagreed reviews.

We also labeled the peer review comments concerning the reasons related to the retraction. Specifically, the two coders first read the peer review comment and the retraction reasons for each retracted paper. They then coded each review comment concerning each retraction reason and "problem detection," "praise," and "solution suggestion" labels (Cho, 2008), based on the following criteria:

- Problem Detection: Concerning each retraction reason, we consider "Problem Detection" to be present (Problem Detection=(1) in a peer review comment if the comment mentions or expresses concerns on the issues leading to the retraction. The reviewer may state the issues in a neutral (if simply stating the problem) or negative (if criticizing the problem) tone. For example, a reviewer raised concerns over a paper that "*some bands in different blots look very similar*" regarding images in the manuscript. According to its retraction notice, the corresponding paper was later retracted due to image manipulation. The review comment will have the "Problem Detection" labeled true in this case.

Table 3
Number of reviews (and associated papers) by type of comments.

	Problem Detection		Praise		Solution Suggestion	
	Yes=1	No=0	Yes=1	No=0	Yes=1	No=0
Number of Papers	59	132	15	150	28	147
Number of Reviews	136	365	33	468	68	433

Note. Because each paper could be retracted for multiple reasons and each review was coded against each reason separately, the number of reviews and papers under each category is the number of times each reason falls into individual categories.

- Praise: Concerning each retraction reason, we consider "Praise" to be present (Praise=1) in a peer review comment if the comment fails to point out the retraction-related issues and uses words expressing gratitude, positivity, admiration, approval, or respect for the very area that later on led to the retraction. For example, a review comment praised the study as "a well-organized, well-studied experimental study." The paper was later retracted due to methods-related issues. The review comment will have the "Praise" label set to value=1.
- Solution Suggestion: Concerning each retraction reason, we consider "Solution Suggestion" to be present (Solution Suggestion=1) in a peer review comment if it points out the retraction-related issues and provides suggestions to address them. For example, for a paper retracted for method-related reasons, its peer review suggested that "given the issues mentioned above, the manuscript would benefit from a reconstruction of the control group by considering…" The review comment will have the "Solution Suggestion" label set to value=1.

Please note that we annotated the review comments by referring to the particular reasons mentioned in the retraction notice of each paper. In case the retraction notice was unavailable, we referred to the brief retraction reasons recorded by RW. Sometimes there could be more than one reason for retraction. Therefore, we coded the review comment based on each retraction reason separately. Two independent coders coded the 260 peer review comments for the type of comment regarding retraction reasons. The Cohen's Kappa values for "Problem Detection," "Praise," and "Solution Suggestion" were 0.83, 0.91, and 0.87, respectively. Similarly, a third coder labeled comments disagreed with by the two coders. The final number of peer reviews and papers in "Problem Detection," "Praise," and "Solution Suggestion is shown in Table 3.

2.4. Reviewer characteristics

To understand the relationship between peer review and retracted science, we examined the relationship between reviewer characteristics and the likelihood of identifying (or praising and suggesting solutions to) issues leading to the retraction. For the 198 individual reviewers of these 260 reviews for 160 retracted papers, we found their review profiles and histories in Publons (anonymized). Using the accessible data, we considered the following reviewer characteristics to approximate their professional expertise, review quality, and review experience while analyzing the linkage between peer review comments and reasons for retraction. Appendix A, Table A.3 shows the Pearson correlation matrix of these variables in our dataset. As no pairs of variables have high correlations, it is less likely to lead to serious multicollinearity in the regression model.

- Topic Similarity: The topic similarity is the topical distance between a review comment and all review comments performed by the same reviewer calculated using the word2vec method (Mikolov et al., 2013). This measures the average similarity between the peer review comment for the retracted paper and all other reviews by the same reviewer, approximating the closeness between the topic of the reviewed manuscript and the areas of expertise of the reviewer. To avoid the similarity being affected by common words such as "figure" in the review comments, we excluded words that appeared in more than 50% of the comments from the calculation to reduce their impact on the results. Furthermore, we calculated the inverse document frequency (IDF) for each lemmatized word in the review comments. The IDF reflects the rarity of a word in the dataset, with more common words receiving lower values. These IDFs were used as weights for computing the document vectors in our topic similarity comparison.
- Average comment length: This measures the average number of words in each peer review comment by the reviewer. The length of a review comment is used as a proxy of the review's quality, thoroughness, and helpfulness (Thelwall, 2022; Zong et al., 2021).
- Acceptance rate: This reflects the percentage of manuscripts published out of the total manuscripts reviewed by a reviewer. A high acceptance rate for a reviewer may indicate that the reviewer is less efficient in gatekeeping the manuscript quality strictly and writing high-quality peer reviews (Kurihara & Colletti, 2013; Ortega, 2017).
- Seniority: This measures the number of years between a reviewer's first and last peer reviews. This measure shows the length of a reviewer's review history and also indicates the reviewer's overall peer review experience from one perspective.
- Number of reviews*:* This measures the annual number of peer reviews performed by a reviewer. This variable quantifies the commitment of a reviewer in the peer-review process and is another indicator of peer-review experience measurement.

2.5. Regression analysis

This study used logistic regression to investigate how various reviewer characteristics contribute to the probability of reviewers identifying issues leading to retractions. The outcome variables include whether the peer review comment of a manuscript detected

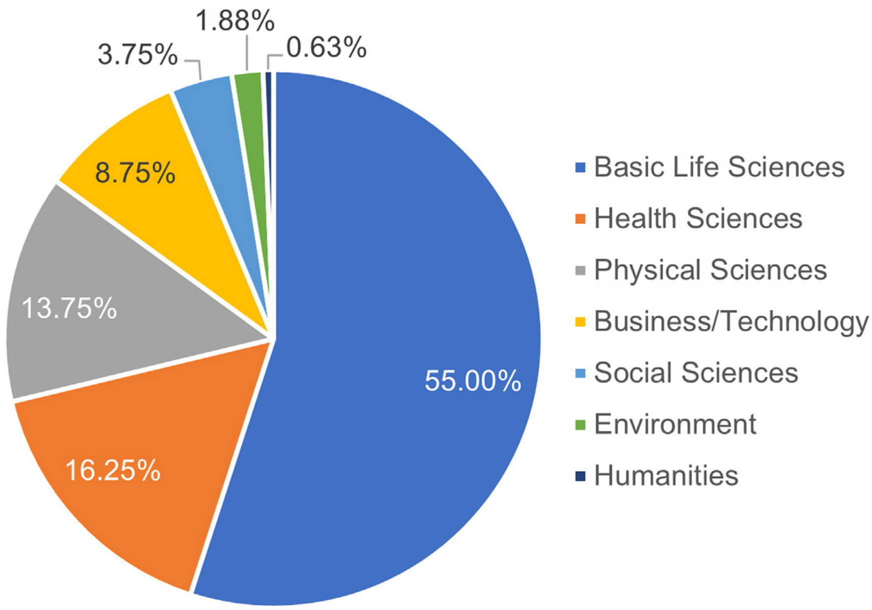

Fig. 3. Discipline distribution of retracted papers.

the issues later leading to the retraction (Problem Detection), whether the peer review comment positively praised the area that later on became reasons for retraction (Praise), and whether the peer review comment provides suggestions to improve the problems leading to the retraction (Solution Suggestion). The independent variables used are the reviewer characteristics mentioned above. Given that all three outcome variables are dichotomous, we chose binary logistic regression for the analysis. We controlled for the disciplines of papers in regression analysis to ensure that the observed relationship is not specific to one particular field (Zhang et al., 2022). We followed the discipline classification by RW. As shown in Fig. 3, the disciplinary distribution in our dataset aligns with He's (2013) finding that retractions are more common in the biomedicine and life science fields. We binned disciplines other than Basic Life Sciences and Health Sciences in the regression to increase statistical power.

The regression specification is as follows.

$$logit(P) = \beta_0 + \beta_1 Seniority + \beta_2 AvgLength + \beta_3 \#reviews + \beta_4 AccRate + \beta_5 TopicSim + \sum \alpha Discipline + \epsilon \quad (1)$$

where P is the probability of detecting problems, praising, or suggesting solutions, and ϵ is the residual. We clustered the standard deviation at the paper level. We rescaled $AccRate$ and $TopicSim$ at the level of 10% in this regression to better display their coefficients in the results.

3. Results

3.1. Reviewer recommendations for retracted papers

Our coding results suggest that most reviewers failed to reject the later-retracted papers. Out of the 260 reviews associated with 160 retracted papers, 128 (49.2%) were perceived to recommend "Acceptance" (55) or "Minor revision" (73) for the manuscripts. 111 (42.7%) of the reviews recommended "Major revision" for their reviewed manuscripts. Only 21 (8.1%) were perceived to recommend "Rejection" for the manuscript. Each paper may have multiple reviews and thus could have different recommendations. In our data, 13 papers (8.1%) received consensus for a "Rejection" from their reviewers, 20 (12.5%) an "Acceptance," 30 (18.8%) a "Minor revision," and 52 (32.5%) a "Major revision." The remaining 45 (28.1%) papers received mixed recommendations from their reviewers (see Table 4).

3.2. Problem detection for retraction reasons

To understand the role of the peer review process in retracted science, we analyzed the effectiveness of peer reviews in identifying issues that were later reasons for retraction – known as "problem detection" in this study. Among the 260 reviews, 192 (73.8%) failed to detect issues related to the retraction of the papers, and 68 (26.2%) detected at least one problem related to the retraction of the papers. Table 5 shows the number and percentage of retracted papers whose retraction reasons were detected in their peer review comments, as well as the number and percentage of reviews that detected the issues related to retraction reasons. Overall, about 24.6% of the reviews identified the issues that are related to reasons for retraction more or less, and 30.9% of papers had at least one review identifying issues related to its retraction. None of the reviews detected the problem for papers later retracted for

Table 4
Number of papers and reviews by recommendation type.

Coded recommendation	Reviews		Papers	
	Number	Percentage	Number	Percentage
Accept	55	21.2%	20	12.5%
Minor revision	73	28.1%	30	18.8%
Major revision	111	42.7%	52	32.5%
Reject	21	8.1%	13	8.1%
Mixed			45	28.1%

Table 5
Problem detection reviews (and associated papers) by reasons for retraction. P= Paper; R=Reviews.

Reasons for retraction	Papers (n=160)		Reviews (n=260)	
	Number	Percentage	Number	Percentage
Plagiarism (P=58; R=103)	10	19.23%	11	11.46%
Data (P=69; R=95)	28	40.58%	32	33.68%
Method/Analysis (P=33; R=55)	14	42.42%	17	30.91%
Result (P=92; R=135)	40	43.48%	48	35.56%
Reference (P=7; R=8)	1	16.67%	1	12.50%
Author (P=3; R=3)	0	0.00%	0	0.00%
Total (P=160; R=260)	59	30.89%	68	24.55%

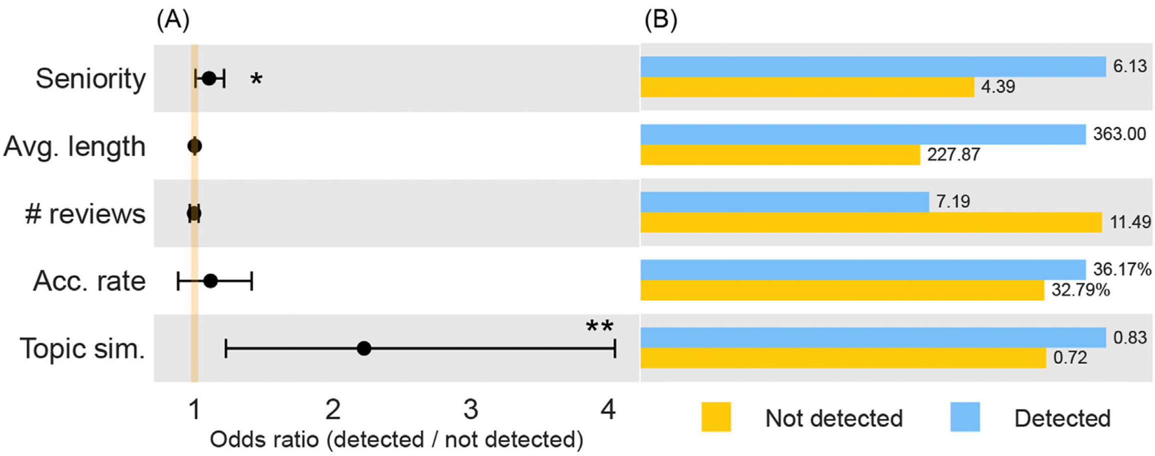

Fig. 4. Logistic regression results for problem detection of retraction reasons. (A) Odds ratio values of reviewer-level factors. (B) Mean values of reviewer-level factors. *** p < 0.001, ** P < 0.01, * P < 0.05.

author-related reasons. Among reasons for retractions, "result" related issues were detected by 35.6% of the peer reviews successfully, followed by "data" (33.7%) and "method/analysis" (30.9%).

We further performed logistic regression analysis to investigate whether the reviewer characteristics are related to the problem-detection chances of peer review comments. Our results show that a reviewer's seniority, average comment length, and topic similarity are significant predictors of problem detection (see Fig. 4). Specifically, Reviewers with higher seniorities are more likely to detect problems that later lead to the retraction of the paper (OR = 1.105, 95%CI [1.006, 1.213], p = 0.037). The average seniority of reviewers is 1.74 years longer in the problem-detected reviewer group (6.13 years) than in the not detected group (4.39 years). The topic similarity (between the current review and all reviews by a reviewer) also contributes significantly to the chance of problem detection: The higher the topic similarity, the more likely a review can identify problems leading to the retraction (OR = 2.227, 95%CI [1.226, 4.043], p = 0.009). The results indicate that reviewers' seniority and expertise in the topic may be associated with the likelihood of detecting potential issues that lead to retraction.

When retraction reasons are aggregated into categories (see Table 1), the reviewer characteristics contributing significantly to the chance of problem detection vary by category (see Table 6). Across reasons for retraction categories, topic similarity contributes significantly to the possibility of detecting issues leading to retractions. For reviews of papers retracted due to data-related and methods and analysis-related issues, the higher the topic similarity (data: OR = 2.264, 95%CI [1.092, 4.694], p = 0.028; method/analysis: OR = 10.284, 95%CI [1.064, 99.428], p = 0.044), the more likely the peer review comment can identify issues leading to retractions.

Table 6
Logistic regression results of problem detection by types of issues.

	Odds ratio	Std. Err	p-value	95% CI Lower	95% CI Upper
Data (n=111)					
Seniority	1.114	0.096	0.210	0.941	1.320
Ave. length	1.001	0.001	0.642	0.998	1.003
# reviews	0.997	0.025	0.901	0.948	1.048
Acc. rate	1.152	0.229	0.477	0.780	1.699
Topic sim.	2.264	0.842	0.028	1.092	4.694
Method/Analysis (n=76)					
Seniority	1.136	0.098	0.139	0.960	1.345
Ave. length	1.002	0.002	0.267	0.998	1.006
# reviews	0.994	0.034	0.855	0.929	1.063
Acc. rate	0.748	0.287	0.448	0.352	1.586
Topic sim.	10.284	11.904	0.044	1.064	99.428
Plagiarism (n=114)					
Seniority	1.044	0.078	0.561	0.902	1.210
Ave. length	0.998	0.002	0.468	0.994	1.003
# reviews	0.937	0.083	0.464	0.788	1.115
Acc. rate	1.811	0.436	0.014	1.130	2.903
Topic sim.	1.460	0.728	0.448	0.549	3.881
Result (n=175)					
Seniority	1.164	0.078	0.022	1.022	1.327
Ave. length	1.000	0.001	0.801	0.998	1.002
# reviews	1.018	0.018	0.306	0.984	1.053
Acc. rate	1.073	0.147	0.604	0.821	1.403
Topic sim.	2.240	0.719	0.012	1.194	4.203

Table 7
Praise and solution suggestion reviews (and associated papers) by reasons for retraction. P= Paper; R=Reviews.

	Praise				Solution Suggestion			
	Paper		Review		Paper		Review	
	Num.	Percent.	Num.	Percent.	Num.	Percent.	Num.	Percent.
Author (P=3; R=3)	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Data (P=69; R=95)	8	11.59%	8	8.42%	13	18.84%	14	14.74%
Method/Analysis (P=33; R=55)	5	15.15%	5	9.09%	10	30.30%	10	18.18%
Plagiarism(P=58; R=103)	5	8.62%	5	4.85%	6	10.34%	6	5.83%
Reference (P=7; R=8)	0	0.00%	0	0.00%	1	14.29%	1	12.50%
Result (P=92; R=135)	9	9.78%	9	6.67%	19	20.65%	19	14.07%
Total (P=160; R=260)	15	9.38%	27	10.38%	28	17.50%	29	11.15%

For reviews of papers retracted due to plagiarism, the acceptance rate is a significant predictor of the chance of detecting plagiarism issues. For reviews of papers retracted due to results-related issues, higher seniority and topic similarity indicates a higher probability of detecting results-related issues (seniority: OR = 1.164, 95%CI [1.022, 1.327], p = 0.022; topic similarity: OR = 2.240, 95%CI [1.194, 4.203], p = 0.012).

3.3. Praise and solution suggestion for retraction reasons

In our data, some review comments praised the areas later on that led to retractions rather than raising concerns, albeit in a small proportion (see Table 7). Among the 260 peer review comments for the 160 retracted papers, 27 (10.4%) reviews (for 15 papers) mentioned the retraction issue with a praising tone. Logistic regression analysis shows that the chance of praising issues leading to retractions is not related to any of the reviewer-level factors (see Fig. 5). Given the limited number of praising peer reviews by retraction reasons, no regression analysis was performed by retraction reasons separately.

We also examined whether peer review comments provide solution suggestions to issues leading to retractions. As shown in Table 7, 29 (11.2%) out of the 260 peer review comments provided suggestions for solving issues leading to retractions, which accounts for 17.50% (28) of the total papers in the sample.

Our results show that average comment length and topic similarity contribute significantly to the chance of providing suggestions to issues leading to retractions (see Fig. 5). Specifically, the longer the reviews written by a reviewer (OR = 1.002, 95%CI [1.000, 1.003], p = 0.020), the higher the topic similarity (OR = 3.395, 95%CI [1.478, 7.800], p = 0.004), the more likely a peer review comment can provide solution suggestions to issues leading to retractions. Given the limited number of solution suggestion reviews by retraction reasons, no regression analysis was performed by retraction reasons separately.

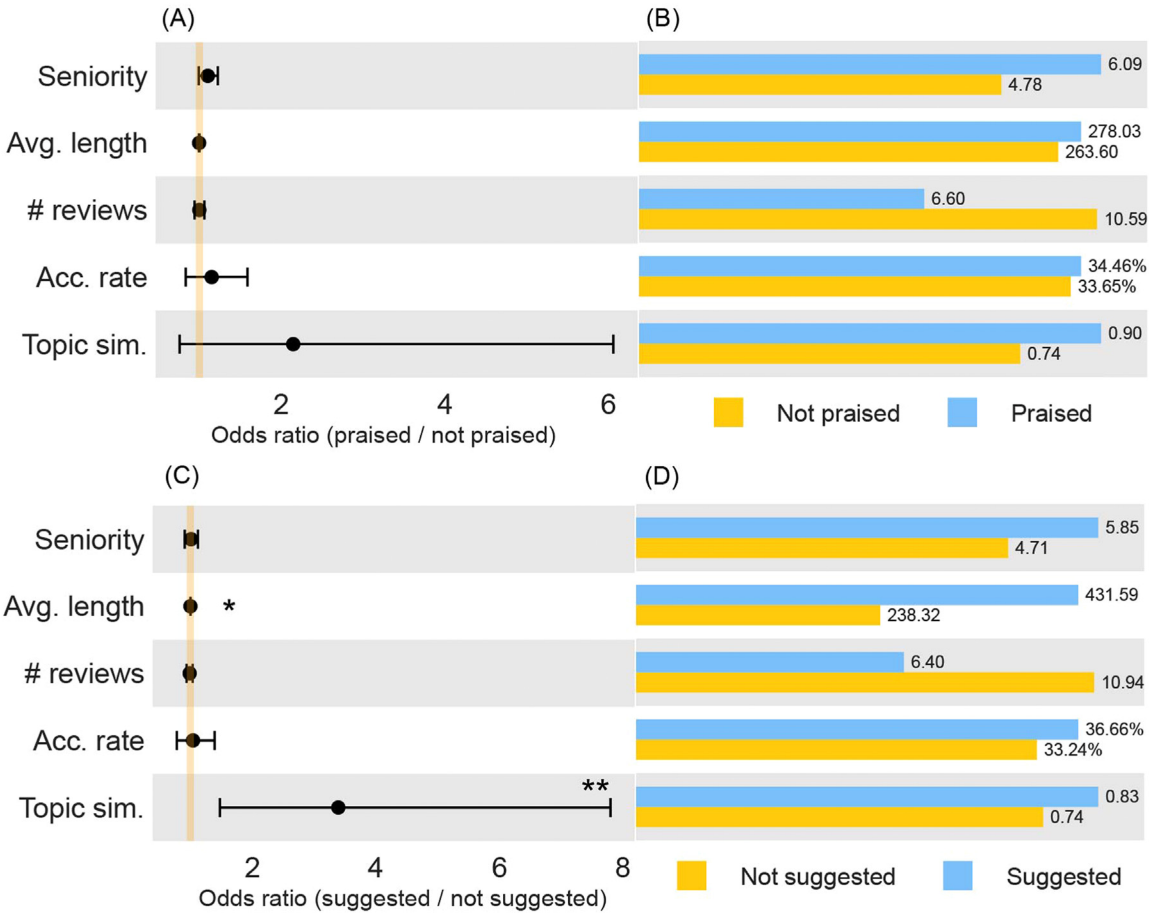

Fig. 5. Logistic regression results for Praise and solution suggestion. (A) Odds ratio values of reviewer-level factors for praise comments. (B) Mean values of reviewer-level factors for praise comments. (C) Odds ratio values of reviewer-level factors for solution suggestion comments. (D) Mean values of reviewer-level factors for solution suggestion comments. *** $p < 0.001$, ** $P < 0.01$, * $P < 0.05$.

4. Discussion and conclusions

This study evaluated the effectiveness of peer-review comments in preventing retractions by analyzing a sample of peer-review comments and comparing them to the reasons for retraction. By manually coding the peer review comments, the study found that only 42.7% of the peer reviews suggested “major revision,” and 8.1% suggested “rejection” for papers that were later retracted. Instead, 49.2% of the peer reviews suggested “minor revision” or “acceptance” without raising concerns about the issues that led to the retraction of the manuscript. The rate at which peer reviews identified issues that led to retractions was not high, with only 24.55% of the 260 peer review comments pointing out issues that later led to retractions, covering 30.9% of the papers. Additionally, 10.40% of the peer reviews praised areas that were later cited as reasons for retractions. 11.15% of peer review comments suggested solutions related to the issues leading to retraction. These findings suggest that while some peer reviews did raise issues and suggest solutions that were later cited as reasons for retraction, the papers still slipped through the editorial peer-review system.

We also found that the effectiveness of the peer-review process in identifying problematic areas varies depending on the type of issue leading to retraction. The study found that issues leading to retractions due to data, methods/analysis, and results were detected by peer reviews at a higher rate than issues leading to plagiarism, author, and reference-related retractions (see Appendix A, Table A.4). As plagiarism is a common reason for retractions (Campos-Varela & Ruano-Raviña, 2019; Steen et al., 2013), attention is needed to improve peer review in detecting plagiarism. The lower detection rate of retractions due to the author or reference-related reasons is possibly due to the double-blind review system, which masks the authors’ real information (Lee et al., 2013; Mulligan et al., 2013), and a lack of focus on references during peer review (Bornmann et al., 2008). However, this finding needs further testing with larger sample sizes.

Matching peer reviewers with submitted manuscripts of similar research topics is crucial in detecting potential issues that may lead to retraction. Our study found that except for the reason of plagiarism, the higher the topic similarity between the current review and all reviews by the same reviewer, the more likely the current peer-review comment will detect potential retraction issues.

Expertise matching is considered essential for fair and professional evaluations of manuscripts and grant submissions (Kelly et al., 2014). However, given the high volumes of submissions in many fields, the limited number of reviewers (Kovanis et al., 2016), and potential mismatching by reviewing systems (Anjum et al., 2019), it is not uncommon for reviewers to evaluate manuscripts that do not fit their areas of expertise exactly. However, our findings suggest that expertise matching should not be sacrificed to meet the needs of increasing peer review but should be emphasized to maintain the research integrity of published work (Resnik & Elmore, 2016).

The study also found that reviewers' seniority is a significant factor in identifying retraction-related issues during the peer-review stage. Senior reviewers are more likely to identify issues in submissions that later lead to retractions, which might be related to their extended reviewing history, confidence, and accumulated research experience (Warne, 2016). Such reviewing experience is valuable and hard to gain through short-term reviewer training (Schroter et al., 2008). Our results suggest that journals and conferences should pay attention to the presence of senior and experienced reviewers when assigning them, especially when multiple reviewers are assigned to a single manuscript. Additionally, the significant association between review length and the possibility of suggesting solutions is not beyond expectation. As longer reviews are more likely to be associated with constructive feedback (Thelwall, 2022; Zong et al., 2021), we infer that reviewers writing longer reviews are more likely to be willing to spend more effort in providing feedback.

Preventing retractions requires intricate and multidimensional efforts involving authors, peer academic institutions, funders, journals, publishers, peer reviewers, and others in the scientific community. The peer review process does seem to detect issues that later lead to retractions and further suggest solutions or recommend rejections to the manuscript. However, we found that peer reviews may also complement and endorse the areas later identified as associated with retractions. Therefore, our findings suggest that the peer review process only detects problematic issues at certain rates, which is subject to peer reviewers' seniority, efforts spent, and the match of expertise. We suggest that editors, or those in charge of consolidating peer review comments for decision-making, should pay close attention to peer review comments and perform additional inspections to trace clues of potential issues from peer review comments. Various parties involved in the peer review and publication process should be vigilant in detecting and addressing potential issues that may lead to retractions. By taking a multifaceted approach to preventing retractions, the scientific community may maximize the effectiveness of peer review to increase the integrity and credibility of published research.

Our findings lead to informed recommendations for journal editors and others about the importance of matching the right peer reviewers with submissions based on areas of expertise. The massive volume of manuscript submissions may drive editors to choose reviewers with limited grounds for assessing their areas of expertise (Kelly et al., 2014) and fail to develop a systematic mechanism to monitor their performance and competence as reviewers (Fox, 1994). However, the integrity of science and the role of peer review should not be sacrificed to accommodate the increasing volume of submissions. Manuscript submission and handling system should be better equipped for matching peer reviewers with submitted manuscripts, in addition to current technologies implemented for plagiarism detection (Ercegovac & Richardson, 2004), image manipulation (Beck, 2021), and reproducibility assessment (Munafò et al., 2017). Given the importance of the thoroughness of comments shown in our study, institutions and funders might also establish procedures and policies to provide the necessary training for reviewers about their responsibilities and how to check scientific misconduct and errors.

This study is not immune to some limitations, primarily due to the data and methods used. Our analyses focused on retractions that could be potentially related to the peer review process and ignored those due to reasons such as authorship disputes, ethical issues, IRB approvals, and peer review fraud. These serious causes of retractions would be hard to prevent during the peer review process but are also critical reasons for many retractions. Additionally, our analysis was restricted to those reviewers who opted to share their reviews on Publons. Therefore, we could not analyze those peer reviews outside of Publons for their relationships with retractions. The anonymity of the reviewers in our dataset prevented us from considering additional characteristics (such as those related to their research practices) in our analysis. However, the current integrated Web of Science researcher profile presents a promising opportunity to link a researcher's Web of Science-indexed publications and their peer review history. We will continue investigating this possibility to improve our understanding of the peer review process.

Furthermore, we manually coded the peer review comments, which could be subject to some common issues with hand coding. On top of that, retraction notices could sometimes be brief without detailing the exact problems causing the retraction. For example, one retraction notice states that the paper was retracted "due to unreliable results" without specifying which result is problematic and why. When we coded the peer review comments, we would label any place in the peer review comments that mentioned the potential issues of unreliable results with "Problem Detection." Therefore, for retractions without detailed notice, there is the possibility that our coding of retraction causes might slightly misalign with the actual causes. Finally, our analysis focused on retractions that happened in recent years due to data availability, which may limit the generalizability of our results when concerning earlier retractions and peer reviews.

Acknowledgments

We gratefully acknowledge the teams of Retraction Watch for providing the retraction records data, and Clarivate Analytics for the anonymized peer review data.

Appendix A

Table A.1
Re-categorization of retraction reasons, and number of papers by retraction reasons (RW).

Class	Reasons by RW	Number of Papers
Author	Author Unresponsive	3
	Breach of Policy by Author	0
	Complaints about Author	1
	Concerns/Issues About Authorship	3
	False Affiliation	0
	False/Forged Authorship	1
	Lack of Approval from Author	0
	Miscommunication by Author	0
	Misconduct - Official Investigation/Finding	6
	Misconduct by Author	10
	Objections by Author(s)	1
Conflict of Interest	Conflict of Interest	0
Data	Concerns/Issues About Data	45
	Contamination of Cell Lines/Tissues	0
	Contamination of Materials (General)	2
	Contamination of Reagents	0
	Duplication of Data	6
	Error in Cell Lines/Tissues	2
	Error in Data	30
	Error in Materials (General)	3
	Falsification/Fabrication of Data	3
	Original data not Provided	5
	Plagiarism of Data	0
	Sabotage of Materials	0
	Unreliable Data	15
Ethical concerns	Breach of Policy by Third Party	0
	Concerns/Issues about Third-Party Involvement	4
	Ethical Violations by Author	3
	Ethical Violations by Third Party	0
	Informed/Patient Consent	4
	Lack of Approval from Company/Institution	0
	Lack of Approval from Third Party	2
	Lack of IRB/IACUC Approval	5
Review	Fake Peer Review	0
Journal/Publisher/third-party issue	Duplicate Publication through Error by Journal/Publisher	0
	Error by Journal/Publisher	0
	Error by Third Party	0
	Miscommunication by Company/Institution	0
	Miscommunication by Journal/Publisher	0
	Miscommunication by Third Party	0
	Misconduct by Third Party	0
Methods/Analysis	Bias Issues or Lack of Balance	3
	Error in Analyses	43
	Error in Methods	31
Other	Civil Proceedings	0
	Complaints about Company/Institution	0
	Complaints about Third Party	0
	Copyright Claims	1
	Criminal Proceedings	0
	Doing the Right Thing	1
	Investigation by Company/Institution	10
	Investigation by Journal/Publisher	23
	Investigation by ORI	0
	Investigation by Third Party	15
	Legal Reasons/Legal Threats	0
	Misconduct by Company/Institution	0
	Not Presented at Conference	0
	Objections by Company/Institution	0
	Objections by Third Party	12
	Publishing Ban	0
	Rogue Editor	5
	Transfer of Copyright/Ownership	2

(*continued on next page*)

Table A.1 (*continued*)

Class	Reasons by RW	Number of Papers
Plagiarism	Duplication of Article	29
	Duplication of Text	2
	Euphemisms for Duplication	2
	Euphemisms for Misconduct	13
	Euphemisms for Plagiarism	0
	Hoax Paper	0
	Paper Mill	19
	Plagiarism of Article	34
	Plagiarism of Text	0
	Salami Slicing	0
	Taken from Dissertation/Thesis	4
Reference	Cites Retracted Work	0
	Concerns/Issues about Referencing/Attributions	5
Results	Concerns/Issues About Image	15
	Concerns/Issues About Results	10
	Duplication of Image	23
	Error in Image	10
	Error in Results and/or Conclusions	40
	Error in Text	7
	Falsification/Fabrication of Image	4
	Falsification/Fabrication of Results	0
	Manipulation of Images	15
	Manipulation of Results	1
	Plagiarism of Image	3
	Results Not Reproducible	11
	Unreliable Image	0
	Unreliable Results	40

Table A.2
Overall recommendation of manuscript – inter-coder agreement.

	Accept	Minor revision	Major revision	Reject	Total
Accept	51	3	2	0	56
Minor revision	2	70	6	1	79
Major revision	0	4	94	8	106
Reject	0	0	4	15	19
Total	53	77	106	24	260
Agreement	51	70	94	15	230

Table A.3
Pearson correlation matrix of regression variables.

	Seniority	Avg. length	# reviews	Acc. rate	Topic sim.
Seniority	1				
Avg. length	0.23	1			
# reviews	-0.16	-0.17	1		
Acc. rate	-0.14	0.21	-0.22	1	
Topic sim.	0.24	0.38	-0.33	-0.04	1

Table A.4
Paper and review distribution by RW reasons for retraction.

Group	Reasons	# Reviews	# Papers
Author (P=3; R=3)	Ethical Violations by Author	3	3
Data (P=69; R=95)	Concerns/Issues About Data	49	35
	Contamination of Materials (General)	2	2
	Duplication of Data	6	3
	Error in Cell Lines/Tissues	2	2
	Error in Data	30	21
	Error in Materials (General)	3	3
	Falsification/Fabrication of Data	3	3
	Original data not Provided	5	2
	Unreliable Data	15	10
Methods & Analysis (P=33; R=55)	Bias Issues or Lack of Balance	3	1
	Error in Analyses	43	24
	Error in Methods	31	18
Plagiarism (P=58; R=103)	Duplication of Article	32	18
	Duplication of Text	2	2
	Euphemisms for Duplication	2	2
	Euphemisms for Plagiarism	15	6
	Paper Mill	19	12
	Plagiarism of Article	40	19
	Plagiarism of Image	3	1
	Taken from Dissertation/Thesis	4	2
Reference (P=7; R=8)	Concerns/Issues about Referencing/Attributions	8	6
Results (P=92; R=135)	Concerns/Issues About Image	15	8
	Concerns/Issues About Results	10	9
	Duplication of Image	27	19
	Error in Image	10	7
	Error in Results and/or Conclusions	40	26
	Error in Text	7	6
	Falsification/Fabrication of Image	4	3
	Manipulation of Images	15	7
	Results Not Reproducible	12	9
	Unreliable Results	41	27

References

Abramo, G., Aguillo, I. F., Aksnes, D. W., Boyack, K., Burrell, Q. L., Campanario, J. M., et al., (2023). Retraction of predatory publishing in scopus: Evidence on cross-country differences lacks justification. *Scientometrics, 128*(2), 1459–1461. 10.1007/s11192-022-04565-6.

Adewoyin, O., & Vassileva, J. (2014). Ethics of scientific peer review: Are we judging or helping the review recipients? In *Proceedings of the IEEE international symposium on ethics in science, technology and engineering* (pp. 1–14). 10.1109/ETHICS.2014.6893385.

Anjum, O., Gong, H., Bhat, S., Hwu, W.-M., & Xiong, J. (2019). PaRe: A paper-reviewer matching approach using a common topic space. In *Proceedings of the conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 518–528). 10.18653/v1/D19-1049.

Azoulay, P., Bonatti, A., & Krieger, J. L. (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy, 46*(9), 1552–1569. 10.1016/j.respol.2017.07.003.

Bar-Ilan, J., & Halevi, G. (2018). Temporal characteristics of retracted articles. *Scientometrics, 116*(3), 1771–1783. 10.1007/s11192-018-2802-y.

Beck, T. S. (2021). Image manipulation in scholarly publications: Are there ways to an automated solution? *Journal of Documentation Scopus*. 10.1108/JD-06-2021-0113.

Bolland, M. J., Grey, A., & Avenell, A. (2022). Citation of retracted publications: A challenging problem. *Accountability in Research, 29*(1), 18–25. 10.1080/08989621.2021.1886933.

Bornmann, L., Nast, I., & Daniel, H. D. (2008). Do editors and referees look for signs of scientific misconduct when reviewing manuscripts? A quantitative content analysis of studies that examined review criteria and reasons for accepting and rejecting manuscripts for publication. *Scientometrics, 77*(3), 415–432. 10.1007/s11192-007-1950-2.

Brainard, J. (2018). Rethinking retractions. *Science, 362*(6413), 390–393. 10.1126/science.362.6413.390.

Brown, S. J., Bakker, C. J., & Theis-Mahon, N. R. (2022). Retracted publications in pharmacy systematic reviews. *Journal of the Medical Library Association, 110*(1) Article 1. 10.5195/jmla.2022.1280.

Byrne, J. (2019). We need to talk about systematic fraud. *Nature, 566*(7742), 9 –9. 10.1038/d41586-019-00439-9.

Campos-Varela, I., & Ruano-Raviña, A. (2019). Misconduct as the main cause for retraction. A descriptive study of retracted publications and their authors. *Gaceta Sanitaria, 33*(4), 356–360. 10.1016/j.gaceta.2018.01.009.

Casnici, N., Grimaldo, F., Gilbert, N., Dondio, P., & Squazzoni, F. (2017). Assessing peer review by gauging the fate of rejected manuscripts: The case of the journal of artificial societies and social simulation. *Scientometrics, 113*(1), 533–546. 10.1007/s11192-017-2241-1.

Chen, C. Y., Kang, Y. N., Kuo, K. N., Glasziou, P., & Chen, K. H. (2021). Increasing retractions of meta-analyses publications for methodological flaw. *Systematic Reviews, 10*(1), 267. 10.1186/s13643-021-01822-2.

COPE Council (2019). COPE Guidelines: Retraction Guidelines. Committee on Publication Ethics. 10.24318/cope.2019.1.4.

Cho, K. (2008). Machine classification of peer comments in physics. *1st International Conference on Educational Data Mining Proceedings*, 192–196.

Clarivate Analytics. (2012). Publons. Retrieved from https://publons.com. Accessed June 23, 2022.

Ercegovac, Z., & Richardson, J. V. (2004). Academic Dishonesty, Plagiarism Included, in the Digital Age: A Literature Review. *College & Research Libraries, 65*(4), 301–318. 10.5860/crl.65.4.301.

Fang, F. C., & Casadevall, A. (2011). Retracted science and the retraction index. *Infection and Immunity, 79*(10), 3855–3859. 10.1128/IAI.05661-11.

Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences, 109*(42), 17028–17033. 10.1073/pnas.1212247109.

Fox, M. F. (1994). Scientific misconduct and editorial and peer review processes. *The Journal of Higher Education, 65*(3), 298–309.

Gerwing, T. G., Allen Gerwing, A. M., Avery-Gomm, S., Choi, C. Y., Clements, J. C., & Rash, J. A. (2020). Quantifying professionalism in peer review. *Research Integrity and Peer Review, 5*(1), 9. 10.1186/s41073-020-00096-x.

He, T. (2013). Retraction of global scientific publications from 2001 to 2010. *Scientometrics, 96*(2), 555–561. 10.1007/s11192-012-0906-3.

Horbach, S. P. J. M., & Halffman, W. (2019). The ability of different peer review procedures to flag problematic publications. *Scientometrics, 118*(1), 339–373. 10.1007/s11192-018-2969-2.

Hsiao, T. K., & Schneider, J. (2021). Continued use of retracted papers: Temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine. *Quantitative Science Studies, 2*(4), 1144–1169. 10.1162/qss_a_00155.

Kelly, J., Sadeghieh, T., & Adeli, K. (2014). Peer review in scientific publications: Benefits, critiques, & a survival guide. *EJIFCC, 25*(3), 227–243.

Kharasch, E. D., Avram, M. J., Clark, J. D., Davidson, A. J., Houle, T. T., Levy, J. H., et al., (2021). Peer review matters: Research quality and the public trust. *Anesthesiology, 134*(1), 1–6. 10.1097/ALN.0000000000003608.

Kovanis, M., Porcher, R., Ravaud, P., & Trinquart, L. (2016). The global burden of journal peer review in the biomedical literature: strong imbalance in the collective enterprise. *PLOS One*, (11), 11. 10.1371/journal.pone.0166387.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. SAGE.

Kühberger, A., Streit, D., & Scherndl, T. (2022). Self-correction in science: The effect of retraction on the frequency of citations. *PLOS One, 17*(12), Article e0277814. 10.1371/journal.pone.0277814.

Kurihara, Y., & Colletti, P. M. (2013). How do reviewers affect the final outcome? Comparison of the quality of peer review and relative acceptance rates of submitted manuscripts. *American Journal of Roentgenology, 201*(3), 468–470. 10.2214/AJR.12.10025.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology, 64*(1), 2–17. 10.1002/asi.22784.

Lu, S. F., Jin, G. Z., Uzzi, B., & Jones, B. (2013). The retraction penalty: Evidence from the web of science. *Scientific Reports, 3*(1) Article 1. 10.1038/srep03146.

Marcovitch, H. (2007). Misconduct by researchers and authors. *Gaceta Sanitaria, 21*(6), 492–499. 10.1157/13112245.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th international conference on neural information processing systems - volume 2* (pp. 3111–3119).

Mulligan, A., Hall, L., & Raphael, E. (2013). Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology, 64*(1), 132–161. 10.1002/asi.22798.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., et al., (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1) Article 1. 10.1038/s41562-016-0021.

Nair, S., Yean, C., Yoo, J., Leff, J., Delphin, E., & Adams, D. C. (2020). Reasons for article retraction in anesthesiology: A comprehensive analysis. *Canadian Journal of Anesthesia/Journal Canadien D'anesthésie, 67*(1), 57–63. 10.1007/s12630-019-01508-3.

Ortega, J. L. (2017). Are peer-review activities related to reviewer bibliometric performance? A scientometric analysis of Publons. *Scientometrics, 112*(2), 947–962. 10.1007/s11192-017-2399-6.

Rennie, D. (2016). Let's make peer review scientific. *Nature, 535*(7610) Article 7610. 10.1038/535031a.

Resnik, D. B., & Elmore, S. A. (2016). Ensuring the quality, fairness, and integrity of journal peer review: A possible role of editors. *Science and Engineering Ethics, 22*(1), 169–188. 10.1007/s11948-015-9625-5.

Resnik, D. B., Gutierrez-Ford, C., & Peddada, S. (2008). Perceptions of ethical problems with scientific journal peer review: An exploratory study. *Science and Engineering Ethics, 14*(3), 305–310. 10.1007/s11948-008-9059-4.

Schroter, S., Black, N., Evans, S., Godlee, F., Osorio, L., & Smith, R. (2008). What errors do peer reviewers detect, and does training improve their ability to detect them? *Journal of the Royal Society of Medicine, 101*(10), 507–514. 10.1258/jrsm.2008.080062.

Serghiou, S., Marton, R. M., & Ioannidis, J. P. A. (2021). Media and social media attention to retracted articles according to Altmetric. *PLOS One, 16*(5), Article e0248625. 10.1371/journal.pone.0248625.

Shamsi, A., Lund, B. D., & SeyyedHosseini, S. (2022). Sharing of retracted COVID-19 articles: An altmetric study. *Journal of the Medical Library Association : JMLA, 110*(1), 97–102. 10.5195/jmla.2022.1269.

Siler, K., Lee, K., & Bero, L. (2015). Measuring the effectiveness of scientific gatekeeping. *Proceedings of the National Academy of Sciences, 112*(2), 360–365. 10.1073/pnas.1418218112.

Steen, R. G. (2011). Retractions in the scientific literature: Is the incidence of research fraud increasing? *Journal of Medical Ethics, 37*(4), 249–253. 10.1136/jme.2010.040923.

Steen, R. G., Casadevall, A., & Fang, F. C. (2013). Why has the number of scientific retractions increased? *PLoS One, 8*(7), e68397. 10.1371/journal.pone.0068397.

Teixeira da Silva, J. A., Bornemann-Cimenti, H., & Tsigaris, P. (2021). Optimizing peer review to minimize the risk of retracting COVID-19-related literature. *Medicine, Health Care, and Philosophy, 24*(1), 21–26. 10.1007/s11019-020-09990-z.

The Center For Scientific Integrity. (2018). The retraction watch database. Retrieved from https://retractiondatabase.org. Accessed June 23, 2022.

The Science of Retraction. (2002). *Nature Neuroscience, 5*(12) Article 12. 10.1038/nn1202-1249.

Thelwall, M. (2022). Journal and disciplinary variations in academic open peer review anonymity, outcomes, and length. *Journal of Librarianship and Information Science* 09610006221079345. 10.1177/09610006221079345.

Van Noorden, R. (2011). Science publishing: The trouble with retractions. *Nature, 478*(7367) Article 7367. 10.1038/478026a.

Warne, V. (2016). Rewarding reviewers – Sense or sensibility? A Wiley study explained. *Learned Publishing, 29*(1), 41–50. 10.1002/leap.1002.

Wolfram, D., Wang, P., Hembree, A., & Park, H. (2020). Open peer review: Promoting transparency in open science. *Scientometrics, 125*(2), 1033–1051. 10.1007/s11192-020-03488-4.

Xu, S. B., & Hu, G. (2022). Retraction stigma and its communication via retraction notices. *Minerva, 60*(3), 349–374. 10.1007/s11024-022-09465-w.

Zhang, G., Xu, S., Sun, Y., Jiang, C., & Wang, X. (2022). Understanding the peer review endeavor in scientific publishing. *Journal of Informetrics, 16*(2), Article 101264. 10.1016/j.joi.2022.101264.

Zong, Z., Schunn, C. D., & Wang, Y. (2021). Learning to improve the quality peer feedback through experience with peer feedback. *Assessment & Evaluation in Higher Education, 46*(6), 973–992. 10.1080/02602938.2020.1833179.