How do journals deal with problematic articles. Editorial response of journals to articles commented in *PubPeer*

José-Luis Ortega; **Lorena Delgado-Quirós**

Nota: Este artículo se puede leer en español en:
https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87096

Recommended citation:

Ortega, José-Luis; **Delgado-Quirós, Lorena** (2023). "How do journals deal with problematic articles. Editorial response of journals to articles commented in *PubPeer*". *Profesional de la información*, v. 32, n. 1, e320118.

https://doi.org/10.3145/epi.2023.ene.18

Article received on August 24th 2022
Approved on December 12th 2022

José-Luis Ortega ✉
https://orcid.org/0000-0001-9857-1511

Instituto de Estudios Sociales Avanzados (IESA-CSIC)
Plaza Campo Santo de los Mártires, 7
14004 Córdoba, España
jortega@iesa.csic.es

Lorena Delgado-Quirós
https://orcid.org/0000-0001-8738-7276

Instituto de Estudios Sociales Avanzados (IESA-CSIC)
Plaza Campo Santo de los Mártires, 7
14004 Córdoba, Spain
ldelgado@iesa.csic.es

Abstract

The aim of this article is to explore the editorial response of journals to research articles that may contain methodological errors or misconduct. A total of 17,244 articles commented on in *PubPeer*, a post-publication peer review site, were processed and classified according to several error and fraud categories. Then, the editorial response (i.e., editorial notices) to these papers were retrieved from *PubPeer*, *Retraction Watch*, and *PubMed* to obtain the most comprehensive picture. The results show that only 21.5% of the articles that deserve an editorial notice (i.e., honest errors, methodological flaws, publishing fraud, manipulation) were corrected by the journal. This percentage would climb to 34% for 2019 publications. This response is different between journals, but cross-sectional across all disciplines. Another interesting result is that high-impact journals suffer more from image manipulations, while plagiarism is more frequent in low-impact journals. The study concludes with the observation that the journals have to improve their response to problematic articles.

Keywords

PubPeer; Bibliometrics; Retractions; Plagiarism; Data manipulation; Editorial notices; Journals; Journal impact; Scholarly communication; Publishing fraud.

Funding

This work was supported by the research project "NewSIS. New scientific information sources: analysis and evaluation for a national scientific information system" (Ref. PID2019-106510GB-I00) funded by the Spanish *State Research Agency* (*AEI*), *Ministry of Science and Innovation*, PN2019.

Acknowledgement

We thank Jaime A. Teixeira da Silva for his comments about the manuscript.

1. Introduction

Research journals are the central element in the current publishing system, where they are critical intermediaries between researchers and their scholarly audiences. This leading role gives them a great responsibility with regard to the research integrity of the published articles (**Marusic** *et al.*, 2007). Editorial boards and invited reviewers are the main gatekeepers to detect and filter erroneous and unreliable publications. However, the role of these actors is to select suitable studies for the journal and to improve the technical quality of the contributions, but not to discover data manipulations or infringements of copyright (**Martin**, 2012).

> The objective of this paper is to study the response of scholarly journals facing reports of research errors or misconduct about their publications

Furthermore, when these problematic articles are identified, many journals lack policies or procedures to handle these publications. In many cases, fraudulent practices go unnoticed because editors and reviewers lack of forensic skills and equipment to detect, for instead, image manipulations (i.e. forensic droplets) or text reusing (i.e. anti-plagiarism software). In other cases, editorial boards are unknown how reporting their suspicious to the authorities (i.e. universities, integrity offices), initiating an internal investigation or answering to their audiences. In consequence, many cases of errors in methods or questionable results have not been adequately treated, by not releasing its corresponding editorial notice (**Wager**, 2015). The creation in 1997 of the *Committee on Publication Ethics* (*COPE*) has helped editors and publishers to palliate this situation, increasing and normalizing the notifications (**Moylan**; **Kowalczuk**, 2016).

However, the absence of notices about investigations makes very hard to understand the real incidence of errors and misconduct in the current scientific literature, because we are not aware of whether an investigation has been carried out and, in that case, what decision was reached (**Smith**; **Godlee**, 2005). A supplementary way to understand this phenomenon is to explore social platforms such as *PubPeer*, where users can critically comment published papers, even report irregularities and fraud. This information allows to be contrasted with the journal responses to obtain a different point of view about the incidence of unreliable science.

This work aims to shed light on this problem exploring the response of journals to article reported in *PubPeer* of errors or misconduct. Using a new approach, comments about publications in that post-publication peer review site were compared with the response of journals by means of editorial notices, with the aim of studying how the journals react to troublesome articles.

2. Literature review

The study of the incidence of misconduct in the scientific literature has been focused on the response of journals releasing editorial notices. The first quantitative studies about these notifications were performed by **Budd** *et al.* (1998; 1999), who summarized the reasons of retractions and the citation of retracted papers. Later, other studies warned of the increase of this type of editorial publications (**Cokol** *et al.*, 2008, **Redman** *et al.*, 2008). **Steen** (2011) also perceived that the levels of misconduct appeared to be higher than in the past when he studied 742 articles from *PubMed*; and two year later, **Steen** *et al.* (2013) concluded that the increment of retractions was mainly due to lower barriers in the detection of bad practices. More recently, **Tripathi** *et al.* (2019) perceived that open access journals had a greater number of retractions as compared to subscription based journals, introducing the influence of the venue in the fraud. In this sense, many studies have observed a positive relationship between journal impact and editorial notices (**Cokol** *et al.*, 2007; **Fang** *et al.*, 2012; **Aspura** *et al.*, 2018; **Faggion** *et al.*, 2018).

Many other studies have focused on the content of the editorial notices, because not all of them are released by fraudulent cases. **Budd** *et al.* (1998; 1999) were the first one in exploring the content of the retractions and they found that 37% of them are due to evident misconduct. **Nath** *et al.* (2006) found that more than the half of the retractions in Medicine were caused by nonintentional mistakes. **Wager** and **Williams** (2011) detected that 28% of retractions were due to research misconduct and 17% to redundant publication. **Fang** *et al.* (2012), studying more than 2,000 research articles, observed that 67.4% of retractions were attributable to misconduct. In a similar study, **Decullier** *et al.* (2013) found that plagiarism (20%) and fraud (14%) were the common motives for retraction. Most recently, **Lei** and **Zhang** (2018) detected that misconduct is the cause of three quarters of the retractions in China; while **Vuong** (2020) observed important inaccuracies in retraction notices, where 10% of them did not contain information related to reasons for retractions. This disparity in the results evidences problems in the definition of misconduct, the creation of a standard taxonomy and the lack of transparency in some notices.

However, fewer articles have treated the response rate of journals to suspicious or problematic articles. **Wager** (2007) was the first one to address this issue, analysing cases submitted to *COPE*. Her results showed that from 79 cases, 49 (62%) were reported in the journal. That same year, **Cokol** *et al.* (2007) estimated the proportion of flawed articles no retracted, finding that 10,000 articles should be retracted, instead of the 596 observed. **Neale** *et al.* (2007) studied the cases of misconduct reported by the *Office of Research Integrity* (*ORI*) annual reports and *National Institutes of Health* (*NIH*) *Guide*, and they found that 83% of articles were noticed in the journals. In a similar study, **Resnik** and **Dinse** (2013) found 127 out of 174 (73%) fraudulent publications with an editorial notice. **Elia** *et al.* (2014) checked the editorial response to articles that warranted retraction from a specific misconduct case, and the result was a 90% of retracted articles. These studies are based on previously investigated publications; accordingly, the degree of response is high. Nevertheless, there are few studies that have explored this issue using external sources. **Brookes** (2014), who analysed anonymous complaints in a specialized blog. His findings were that only 23% of the reported articles were later corrected or retracted. More recently, **Bik** *et al.* (2018) manually inspected image manipulation in *Molecular and cellular biology* journals, and they found that approximately 10% of the papers with demonstrated image manipulation were retracted.

The post-publication review site, *PubPeer*, is becoming an interesting source for studies on scientific misconduct due to the posting of comments on questionable practices. The first studies have focused on discussing conceptual and ethical

issues. **Blatt** (2015) and *PubPeer* (2015) debated on the risks of anonymous comments; while **Da-Silva** (2018a) criticized the ownership and copyright of these comments. However, more and more studies have explored its influence on the elucidation of suspicious publishing practices. **Wager** and **Veitch** (2017) used that platform to test its ability in reporting fraudulent cases, and they concluded that only 9% of comments required a journal reaction. **Ortega** (2021) analysed the coverage of editorial notices in this site, finding that the relationship between them is scant. More recently, **Ortega** (2022) categorized the content of *PubPeer* posts, observing that more than two-thirds of comments are posted to report some type of misconduct.

> This work sheds light on the response of journals to articles reported for errors or misconduct on the post-publication peer review site *PubPeer*

3. Objectives

The objective of this paper is to study the response of scholarly journals facing reports of research errors or misconduct about their publications. Using a descriptive approach, the study attempts to analyse the number and types of editorial notices released by issue type and journal. In addition, the incidence of these responses is analysed regarding to the academic impact and the research area of journals. Four research questions were addressed in this study:

- How often do journals react on articles reported of errors or misconduct? And what type of editorial notice do they release?
- How does evolve the proportion of editorial notices throughout the years? Is it improving the identification and correction of suspect literature?
- Is there any relationship between the academic impact of journals and their editorial response?
- Is this response different according to research areas?

4. Methods

4.1. Sources

PubPeer defines itself as journal club where scholarly documents can be discussed after being published or uploaded to the Web. Created in October 2012, the success of this post-publication peer review site resides in the possibility of posting comments anonymously. This particular feature has caused the specialization of the site in reporting of misconduct and errors in the scientific literature. This fact is generating considerable controversy because many authors feel defenceless in the face of unknown accusers (**Torny**, 2018). On the contrary, research integrity is benefiting from this format because allows the uncovering of bad practices with no reprisals. *PubPeer* also included comments from external sources such as *Twitter* and *PubMed Commons* (**Da-Silva**, 2018b).

Retraction Watch is a web blog created in 2010 by two scientific journalists, Ivan Oransky and Adam Marcus, concerned on the overall absence of transparency in the investigation of misconduct in science and, concretely, on the lack of information in retractions. In this manner, the blog investigates the hidden reasons behind inscrutable retraction notices with interviews and inquiries about retraction cases. These retractions are storage in a publicly accessible bibliographic database, *Retraction Watch Database*:
https://www.retractiondatabase.org

This is an exhaustive list of retracted publications, which includes the reasons of the retraction from the retraction notice or as result of their investigations.

PubMed is an academic search engine created by the *National Institutes of Health* (*NIH*). Launched in 1997, the engine connects with *Medline* database to retrieve research publications on Biomedicine and related disciplines. *PubMed* was used because is one of the few search engines that links the original publication with the associated editorial notices.

4.2. Data access and extraction

PubPeer does not provide open access to their data. Due to this, information about publications and associated comments were directly extracted from the website (*pubpeer.com*) using web scraping techniques.

Two samples were extracted for this study in different moments. In March 2019, 32,097 threads and 65,179 posts were obtained. This sample was enlarged and updated with a second sample in January 2020, which included 7,659 threads and 21,200 posts. In total, 86,379 posts from 39,757 threads associated with 24,779 publications were retrieved. *PubPeer* does not provide a full list of the commented publications. Then, a search strategy was designed to retrieve the largest sample of publications but without committing any bias in the selection process. The best option was to select neutral terms that retrieve documents from any disciplines and written in any alphabetic language. The first letters of the alphabet (a, b, and c) were searched in the standard search box to ensure the randomness of the sample. Only these first letters were used because the results showed a high overlap and it was estimated that querying for the other letters would report similar results, but with more effort.

> The absence of notices about investigations makes it very hard to understand the real incidence of errors and misconduct in the current scientific literature

> An editorial notice is released by a research journal to correct or highlight any problem with a published research article, and is associated with a decision of the editorial board

These queries retrieved comments to publications, including the internal ID of each paper. A web crawler was designed to retrieve this information. The crawler code consisted in pasting a base URL (i.e., *https://blog.pubpeer.com/publications*) and the ID of each publication previously retrieved (i.e. CF52AD098D3AC462697D-50B97B3105). Next, from each URL, bibliographic metadata and information about the comments associated to those publications (user, text, date, etc.) were extracted and storage in a csv file. *WebQL Studio* was used for this task:
https://www.ql2.com

The sample was cleaned removing comments generated by robots (11,469, 13.3%) when the same text was repeatedly posted by the same account. Concretely, that is the case of "statcheck" user who checked statistical inconsistencies in thousands of articles and then included automatic posts in *PubPeer* about the resulting test for each publication:
https://retractionwatch.com/2016/09/02/heres-why-more-than-50000-psychology-studies-are-about-to-have-pubpeer-entries

Publications without user comments were also removed (6,328, 7.3%). Finally, 68,595 (79.4%) posts about 26,133 research documents published after 2000 were selected. This cut-off was set because the number of discussed papers in *PubPeer* published before 2000 is very low, which could produce statistical distortions in the longitudinal analysis (Figure 2).

4.3. Classification and selection criteria

From these records a sample of 17,244 (66%) articles were classified according to the content of the comments. The remaining publications (8,889, 44%) were rejected due to the comments were not sufficiently explanatory (e.g., very short comments, little reasoning) or they do not fit with the classification scheme. The classification process was based on the extraction of keywords that described the content of the comments. Then, publications through the keyword's comments, were grouped in seven categories (**Ortega**, 2022):

- Positive review: Comments that praise and highlight publications according to the reach and importance of the results.
- Critical review: Comments that discuss the methods and results and their interpretations. This group includes discussions about theoretical implications and scientific disagreements.
- Lack of information: Inside Critical review, this is a sub-category that addresses the problematic absence of information about how the study was performed, the availability of raw data, and lack of relevant bibliographic references.
- Honest errors (**Resnik**; **Stewart**, 2012): They could be rectifiable mistakes (e.g., erratum) due to confusion and oversight in the writing of the paper.
- Methodological flaws: They are motivated by a lack of awareness of statistical or other scientific techniques (e.g., western blots, spectroscopy) that throw up wrong results (e.g., correlation fishing, bar errors, loading controls). This category could be bordering on fraud, because this confusion could be intended to obtain the desired results. However, such intentionality is not always evident, and these issues are given the benefit of doubt.
- Publishing fraud: Interference with the publishing system to increase production and impact. It mainly includes plagiarism, reused text, ghost authorship and fake peer review.
- Manipulation: Intentional edition and manipulation/fabrication of data and images to obtain better results than those expected, to corroborate the desired hypothesis.

Finally, to validate the accuracy of this classification procedure, a sub-sample of comments (4,000) were manually classified and compared with the original procedure. A confusion matrix showed a high overall precision (88.1%), demonstrating that close to nine out of ten posts were correctly assigned (**Ortega**, 2022).

Subject matter classification and impact quartiles we obtained from *SCImago Journal & Country Rank* (*SJR*) portal, 2020 version:
https://www.scimagojr.com

This site uses *All Science Journal Classification* (*ASJC*) to categorize and rank journals. If a journal is assigned to more than one discipline, and therefore more than one quartile, then the discipline with the best quartile is selected. This ranking was used because it includes more journals than others (i.e., *Journal Citation Reports*).

When a publication had generated several editorial notices, the most serious one was selected. The importance goes from Erratum, Expression of Concern to Retraction. Thus, whether an article has been corrected with an erratum, and later was finally retracted, we have then considered this paper as retracted.

Data sets about this study are openly available in:
https://osf.io/hecbg

4.4. Editorial notices coverage

First, it is important to verify whether a publication has received an editorial notice. An editorial notice is a publication released by a research journal to correct or notice any problem about a published research article, and associated to a decision of the editorial board. These editorial notices are mainly erratum, expression of concern and retraction. *PubPeer* indicates when a publication has been subject of an editorial notice, such as retraction, expression of concern and erratum (**Ortega**, 2021). However, we are not aware of the reliability of this platform detecting when an editorial notice is released and it is associated to a publication. A way to test this ability is matching the sample of publications commented in *PubPeer* with the *Retraction Watch*'s database (24,421 publications) and the set of publications with editorial notices in *PubMed* (8,621 publications).

This comparison was performed with the initial sample of 26,133 publications because the coverage test does not depend on the thematic classification of the publications, bringing a most reliable view about the coverage. 3,076 (11.8%) were subject of an editorial notice according *PubPeer*. The search in *Retraction Watch* and *PubMed* achieved to detect 370 (12%) additional publications with notices. 366 (11.9%) from *Retraction Watch* and 288 (9.4%) from *PubMed*, being 3,445 (13.2%) the total set of articles with editorial notices. This means that *PubPeer* has an efficacy of 89.3% detecting this type of editorial publications.

5. Results

Finally, this study is based on 17,244 publications thematically classified according to their comments in *PubPeer*, from which 3,203 (18.6%) received an editorial notice. 14,290 (82.9%) are considered troublesome articles, publications that are reported of Publishing fraud, Manipulation, Methodological flaws, or Honest errors and therefore they could be subject of an editorial notice.

Table 1. Distribution of publications and editorial notices according to type of comment

Type of comments	Articles		With editorial notices	
	n	%	n	%
Positive review	131	0.8	2	1.5
Critical review	2,539	14.7	108	4.3
Lack of information	284	1.6	23	8.1
Honest errors	408	2.4	86	21.1
Methodological flaws	1,195	6.9	229	19.2
Manipulation	10,989	63.7	2,256	20.5
Publishing fraud	1,698	9.8	499	29.4
Total	17,244	100.0	3,203	18.6

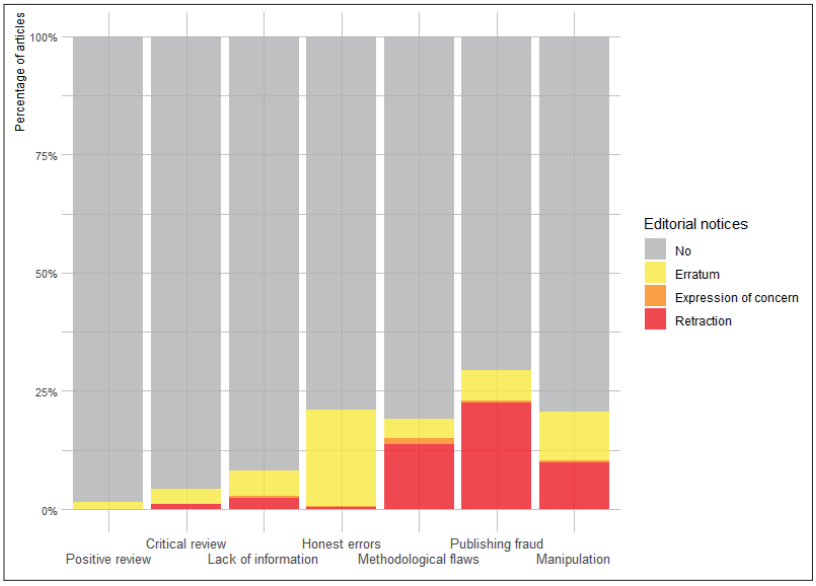

Figure 1. Distribution of editorial notices by type of comment

Table 1 shows the number of publications commented in *PubPeer* according to the type of comment posted on *PubPeer* and the consequent editorial notice. Manipulation (63.7%) and Critical review (14.7%) are the categories with the most publications, followed by Publishing fraud (9.8%) and Methodological flaws (6.9%). These results confirm that *PubPeer* is used almost exclusively for reporting errors and misconduct cases. The high incidence of image manipulations has been previously reported (Bik *et al.*, 2016; 2018), and confirms that this problem is spread across the biology research. Publications with comments related to misconduct such as Publishing fraud (29.2%) and Manipulation (20.5%) receive the largest number of editorial notices. But also, Honest errors (21.1%) and Methodological flaws (19.2%) gather a considerable proportion of editorial notices. In general, only 18.6% of the publications have been subject of an editorial notice. If this percentage is limited to publications accused of misconduct or errors (troublesome articles), the percentage climbs to 21.5%. This could mean that almost only one out of five research papers suspected of errors or misconduct according to *PubPeer* received a notification from the editorial board of the journal.

In detail, Figure 1 shows the percentage of different editorial notices by type of comment. This allows us to know what type of reaction is more common according to the type of problem. The bar graph depicts that Errata are mainly published for Honest errors (20.3%) and Manipulation (10.1%). In the first case, minimum and honest errors can easily be solved with an erratum. However, the second case, illustrates that a considerable proportion of manipulations could

be due to unintentional mistakes in the use of images or that troublesome images are just retired to avoid the entire retraction of the publication. Publishing fraud (22.5%), Methodological flaws (13.8%) and Manipulation (9.8%) are the type of comments that generate most retractions. The high proportion of Publishing fraud can only be explained by the great difficulty of correcting plagiarism, compromised peer review or ghost authorship.

Figure 2 displays the proportion of editorial notices according to the publication date of publications. This graph aims to show if the percentage of editorial notices increases or decreases as time goes on. The trend of publications without editorial notices is descending up to 2011, while Retraction increases up to 2012 and Erratum up to 2015. This change of tendency could be mainly due to the great delay in the release of these notices (**Stricker**; **Günther**, 2019; **Ortega**, 2021). Considering the time delay between the moment in which a paper is commented in *PubPeer* and then is subject of an Erratum (296 days) or Retraction (541 days), it is necessary to estimate the influence of this delay. A linear fit allows us to estimate the percentage of editorial notices for the most recent publications. Dot lines show the estimation, R^2 the goodness of fit and β_1 the slope coefficient. Thus, for articles published in 2019, we estimate a proportion of 34.1% of editorial notices for troublesome publications (R^2=.79 and β_1=-.0115), whereas the proportion of retractions climbs to 19.6% (R^2=.71 and β_1=.0068) and errata to 14.4% (R^2=.85 and β_1=.0051).

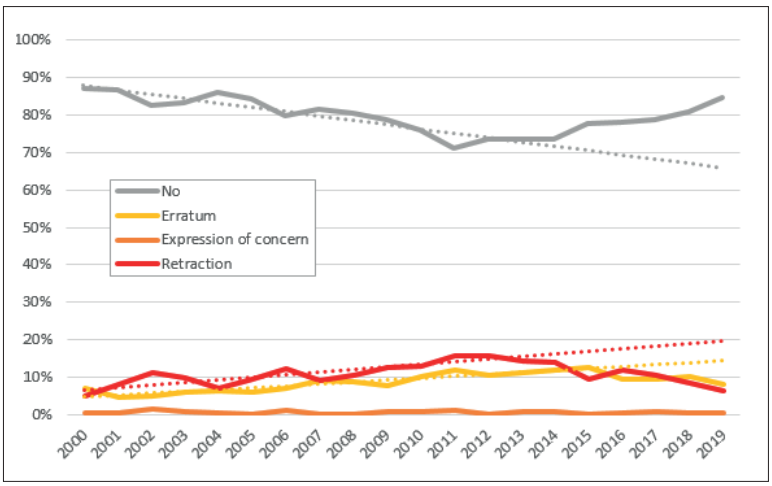

Figure 2. Proportion and estimation of editorial notices by publication date (dot lines show the linear estimation)

Table 2. The ten journals with most troublesome articles in *PubPeer* and the proportion of editorial notices

Journal	Troublesome articles		With editorial notices	
	n	%	n	%
Journal of biological chemistry	751	5.3	287	38.2
PLoS one	526	3.7	191	36.3
Oncotarget	407	2.8	53	13.0
Cancer research	395	2.8	103	26.1
Oncogene	335	2.3	48	14.3
Proceedings of the National Academy of Sciences (*PNAS*)	266	1.9	60	22.6
Molecular and cellular biology	220	1.5	47	21.4
Blood	189	1.3	29	15.3
Scientific reports	186	1.3	35	18.8
Clinical cancer research	176	1.2	51	29.0
Total artículos en las diez revistas	10,839	24.2	904	26.2
Total	14,290	100	3,070	21.5

Table 2 shows the ten journals with the highest number of problematic articles in *PubPeer* along with the proportion of editorial notices. The purpose of this table is to describe differences in the editorial management of suspicious articles at journal level. *Journal of biological chemistry* (5.3%) and *PLoS one* (3.7%) are the journals that have the most suspect publications, and also those that react the most to this type of publications, correcting 38.2% of the publications in *Journal of biological chemistry* and 36.3% in *PLoS one*. However, the journals that few respond to troublesome articles are *Oncotarget* (13%) and *Oncogene* (14.3%).

Figure 3 details the type of editorial notice released by each journal according to the different problems identified in *PubPeer*. This picture enables to ascertain what is the particular response of the journals that suffer the most from suspect articles. All the journals correct honest errors with errata in a high proportion (53.7%), being *Oncogene* and *Clinical cancer research* the journals that release errata in every case of honest errors. However, this response ratio drops significantly in the other more serious issues. In Methodological flaws, journals respond with an editorial notice in 26.9% of the cases, mainly retractions (16.8%). *PLoS one* (60.9%) and *Oncotarget* (36.4%) are the journals with the most editorial notices about this issue. A similar proportion of editorial notices is found in Manipulation (23%), with 11.6% of errata and 9.5% of retractions. *Journal of biological chemistry* (34.7%) and *PLoS one* (28.6%) are the journals with the highest proportion of editorial notices. Finally, Publishing fraud is the second type of complaint with most editorial responses (43.8%), being the retraction the most frequent (21.8%). Again, *Journal of biological chemistry* (79.2%) is far and away the journal that more respond to publishing misconduct.

> Almost only one out of five research papers suspected of errors or misconduct according to *PubPeer* received a notification from the editorial board of the journal

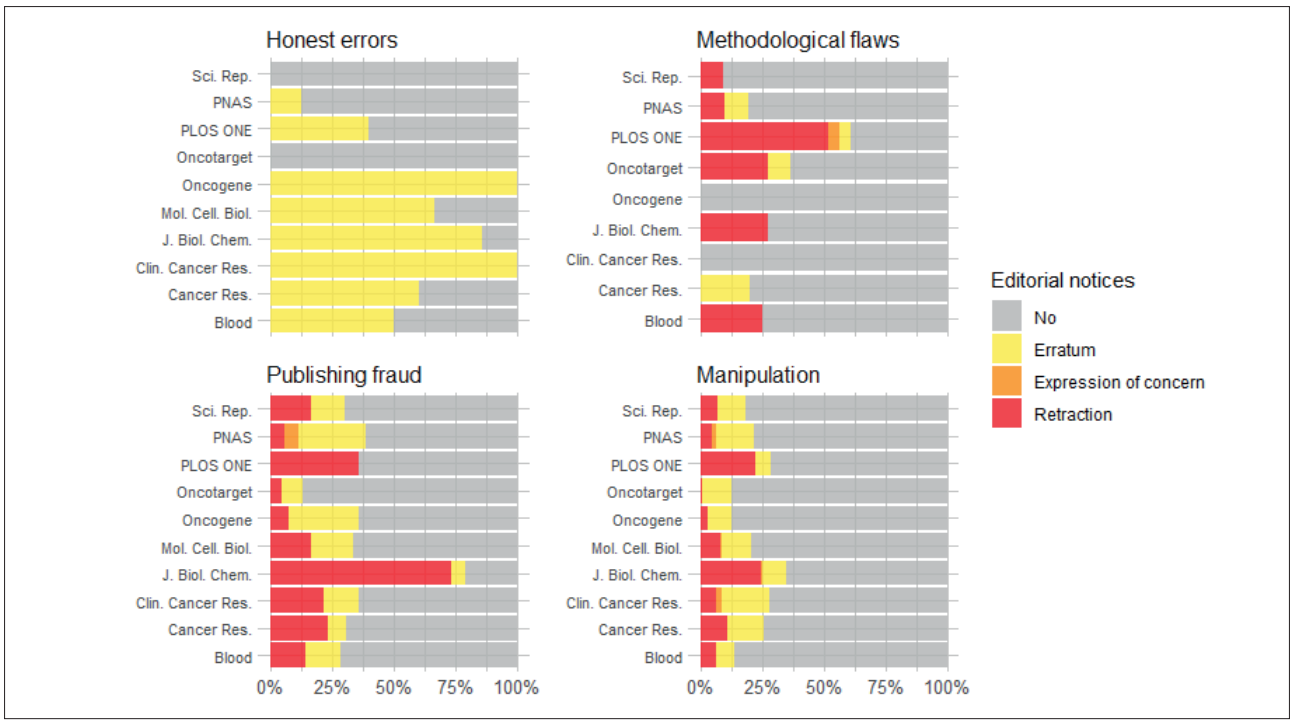

Figure 3. Distribution of editorial notices by type of comment in the ten journals with the most troublesome articles

Table 3. Distribution of troublesome articles and editorial notices by impact quartile

Quartile	Journals		Troublesome articles		With editorial notices	
	n	%	n	%	n	%
Q1	7,646	29.5	11,328	79.3	2,617	23.1
Q2	6,527	25.2	2,091	14.6	333	15.9
Q3	6,030	23.3	410	2.9	71	17.3
Q4	5,684	22.0	122	0.9	17	13.9
No indexed			339	2.4	34	10.0
Total	25,887	100	14,290	100	3,072	21.5

Table 3 describes the proportion of troublesome articles according *PubPeer* and their editorial notices in each impact quartile. The impact of journals is measured by *SJR* and grouped by quartiles. Not indexed journals in *SJR* are grouped as Not indexed. The aim is observing the incidence of these publications according to the research quality of the journals. The results show a high concentration of troublesome articles in Q1 journals (79.3%), which could suggest that fraudulent publications could be attracted by the prestige of high impact journals or that that category of journals could be more exposed to the public scrutiny. The proportion of responses to these papers slightly falls from the 23.1% in Q1 to the 10% of non-indexed journals in *Scopus*, which would indicate an association between the impact of the journal and its ability to detect suspect articles.

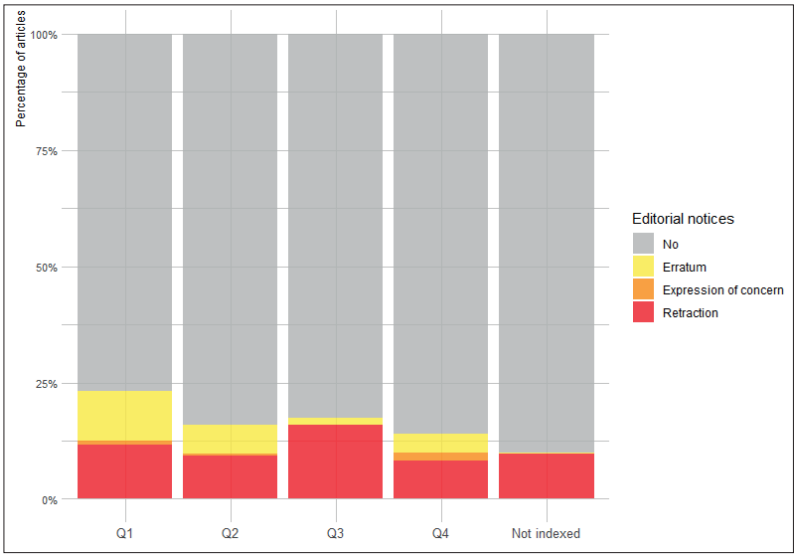

Figure 4. Distribution of editorial notices by impact quartile

Figure 4 illustrates the distribution of the different editorial notices in the impact quartiles. The aim is to appreciate if there is any relationship between editorial notices and the impact of the journals. The bar graph shows that the number of editorial notices drops as the impact decreases. This decline of editorial notices is mainly due to errata, which go from

10.7% in Q1 to 4.1% in Q4 and 0.2% in Not indexed. This descending pattern evidences that the use of errata as correction mechanism is more associated to high impact journals. However, retractions are more transverse, with a constant proportion in all the quartiles, going from the 15.9% of Q3 to the 8.2% of Q4. These results are in line with **Campos-Varela** *et al.* (2021). A possible explanation of these tendencies could be found in Figure 4.

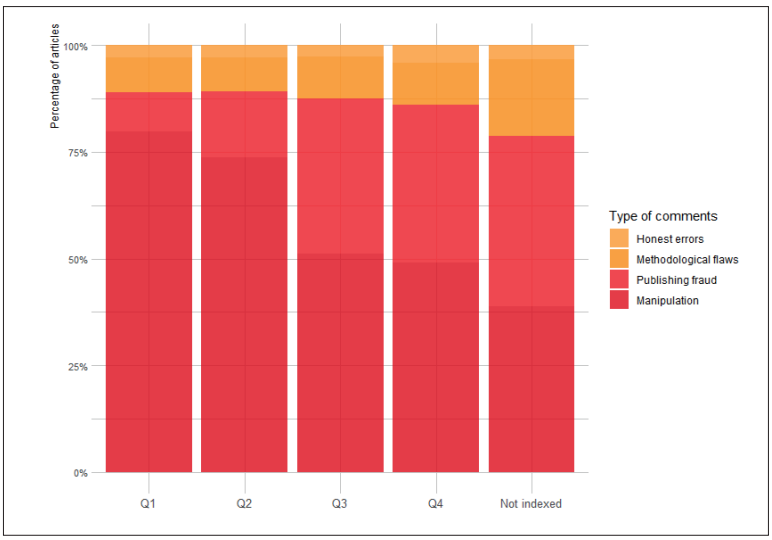

Figure 5. Distribution of troublesome articles by type of error or misconduct in each impact quartile

Figure 5 depicts the distribution of type of error or misconduct by impact quartile of the journal. This result aims to show how the incidence of different problems in *PubPeer* changes with the impact of the journal. It is interesting to notice that the two types of misconduct (Publishing fraud and Manipulation) evolve in a contrary direction. Manipulation is a very frequent problem in high impact journals (Q1=79.9%), while it considerably falls in Q4 (49.2%) and more than the half in not indexed journals (38.9%). Contrarily, Publishing fraud is little significant in Q1 (9.2%), but it gains importance in Q3 (36.3%), Q4 (36.9%) and not indexed journals (39.8%). These opposed trends could be interpreted in different ways. Image or data manipulation is a complex practice difficult to uncover and focused on justifying significant advances that increase the prestige of researchers. It is possible that these practices could be more focused on high impact journals, because these venues publish the most important discoveries in each discipline. On the other hand, Publishing fraud is mainly oriented to increase the production and not really the prestige. In addition, a great part of publishing fraud is plagiarism, and it is possible that this practice could be addressed to low impact journals with less editorial control. As we have observed in Figure 1, Errata are very frequent in cases of manipulation. Then, the great presence of manipulations in Q1 and Q2 would explain the results of Figure 4, where the high presence of errata in high impact journals could be due to these journals have more cases of manipulation.

Table 4. Distribution of troublesome articles and editorial notices by research area

Research areas	Journals		Troublesome articles		With editorial notice	
	n	%	n	%	n	%
Health Sciences	5,497	16.8	3,289	19.6	636	19.3
Life Sciences	3,601	11.0	9,510	56.6	2,136	22.5
Multidisciplinary	102	0.3	1,446	8.6	413	28.6
Physical Sciences	13,607	41.6	2,025	12.0	369	18.2
Social Sciences & Humanities	9,921	30.3	539	3.2	113	21.0
Total	32,728	100	16,809	100	3,667	21.8

Table 4 depicts the reaction of journals to troublesome articles according to the main disciplinary area in *ASJC*. Note that articles and journals could be classified in more than one research area. Life Sciences (56.6%) and Health Sciences (19.6%) are the disciplines with the most suspect publications. However, Multidisciplinary (28.6%) journals release more editorial notices than Life Sciences (22.5%) and Social Sciences & Humanities (21%). This suggests that journals in Multidisciplinary category would have more editorial control on problem publications, whereas Health Sciences (19.3%) journals could detect less erroneous articles.

Figure 6 depicts the proportion of editorial notices by troublesome articles grouped by the main disciplinary area. There are not important differences between disciplines, which would mean that journals react in the same manner independently of the research field. Thus, in Publishing fraud, Life Sciences shows 36.6% of editorial notices, while Physical Sciences only 25.3%. Regarding to Manipulation, there are also little differences, outstanding Multidisciplinary (28.3%) and Social Sciences & Humanities (27.7%) as the fields with the most editorial notices. Honest errors and Methodological flaws describe more differences, being Life Sciences (33.3%) the area with the most editorial releases for the first issue, and Multidisciplinary (32.6%) for the second one. According to the type of notice, the picture shows that errata is more frequent for Honest errors (20.2%) and Manipulation (10.3%).

> This low response rate suggests that the publishing system in general is not aware of external investigations, mainly from web platforms such as *PubPeer*, which can negatively affect the correction of misconduct in science

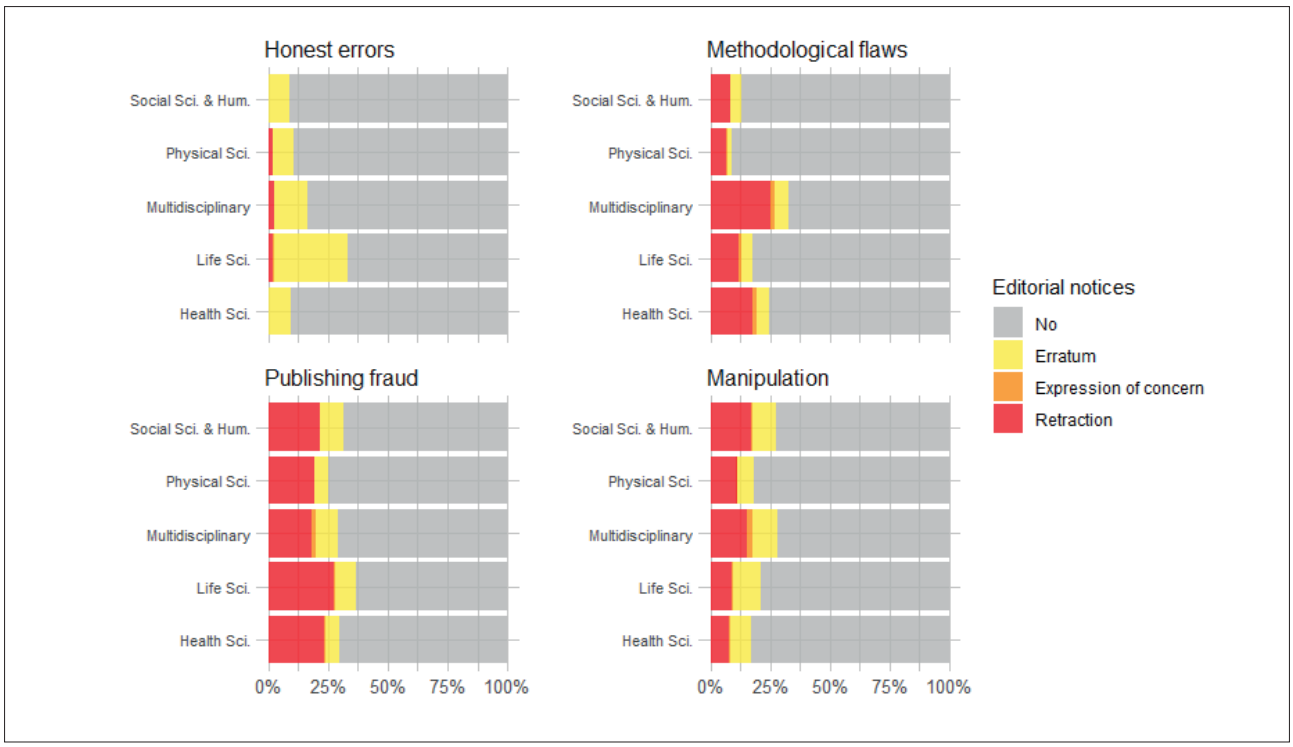

Figure 6. Distribution of editorial notices by type of comment according the main disciplinary subject area of the journal

6. Discussion

This study provides the opportunity to relate the complaints expressed in *PubPeer* on errors and misconduct of research papers with the consequent editorial response of journals. The most surprising result is the low response rate of research journals when a scientific paper is reported of misconduct or error in *PubPeer*. Only 21.5% of the papers that should deserve a notification were subject of an editorial notice. This means that almost only one out of five reported articles in *PubPeer* are corrected by journals. A similar proportion (23%) was also found by **Brookes** (2014) in articles publicly discussed in a blog. These proportions depend on what type of errors or misconduct we think that deserve an editorial notice. For example, the early *PubPeer* comments classification conducted by **Wager** and **Veitch** (2017) concluded that only 9 (7%) articles required a journal action. However, analysing their classification, we found that this figure should be 31 (4 for Fabrication, *2* for other misconduct, 5 for honest errors and 20 for methodological issues). If we compare this amount with the number of editorial notices released (5), the response rate is even lower than in our study (16%). In line with this result, the manual image inspection of **Bik** *et al.* (2018) verified that approximately only 10% of the papers with evident manipulations were retracted. This low response rate suggests that the publishing system in general is not aware of external investigations, mainly from web platforms such as *PubPeer*, which can get worse the correction of misconduct in science.

However, other studies based on already investigated cases show a higher journal response, suggesting that the reaction of journals is determined by the conclusions of official investigations (**Wager**, 2014). Even though, this response is not complete and significant number of articles do not receive any alert. **Neale** *et al.* (2007) found that 83% of articles reported of misconduct were noticed, and **Resnik** and **Dinse** (2013) found 127 out of 174 (73%) fraudulent publications had an editorial notice. These studies show that a non-trivial proportion of articles that have already been investigated and with a clear statement of misconduct (17% in the first, and 27% in the second one), have not been notified by their journals. These low response rates could indicate some disconnection of the publishing system with other scholarly environments (research organizations, integrity offices, web platforms, etc.), that impede be aware of misconduct reporting. Another possible causes would be a deliberate neglect of editorial duties (**Shelomi**, 2014), lack of misconduct policies (**Bosch** *et al.*, 2012) or absence of investigation by research organizations (**Wager**, 2015). Independent of the reasons, these results evidence that the identification and correction of erroneous literature is far from being properly addressed because it involves the responsible engagement of multiple agents (authors, organizations, journals, public) to create an efficient correction system.

> The identification and correction of erroneous literature is far from being properly addressed because it involves the responsible engagement of authors, organizations, journals, and public to create an efficient correction system

The positive trend observed in Figure 2 allows us to be slightly optimistic about the improvement in the detection of fraudulent practices in research publications, because the share of suspect publications with an editorial notice increases .07% each year. However, this growth

rate is still slow and more actions would be needed to close the current gap in the correction of erroneous and fraudulent literature.

> There is a high concentration of troublesome articles in Q1 journals (79.3%), which could suggest that fraudulent publications could be attracted by the prestige of high impact journals, or that that category of journals could be more exposed to public scrutiny

It is also interesting to discuss how editorial notices are used for different problems. Errata, which is accustomed to be used for minor cases (96.5% in Honest errors), is also used for serious cases (49.3% of Manipulation), while retractions are also released for non-intentional errors (72% for Methodological flaws and 30% for Lack of information). These different uses of editorial notices can cause confusion about their meaning when they are utilized for research integrity studies (**Da-Silva**, 2022). This result suggests that the study of research misconduct cannot be based only on retractions because there is an important amount of, for instance, manipulations that are corrected with errata (false negatives) and retractions due to non-fraudulent practices (false positives) (**Nath** *et al.* 2006; **Campos-Varela** *et al.*, 2021).

Other significant result is the high proportion of problematic articles in Q1 journals (79%), when the general proportion of articles is almost the half (44%) in this quartile (**De-Moya-Anegón**, 2020). This great concentration evidences that the publication of erroneous or fraudulent studies mainly occurs in high impact journals (**Steen**, 2011). The list of journals with the most erroneous articles includes important multidisciplinary journals (i.e., *PLoS one*, *PNAS*, *Scientific reports*) and reputed journals in their fields (i.e., *Cancer research*, *Oncogene*, *Blood*). This high incidence seems to indicate that fraudulent publications target top journals because these venues receive more attention, which could increase their prestige and fame. This great visibility would also attract potential commenters, overestimating the reporting of problem papers in these journals. Otherwise, these journals compete in publishing adventurous and attractive studies, which would increase the risk of accepting troublesome articles.

All these factors could explain the high incidence of errors and misconduct in high impact journals. However, Figure 4 shows a more nuanced view, in which the image manipulation is more frequent in high impact journals, while publishing fraud occurs more often in average journals. This fact was already observed by **Fang** *et al.* (2012). These opposed patterns would suggest that manipulation is used to achieve success, while plagiarism and reutilization to increase the production, independent of the quality of the publication. In addition, the high disproportion between Manipulation (63.7%) and Publishing fraud (9.8%) evidences that the scientific success is more attractive than the mere production. This landscape would explain that Q1 journals (23.1%) react more than other categories, due perhaps to a higher control by their editorial boards or simply because they are the object of the majority of the misconduct practices (**Corbyn**, 2012).

7. Limitations

An important limitation that could be attributed to this study is that not all the complaints in *PubPeer* are based on evidences. This platform does not check the veracity of the reports and therefore it could be possible to observe unfound accusations that do not deserve a statement. In that case, the proportion of articles that deserve an editorial notice could be lower. However, this fact only can be demonstrated when precisely journals initiate investigations and release editorial notices. Only in those cases, when an investigation has been accomplished, and a notification is released, we can assume veracity or falseness to these comments.

Another important limitation is the slowness of the editorial actions, which could delay the release of editorial notices (**Stricker** and **Günther**, 2019; **Ortega**, 2021). This problem causes that some recent publications that would deserve an editorial notice could be still under investigation. Figure 2 has evidenced this delay in more than 5 years and has estimated that the response to these troublesome articles would climb to 34% if this delay is taking into account. Therefore, forthcoming studies are necessary to confirm the reliability of these findings with more robust data. A second question is related to *PubPeer* as data source, because publications discussed in this site would influence on their investigation and the release of editorial notices. Although we have not found evidences of that, it is a question that we must consider in the interpretation of the results.

8. Conclusions

The main conclusion of this study is that, in average, only 21.5% publications that should deserve a correction or retraction according to *PubPeer* are subject of an editorial notice. Although this proportion would climb to 34.1% in recent publications, this result demonstrates that there is still a long way to reach a successful detection and investigation of misconduct publications. Errata are principally published for Honest errors (20.3%) and Manipulation (10.1%), while Retractions are mainly used for Publishing fraud (22.5%). The evolution of this editorial control improves throughout the years, with a steady increase of 0.07% each year. However, this rate is very slow and more actions are necessary to improve this share.

> It seems that manipulation is used to achieve success, while plagiarism and reutilization are used to increase production, irrespective of the quality of the publication

The study has shown that there is a positive relationship between the impact of a journal and the proportion of publications with an editorial notice. High-impact journals publish more problematic publications, which causes more editorial notices, mainly errata. It is also interesting to appreciate that high-impact journals are more damaged by manipulation, while the low-impact journals suffer from Publishing fraud.

> The appearance of post-publication peer review sites such *PubPeer* or *Publons*, and other social networks, bring a new perspective to misconduct in the scientific literature

Finally, the disciplinary analysis does not have disclosed any important thematic difference. Life Sciences (56.6%) and Health Sciences (19.6%) are the disciplines with the most problematic publications. Multidisciplinary (28.6%) category has more editorial control on erroneous articles, being the strictest field detecting manipulations (28.3%). Life Sciences (33.3%) is the research area that more detects honest errors, while Physical Sciences (25.3%) releases fewer editorial notices for Manipulation.

In general, we can conclude that the appearance of post-publication peer review sites such *PubPeer* or *Publons* and other social networks open a new perspective about the misconduct in the scientific literature. They provide an open and participatory environment to audit publications, questioning the role of institutions and fostering the demands of the scholarly community to find better correction mechanisms in science.

9. References

Aspura, M. Yanti-Idaya; **Noorhidawati, Abdullah**; **Abrizah, Abdullah** (2018). "An analysis of Malaysian retracted papers: Misconduct or mistakes?". *Scientometrics*, v. 115, n. 3, pp. 1315-1328.
https://doi.org/10.1007/s11192-018-2720-z

Bik, Elisabeth M.; **Casadevall, Arturo**; **Fang, Ferric C.** (2016). "The prevalence of inappropriate image duplication in biomedical research publications". *mBio*, v. 7, n. 3, e00809-16.
https://doi.org/10.1128/mBio.00809-16

Bik, Elisabeth M.; **Fang, Ferric C.**; **Kullas, Amy L.**; **Davis, Roger J.**; **Casadevall, Arturo** (2018). "Analysis and correction of inappropriate image duplication: The molecular and cellular biology experience". *Molecular and cellular biology*, v. 38, n. 20, e00309-18.
https://doi.org/10.1128/MCB.00309-18

Blatt, Michael R. (2015). "Vigilante science". *Plant physiology*, v. 169, n. 2, pp. 907-909.
https://doi.org/10.1104/pp.15.01443

Bosch, Xavier; **Hernández, Cristina**; **Pericas, Juan M.**; **Doti, Pamela**; **Marušić, Ana** (2012). "Misconduct policies in high-impact biomedical journals". *PloS one*, v. 7, n. 12, e51928.
https://doi.org/10.1371/journal.pone.0051928

Brookes, Paul S. (2014). "Internet publicity of data problems in the bioscience literature correlates with enhanced corrective action". *PeerJ*, n. 2, e313.
https://doi.org/10.7717/peerj.313

Budd, John M; **Sievert, Mary Ellen**; **Schultz, Tom R.** (1998). "Phenomena of retraction: Reasons for retraction and citations to the publications". *Journal of the American Medical Association, v. 280, n.* 3, pp. 296-297.
https://doi.org/10.1001/jama.280.3.296

Budd, John M.; **Sievert, Mary-Ellen**; **Schultz, Tom R.**; **Scoville, Caryn** (1999). "Effects of article retraction on citation and practice in medicine". *Bulletin of the Medical Library Association*, v. 87, n. 4, pp. 437-443.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC226618/

Campos-Varela, Isabel; **Villaverde-Castañeda, Ramón**; **Ruano-Raviña, Alberto** (2021). "Retraction of publications: a study of biomedical journals retracting publications based on impact factor and journal category". *Gaceta sanitaria*, n. 34, pp. 430-434.
https://doi.org/10.1016/j.gaceta.2019.05.008

Cokol, Murat; **Iossifov, Ivan**; **Rodríguez-Esteban, Raúl**; **Rzhetsky, Andrey** (2007). "How many scientific papers should be retracted?". *EMBO reports*, v. 8, n. 5, pp. 422-423.
https://doi.org/10.1038/sj.embor.7400970

Cokol, Murat; **Ozbay, Fatih**; **Rodríguez-Esteban, Raúl** (2008). "Retraction rates are on the rise". *EMBO reports*, v. 9, n. 1, p. 2.
https://doi.org/10.1038/sj.embor.7401143

Corbyn, Zoë (2012). "Misconduct is the main cause of life-sciences retractions". *Nature news*, n. 490, 21.
https://doi.org/10.1038/490021a

Da-Silva, Jaime A. Teixeira (2018a). "The issue of comment ownership and copyright at *PubPeer*". *Journal of educational media & library sciences*, v. 55, n. 2, pp. 181-191.
http://doi.org/10.6120/JoEMLS.201807_55(2).e001.BC.BE

Da-Silva, Jaime A. Teixeira (2018b). "PubMed Commons closure: a step back in post-publication peer review". *AME medical journal*, v. 30, n. 3.
https://doi.org/10.21037/amj.2018.02.07

Da-Silva, Jaime A. Teixeira (2022). "A synthesis of the formats for correcting erroneous and fraudulent academic literature, and associated challenges". *Journal for general philosophy of science* (in press).
https://doi.org/10.1007/s10838-022-09607-4

Decullier, Evelyne; **Huot, Laure**; **Samson, Géraldine**; **Maisonneuve, Hervé** (2013). "Visibility of retractions: a cross-sectional one-year study". *BioMed Central research notes*, n. 6, e238.
https://doi.org/10.1186/1756-0500-6-238

De-Moya-Anegón, Félix (2020). "Research evaluation entities cause a shift of publication to Q1 journals". *Profesional de la información*, v. 29, n. 4, e290431.
https://doi.org/10.3145/epi.2020.jul.31

Elia, Nadia; **Wager, Elizabeth**; **Tramèr, Martin R.** (2014). "Fate of articles that warranted retraction due to ethical concerns: A descriptive cross-sectional study". *PLoS one*, v. 9, n. 1, e85846.
https://doi.org/10.1371/journal.pone.0085846

Faggion Jr., Clovis-Mariano; **Ware, Robert S.**; **Bakas, Nikolaos**; **Wasiak, Jason** (2018). "An analysis of retractions of dental publications". *Journal of dentistry*, n. 79, pp. 19-23.
https://doi.org/10.1016/j.jdent.2018.09.002

Fang, Ferric C.; **Steen, R. Grant**; **Casadevall, Arturo** (2012). "Misconduct accounts for the majority of retracted scientific publications". *Proceedings of the National Academy of Sciences*, v. 109, n. 42, pp. 17028-17033.
https://doi.org/10.1073/pnas.1212247109

Gasparyan, Armen-Yuri; **Ayvazyan, Lilit**; **Akazhanov, Nurbek A.**; **Kitas, George D.** (2014). "Self-correction in biomedical publications and the scientific impact". *Croatian medical journal*, v. 55, n. 1, pp. 61-72.
https://doi.org/10.3325/cmj.2014.55.61

Lei, Lei; **Zhang, Ying** (2018). "Lack of improvement in scientific integrity: An analysis of WoS retractions by Chinese researchers (1997-2016)". *Science and engineering ethics*, v. 24, n. 5, pp. 1409-1420.
https://doi.org/10.1007/s11948-017-9962-7

Martin, Ben R. (2012). "*Does peer review work as a self-policing mechanism in preventing misconduct: a case study of a serial plagiarist".* In: Mayer, Tony; Steneck, Nicholas. *Promoting research integrity in a global environment*, pp. 97-114. ISBN: 978 9814340977
http://sro.sussex.ac.uk/id/eprint/18580

Marusic, Ana; **Katavic, Vedran**; **Marusic, Matko** (2007). "Role of editors and journals in detecting and preventing scientific misconduct: strengths, weaknesses, opportunities, and threats". *Medicine and law*, n. 26, pp. 545-566.
https://heinonline.org/HOL/LandingPage?handle=hein.journals/mlv26&div=51

Moylan, Elizabeth C.; **Kowalczuk, Maria K.** (2016). "Why articles are retracted: a retrospective cross-sectional study of retraction notices at BioMed Central". *British Medical Journal open*, v. 6, n. 11, e012047.
https://doi.org/10.1136/bmjopen-2016-012047

Nath, Sara B.; **Marcus, Steven C.**; **Druss, Benjamin G.** (2006). "Retractions in the research literature: misconduct or mistakes?" *Medical journal of Australia*, v. 185, n. 3, pp. 152-154.
https://doi.org/10.5694/j.1326-5377.2006.tb00504.x

Neale, Anne-Victoria; **Northrup, Justin**; **Dailey, Rhonda**; **Marks, Ellen**; **Abrams, Judit** (2007). "Correction and use of biomedical literature affected by scientific misconduct". *Science and engineering ethics*, v. 13, n. 1, pp. 5-24.
https://doi.org/10.1007/s11948-006-0003-1

Ortega, José-Luis (2021). "The relationship and incidence of three editorial notices in *PubPeer*: Errata, expressions of concern, and retractions". *Learned publishing*, v. 34, n. 2, pp. 164-174.
https://doi.org/10.1002/leap.1339

Ortega, José-Luis (2022). "Classification and analysis of *PubPeer* comments: How a web journal club is used". *Journal of the Association for Information Science and Technology*, v. 73, n. 5, pp. 650-670.
https://doi.org/10.1002/asi.24568

PubPeer (2015). *Vigilant scientists*. *PubPeer*.
https://blog.pubpeer.com/publications/B6CF3DB974A8ECC64B1A0303BBCD6F#16

Redman, Barbara K.; **Yarandi, Hossein N.**; **Merz, Jon F.** (2008). "Empirical developments in retraction". *Journal of medical ethics*, v. 34, n. 11, pp. 807-809.
https://doi.org/10.1136/jme.2007.023069

Resnik, David B.; **Dinse, Gregg E.** (2013). "Scientific retractions and corrections related to misconduct findings". *Journal of medical ethics*, v. 39, n. 1, pp. 46-50.
https://doi.org/10.1136/medethics-2012-100766

Smith, Jane; **Godlee, Fiona** (2005). "Investigating allegations of scientific misconduct". *British medical journal*, v. 331, n. 7511, pp. 245-246.
https://doi.org/10.1136/bmj.331.7511.245

Steen, R. Grant (2011). "Retractions in the scientific literature: do authors deliberately commit research fraud?". *Journal of medical ethics*, v. 37, n. 2, pp. 113-117.
https://doi.org/10.1136/jme.2010.038125

Steen, R. Grant; **Casadevall, Arturo**; **Fang, Ferric C.** (2013). "Why has the number of scientific retractions increased?". *PloS one*, v. 8, n. 7, e68397.
https://doi.org/10.1371/journal.pone.0068397

Stricker, Johanes; **Günther, Armin** (2019). "Scientific misconduct in psychology". *Zeitschrift für psychologie, v. 227, n.* 1, pp. 53-63.
https://doi.org/10.1027/2151-2604/a000356

Torny, Didier (2018). "*PubPeer*: Vigilante science, journal club or alarm raiser? The controversies over anonymity in post-publication peer review". In: *Peere international conference on peer review*, Peerecost network, Mar 2018, Rome, Italy.
https://shs.hal.science/halshs-01700198

Tripathi, Manorama; **Sonkar, Sharad Kumar**; **Kumar, Sunil** (2019). "A cross sectional study of retraction notices of scholarly journals of science". *Desidoc journal of library & information technology*, v. 39, n. 2, pp. 74-81.
https://doi.org/10.14429/djlit.39.2.14000

Vuong, Quan-Hoang (2020). "The limitations of retraction notices and the heroic acts of authors who correct the scholarly record: An analysis of retractions of papers published from 1975 to 2019". *Learned publishing*, v. 33, n. 2, pp. 119-130.
https://doi.org/10.1002/leap.1282

Wager, Elizabeth (2007). "What do journal editors do when they suspect research misconduct". *Medicine and law*, v. 26, n. 3, pp. 535.
https://pubmed.ncbi.nlm.nih.gov/17970251

Wager, Elizabeth (2014). "How should journal editors respond to cases of suspected misconduct?". *Journal of microbiology & biology education*, v. 15, n. 2, pp. 146-150.
https://doi.org/10.1128/jmbe.v15i2.829

Wager, Elizabeth (2015). "Why are retractions so difficult?". *Science editing*, v. 2, n. 1, pp. 32-34.
https://doi.org/10.6087/kcse.34

Wager, Elizabeth; **Veitch, Emma** (2017). "The role of *PubPeer* comments in alerting editors to serious problems with clinical research publications". In: *8th International congress on peer review and scientific publication*, Chicago.
https://peerreviewcongress.org/abstract/the-role-of-pubpeer-comments-in-alerting-editors-to-serious-problems-with-clinical-research-publications

Wager, Elizabeth; **Williams, Peter** (2011). "Why and how do journals retract articles? An analysis of Medline retractions 1988-2008". *Journal of medical ethics*, v. 37, n. 9, pp. 567-570.
https://doi.org/10.1136/jme.2010.040964