

RESEARCH

Open Access



Identifying common patterns in journals that retracted papers from paper mills: a cross-sectional study

Noa Mascato Fontaiña¹, Cristina Candal-Pedreira^{1,2,3*} , Guadalupe García^{1,2}, Joseph S. Ross^{4,5}, Alberto Ruano-Ravina^{1,2,3} and Lucía Martín-Gisbert^{1,2,6}

Abstract

Objectives To characterize journals that published and retracted articles retracted for having originated from paper mills and examine associations between paper mill retraction frequency and journal characteristics.

Methods Retraction Watch database was used to identify papers retracted due to originating from paper mills and journals, between January 2020 and December 2022. Data on the total number of articles and journal characteristics were obtained from Web of Science and Journal Citation Reports. Journals were classified based on the frequency of retracted paper mill papers (1, 2–9, ≥ 10 retractions). Logistic regressions were conducted to explore associations between retraction frequency and journal characteristics.

Results One hundred forty-two journals were identified that retracted 2,051 articles from paper mills. Among these, 71 (50%) journals had 1 retraction, 36 (25.4%) had 2–9 retractions, and 35 (24.6%) had ≥ 10 retractions; 4 (2.8%) journals had > 100 retractions. These journals, regardless of paper mill retraction number, were mainly in the second (35.2%) and third (29.6%) quartiles by impact factor. Medicine and health emerged as the predominant subject area, comprising 61.2% of all indexed journal categories. Comparing journals with one retraction to those with ten or more, the proportion of open access articles (72.6% vs. 19.2%) and median editorial times (86 vs. 116 days) differed across groups, although these differences were not statistically significant. An inverse correlation was observed between the proportion of paper mill papers and original articles (Spearman's $Rho = -0.1891$, 95%CI -0.370 to -0.008). Logistic regressions found no significant association between paper mill retraction number and other variables.

Conclusion This study suggests that paper mill retractions are concentrated in a small number of journals with common characteristics: high open access rates, intermediate impact factor quartiles, a high volume of citable items, and classification in medicine and health categories. Short editorial times may indicate a higher presence of paper mill publications, but more research is needed to examine this factor in depth, as well as the possible influence of acceptance rates.

Keywords Paper mills, Retractions, Academic journals, Publication ethics

*Correspondence:

Cristina Candal-Pedreira
c.candal.pedreira@usc.es

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Scientific literature and academic integrity are fundamental pillars for the advancement of knowledge. In recent years, however, this has been threatened by the surge in so-called “paper mills”, organizations devoted to mass production of fraudulent or plagiarized scientific papers, then sold to researchers to publish under their names [1, 2]. The “publish or perish” axiom, the need to publish for graduation purposes and for career progression, along with financial and promotion incentives have led to a situation where researchers resort to buying the authorship of these paper mill papers to secure paper authorship, which are delivered by the paper mills [1, 3]. There is evidence to show that, in addition to offering services such as ghost writing, intermediation of authorship, and the falsification and plagiarism of data and figures [4], paper mills manipulate the peer review process, supplant real reviewers, and are sometimes associated with editorial teams to ensure faster, and easier, publication [5–7].

Some of the greatest damage caused by these organizations is generated when false research infiltrates the research enterprise, such as by being aggregated into systematic reviews and meta-analyses, falsifying scientific evidence and potentially affecting clinical practice. This amounts to a waste of time and valuable resources, induces dangers for public health and medicine, and can even put the lives of patients at risk [8–10]. The paper mill phenomenon has experienced substantial growth in recent years and is envisaged to continue growing [3, 11–13]. A report by the Committee on Publication Ethics (COPE), based on data from six publishers covering over 53,000 submitted manuscripts across various disciplines, found that the proportion of suspected papers ranged from 2 to 46%, depending on the journal’s level of targeting and vulnerability to such practices [1]. While these figures are not representative of the entire publishing landscape, they illustrate the potentially widespread and uneven impact of paper mills on scientific publishing.

Previous investigations have identified certain characteristics related to paper mills [14], but there is still little information about some characteristics of journals that publish papers from paper mills, such as number of editors or published papers. However, understanding the characteristics and features of paper mills, including the journals in which they publish, may provide valuable insights and guide the development of strategies to identify such studies. Our aim was therefore to characterize journals that had retracted papers having originated from paper mills (hereafter paper mill papers) and examine associations between paper mill retraction frequency and journal characteristics, such as impact factor, number of editors, areas of knowledge, number of papers published

and open access, among others. With a better understanding of the nature and scope of this problem, more effective strategies could be developed to protect the integrity of scientific publication and preserve the quality of scientific evidence.

Methods

This paper reports on a cross-sectional study conducted to investigate journals that retracted at least one paper originating from paper mills. We included all journals that had at least one paper mill paper retracted over the period January 1st, 2020, through December 31st, 2022. Corrections, expressions of concern, and communications to conferences were excluded.

We used the Retraction Watch database to identify retracted paper mill papers and the journals in which they were published (last access to the database January 23rd, 2023). A number of journal characteristics were collected as follows. First, the total number of papers published by each journal from 2020 through 2022 was extracted from Web of Science. Next, the following information was sourced from Journal Citation Reports (JCR): total of citable items; journal impact factor (JIF); quartile (Q); percentage of papers published by the journal, with respect to the total of citable items; percentage of publications in open access; and the JCR categories of each journal. Because the journals possessed more than one category, we recorded all categories belonging to the journals analyzed: the JIF and quartile selected were the most favorable one of the categories. The date of the last search of the JCR database was October 21, 2023. Finally, official journal websites were searched to obtain information on the country, publisher, total number of editors (excluding senior and honorary editors), editorial times (from submission until first and second editorial decisions), and acceptance rate. These data are reported by the journals themselves.

Statistical analysis

The journals included were classified into three groups based on the number of retracted paper mill papers that they published over the period analysed, using the 50th and 75th percentiles of the entire sample as cut-off points. Journals that had published only one retracted paper mill paper were classified in Group 1, journals with between 2 and 9 papers in Group 2, and those with 10 or more papers in Group 3.

First, we performed a descriptive analysis of the journals, divided into groups according to the following qualitative variables: publisher, country, quartile and category. For each variable, we recorded the number of journals and the number of retracted paper mill papers published. Additionally, a descriptive analysis of the

journals categorised by group was performed for the following quantitative variables: number of paper mill articles; total citable articles; percentage of open access; JIF; number of JCR subject categories assigned to the journal; percentage of articles out of the total citable articles; and number of editors. For each of the quantitative variables, the median, interquartile range, minimum and maximum values were obtained according to the group. Quantitative variables were compared between groups using analysis of variance (ANOVA), while qualitative variables were compared using the chi-square test. Where differences were indicated by overall comparisons between the three groups, multiple comparison analyses were conducted between pairs of groups using the Bonferroni correction.

The descriptive analysis was replicated in the 20 journals with the most retracted paper mill papers.

The correlation between the quantitative variables and the percentage of retracted paper mill papers was assessed. To generate the percentage of retracted paper mill papers for each analyzed journal, the variable 'number of retracted paper mill papers' was standardized by dividing it by the total number of citable articles published by each journal. This variable gives the proportion of paper mill papers in a journal out of the total number of citable articles published by that journal. Citable articles, also called citable items, are used to calculate the JCR impact factor and include original articles and reviews. This variable facilitates the assessment of the relative magnitude of paper mill penetration in the analysed journals. The Spearman's Rho is presented for each comparison, along with the 95% confidence interval and the p-value.

An ordinal logistic regression model was performed to examine the relationship between the dependent variable (Group, with three ordered levels) and several independent variables. This approach allowed us to identify whether the independent variables were associated with the likelihood of a journal being classified in Groups 1, 2 or 3. Due to multicollinearity among the independent variables, a principal component analysis (PCA) was conducted to identify combinations of variables that explained the largest variance in the data. Those components with eigenvalues equal or greater than 1 were included. The included independent variables were the total of citable items, JIF, percentage of original papers published by the journal and number of JCR categories.

Additionally, a binary logistic regression was performed. To do this, the variable 'Group' was recategorized into a dichotomous variable with a cut-off point set at ≥ 10 retracted paper mill papers. The dependent variable was the dichotomous variable (having published < 10 paper mill papers / ≥ 10 paper mill papers) and the

independent variables were the total of citable items, JIF, percentage of original papers published by the journal and number of JCR categories.

All statistical analyses were performed using the STATA v.16 software package. Statistical significance was set at a p -value < 0.05 .

Due to the nature of the study, no ethics committee approval was required.

Results

Description of journals categorized by group

The study included 2,051 papers retracted for originating from paper mills from January 1st, 2020, through December 31st, 2022, published in 142 different journals. Of these, 71 journals (50.0%) were classified in Group 1, 36 (25.4%) in Group 2, and 35 (24.6%) in Group 3. A total of 93.7% were indexed in JCR; the 9 unindexed journals were registered as missing values, with one belonging to Group 1 (1 retracted paper mill paper), three to Group 2 (2–9 retracted paper mill papers), and the remaining five to Group 3 (10 or more retracted paper mill papers). Group 3 contained most retractions (89.4%).

Among all the indexed journals included, journals most commonly were based in the United States (USA) (Supplementary Table 1), with a total of 42 journals (29.6%), followed by the United Kingdom, with 35 (24.6%). Table 1 shows the distribution of journals by publisher and the number of paper mill-related retractions. Eight main publishers accounted for 65.5% of the journals with most retracted paper mill papers and 42.8% of total retractions. Most of the indexed journals' impact factors were in Q2 (35.2%) and Q3 (29.6%) (Table 2).

A total of 67 categories were identified in all the JCR-indexed journals, and of these, 41 (61.2%) were related to medicine and health. The remaining categories were Engineering and Technology (10.4%), Economic and Social Sciences (10.4%), Chemistry (6.0%), Physics (4.5%), Material Sciences (3.0%), and Education, and Biological and Environmental Sciences (1.5% each).

Table 3 lists the characteristics of the 20 journals with most paper mill-related retractions. Four journals exceeded 100 retractions. Ten journals accounted for 1,276 (62.2%) retracted papers. The *Journal of Physics: Conference Series* led with 494 retractions, followed by the *European Review for Medical and Pharmacological Sciences* with 165. All the journals with high retraction rates had an open access rate above 60%. After mitigating the effect of the journal's publication volume, the *International Journal of Immunopathology and Pharmacology* was observed to have the greatest density of paper mill-related retractions, with 8.34 per 1,000 citable items, followed by the *European Review for Medical and*

Table 1 Description of the number of journals and number of retracted paper mill papers, by publisher

Publisher	Group 1 (N=71)		Group 2 (N=36)		Group 3 (N=35)		Total Journals (PPM)
	Journals (%)	PPM	Journals (%)	PPM	Journals (%)	PPM	
Elsevier	13 (13)	25	9 (44)	52	4 (135)	547	26 (192)
Springer	13 (13)	21	5 (22)	38	0 (0)	402	18 (35)
Wiley	7 (7)	18	3 (13)	25	5 (216)	141	15 (236)
Springer—Nature Publishing Group	5 (5)	6	3 (6)	6	0 (0)	0	8 (11)
Springer—Biomed Central (BMC)	4 (4)	4	4 (13)	17	0 (0)	0	8 (17)
SAGE Publications	2 (2)	2	1 (3)	3	3 (163)	163	6 (168)
Taylor and Francis—Dove Press	3 (3)	3	1 (4)	4	2 (62)	62	6 (69)
Spandidos	0 (0)	0	0 (0)	0	6 (210)	210	6 (210)
Frontiers	4 (4)	4	1 (8)	8	0 (0)	0	5 (12)
Taylor and Francis	2 (2)	2	0 (0)	0	2 (66)	66	4 (68)
Karger	1 (1)	1	2 (4)	4	0 (0)	0	3 (5)
e-Century Publishing Corporation	0 (0)	0	2 (5)	5	1 (10)	10	3 (15)
Mary Ann Liebert	1 (1)	1	1 (2)	2	1 (18)	18	3 (21)
Royal Society of Chemistry (RSC)	2 (2)	2	0 (0)	0	1 (70)	70	3 (72)
Hindawi	2 (2)	2	0 (0)	0	0 (0)	0	2 (2)
IOS Press	0 (0)	0	1 (6)	6	1 (11)	11	2 (17)
De Gruyter	2 (2)	2	0 (0)	0	0 (0)	0	2 (2)
IOP Publishing	1 (1)	1	0 (0)	0	1 (494)	494	2 (495)
Impact Journals	1 (1)	1	1 (4)	4	0 (0)	0	2 (5)
Elsevier—Cell Press	0 (0)	0	0 (0)	0	1 (21)	21	1 (21)
Assoc. Brasileira de Divulgação Científica	0 (0)	0	1 (9)	9	0 (0)	0	1 (9)
Cellular Physiol Biochem Press	0 (0)	0	0 (0)	0	1 (64)	64	1 (64)
BAKIS Productions LTD	0 (0)	0	0 (0)	0	1 (23)	23	1 (23)
International Scientific Information, Inc	0 (0)	0	1 (3)	3	0 (0)	0	1 (3)
Ingenta	0 (0)	0	0 (0)	0	1 (12)	12	1 (12)
International Assoc. of Online Engineering	0 (0)	0	0 (0)	0	1 (29)	29	1 (29)
PLoS	0 (0)	0	0 (0)	0	1 (17)	17	1 (17)
Verduci Editore	0 (0)	0	0 (0)	0	1 (165)	165	1 (165)
Portland Press	0 (0)	0	0 (0)	0	1 (48)	48	1 (48)
Otras (N=8)	8 (8)	8	0 (0)	0	0 (0)	0	8 (8)

Group 1: one retracted paper mill paper. Group 2: 2–9 retracted paper mill papers. Group 3: 10 or more retracted paper mill papers
 The classification "others" includes all publishers having only one journal with a single-mill paper published. PPM = retracted paper mill papers

Table 2 Description of the number of journals and number of retracted paper mill papers by quartile

Quartile	Group 1 (N=71)		Group 2 (N=36)		Group 3 (N=35)		Total Journals (%)
	Journals (%)	PPM	Journals (%)	PPM	Journals (%)	PPM	
Q2	25 (35.2%)	25	11 (30.6%)	52	14 (40%)	547	50 (35.2%)
Q3	21 (29.6%)	21	11 (30.6%)	38	10 (28.6%)	402	42 (29.6%)
Q1	18 (25.4%)	18	6 (16.7%)	25	5 (14.3%)	141	29 (20.4%)
Q4	5 (7.0%)	5	4 (11.1%)	17	1 (2.9%)	12	10 (7.0%)
Quartile unknown	2 (2.8%)	2	4 (11.1%)	14	5 (14.3%)	732	11 (7.7%)

PPM = retracted paper mill papers, Q = Quartile
 Group 1: one retracted paper mill paper. Group 2: 2–9 retracted paper mill papers. Group 3: 10 or more retracted paper mill papers

Table 3 (continued)

PPM	Journal	Publisher	Country	Category	Nº of categories	JIF 2022	Quartile	Citable items	Open access (%)	% Of Articles	Total publishers	2ª Editorial decision
28	International Journal of Immunopathology and Pharmacology	SAGE Publications	Italy	Immunology; Pathology; Pharmacology & Pharmacy	3	3.5	Q2/Q3	3359	92.17%	94.19%	25	-
25	Cancer Management and Research	Taylor and Francis - Dove Press	UK	Oncology	1	3.3	Q3	13689	97.36%	77.30%	32	73
25	Oncology Letters	Spandidos	Greece	Oncology	1	2.9	Q3	28011	98.18%	84.33%	603	-
23	Journal of Balkan Union of Oncology****	BAKIS Productions LTD	-	-	-	-	-	-	-	-	-	-

PPM= Paper Mills; *Last record in JCR: 2005; **Last record in JCR:2020; ***Last record in JCR: 2007; **** Last record in JCR:2021

Pharmacological Sciences (7.23) and *Cancer Biotherapy & Radiopharmaceuticals* (6.30) (Fig. 1).

Table 4 shows the distribution of the journals by group, by reference to the quantitative variables. A gradual increase was observed in the median of citable items as the group grew larger (Fig. 2c) (p -value = 0.41). The journals with most retracted paper mill papers published were mostly open access, with a median of 72.6%. In terms of JIF, 85.7% of the journals had an impact factor of less than 6. *Molecular Cancer* was the journal with the highest impact factor (37.3). The median impact factor remained at around 4 in all groups, and the median number of categories was 1 to 2 for the three groups.

While Group 1 journals (1 retracted paper mill paper) had a median of 65 editors, Groups 2 (2–9 retracted paper mill papers) and 3 (10 or more retracted paper mill papers) had a median of 59 (p -value = 0.88). The percentage of papers retracted paper mill papers, relative to the total of citable items, was lowest in Group 3 (85.8%) and highest in Group 2 (92.0%) (Fig. 2a). In terms of time until the second editorial decision, construed as the median number of days from a paper's submission until its formal acceptance, editorial time showed a possible tendency to become shorter as the Group became higher (Fig. 2b).

Correlation and regression analysis

Figure 3 shows the correlation matrix between the quantitative variables. There was an inverse correlation between the percentage of papermill papers and percentage of original papers published, with a Spearman's Rho of -0.1891 (95%CI -0.370 to -0.008 ; p -value: 0.0293). The correlation between the percentage of articles and the JIF, as well as with the number of editors, could be an effect of the standardization of the variable "paper mill papers". Supplementary Table 2 includes the correlation values.

The ordinal logistic regression found no significant association between the journals categorized into groups and the independent variables. Similarly, the binary logistic regression showed no significant association between the "high volume of paper mill papers retracted" and these variables. The results of both regressions can be consulted in Supplementary Table 3.

Discussion

Main findings

The results of this paper suggest that retracted paper mill papers tend to be grouped within a small number of journals. These journals are typically in intermediate quartiles, are mostly open access, and have an inverse correlation with the percentage of original articles published (i.e. the fewer original articles a journal publishes,

the more paper mill retractions it has). Four journals were identified that were particularly implicated, with over 100 retracted paper mill papers in a period of three years: *Journal of Physics: Conference Series*, *European Review for Medical and Pharmacological Sciences*, *Journal of Cellular Biochemistry* and *International Journal of Electrical Engineering and Education*. There is an urgent need to address the systemic problems that lead to these high retractions rates in certain journals and publishing houses.

The findings of this study indicate that most journals impacted by paper mill papers are from the United Kingdom and the USA and associated with a limited group of publishers, including Elsevier, Springer, and Wiley. These publishers have responded to calls from the Committee on Publishing Ethics (COPE), which has urged the academic publishing community to take coordinated action against paper mills. In response, they have introduced measures to detect problems in their manuscripts, such as training for editors or investing in suspicious-paper detection systems [11, 15–18]. Regarding the identification of the four journals with over 100 paper mill papers, most of the retractions in these journals, all currently relatively reputable, have been thanks to the work of sleuths (external researchers), who have publicly highlighted suspicious characteristics in the published studies [18–20].

Our results are in line with previous research and confirm that journals with retracted paper mill papers are predominantly linked to the field of health sciences [1, 8–10, 21]. Nevertheless, it highlights the fact that, though in a minority of cases, paper mills also affect other areas of knowledge, such as engineering, economic and social sciences, physics, chemistry, and even education, something that should not be ignored when it comes to studying this phenomenon.

Previous studies, such as that by Pérez-Neri et al. [10], reported a median impact factor of 3.119 for journals with retracted paper mill papers, and Qi et al. [16] observed that almost all the journals displayed an impact factor lower than 5. Furthermore, these journals generally tend to be positioned in Q2 [17]. The results of this study are in line with previous evidence, indicating that the majority of the journals analyzed belong to intermediate quartiles. Yet, the position in the quartiles does not vary in relation with the number of retractions, which suggests that impact factors do not seem to be clear indicators of the magnitude of paper mill-related retractions. This may be attributed to the complexity of the factors that influence such a retracted paper, as outlined in the journal's position. Firstly, it might be expected that journals of greatest impact, subjected to more rigorous scrutiny, would have a greater ability to detect these papers produced by paper mills, but the evidence suggests that

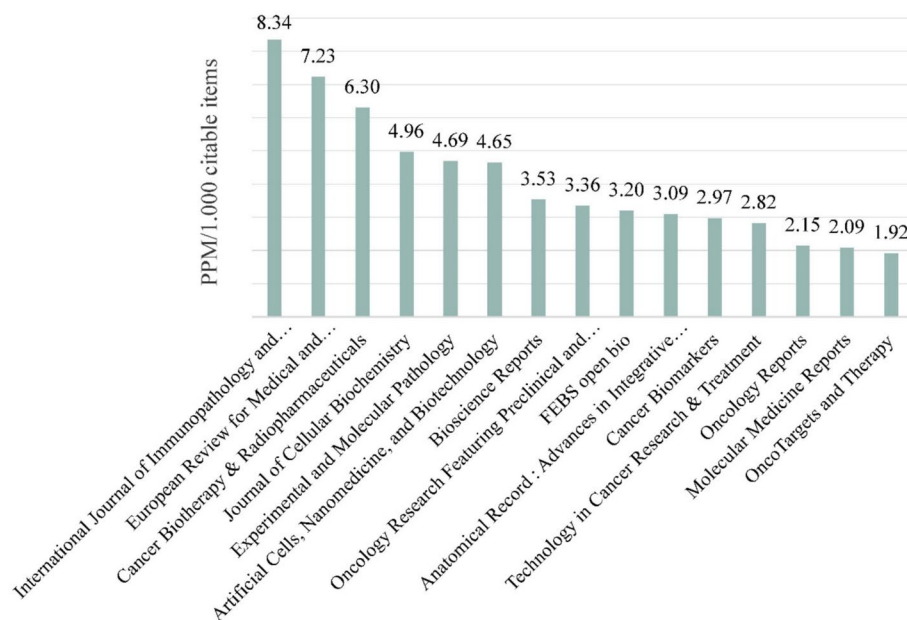


Fig. 1 Distribution of journals by density of retracted paper mill papers (number of retracted paper mill papers per 1,000 citable items)

they tend to register a lower number of retractions [22]. This may be because they do not publish a significant number of such papers in the first place. In contrast, journals with a lower impact factor may be more vulnerable to a greater volume of fraudulent papers submission, which could result in a higher proportion of paper mill papers being published and subsequently retracted [23].

The analysis by group suggests an accumulation of paper mill-related retractions in open access journals, though the publishing model does not seem to be a related factor. Even so, journals, known as predatory (those which prioritize profit over scientific quality by publishing low-standard content) have arisen in the last decades, taking advantage of the open access model [24–28]. These types of journals could afford paper mills as a means for publishing fraudulent content, which would account for the high publication rate of related retractions.

Special numbers are particularly vulnerable to paper mill papers because they are edited and published separately from the regular journal, usually by guest editors [15]. The former Hindawi publishing house withdrew thousands of papers in special editions, due to compromised peer review [29]. The results of our study show that the higher the proportion of content such as reviews, short communications, letters to the editor, and editorials (i.e., a lower number of original papers published), the higher the number of retracted paper mill papers that a journal will contain. This therefore gives rise to the hypothesis that journals which publish more special

numbers might be more susceptible to the introduction of false papers.

Although some studies suggest that a small number of editors may be representative of non-ethical practices [30], our study found no differences in the number of editors between journals with a single retraction and those with more than 10. It is, however, relevant to stress the surprising size of the editorial team in certain journals, which can rise to as many as 13,755 editors. The scientific community has warned of paper mills paying bribes to editors and including their own agents on editorial boards to facilitate the publication of their manuscripts [15]. Reputable publishers such as Wiley and Elsevier have already been involved in such matters [15, 29], which underscores the need for greater surveillance and regulation of editorial-team integrity, and poses the question of whether an editorial volume of such magnitude can be supervised in a way that will bolster ethical editorial practices.

Lastly, the data gathered could suggest a possible reduction in editorial times in the group with the highest number of retractions, indicating that these journals would be opting for swifter editorial processes and a potential dubious review processes [25, 30].

Advantages and limitations

This study has a series of limitations. Firstly, it was only possible to include formally retracted paper mill papers, which may underestimate the magnitude of the problem and complicate the identification of associations.

Table 4 Characteristics of journals analyzed by group

	N	Median	P25-P75	min-max
Group 1 (N=71)				
Number of PPM	71	1	1-1	1
Total of citable items	70	8,797.5	4,123-22,952	620-738,367
% of open access	70	19.2	10.3-97.55	1.4-100
JIF 2022	70	4	2.9-5.4	1.4-37.3
Number of categories	70	1.5	1-2	1-4
% of papers	70	91.5	81.4-96.1	47.4-100
Number of editors	70	65	45-115	4-10,111
Days until 2nd editorial decision	25	116	82-142	27-181
Group 2 (N=36)				
Number of PPM	36	3	2-6	2-9
Total of citable items	33	13,883	6,420-27,506	1,622-110,561
% of open access	33	22	10.97-92.55	0-99.7
JIF 2022	33	3.6	3.1-5	1.4-11.3
Number of categories	33	2	1-2	1-4
% of papers	33	92	84.12-96.42	68.5-100
Number of editors	33	59	46-86	9-13,755
Days until 2nd editorial decision	16	103.5	67-127.5	16-164
Group 3 (N=35)				
Number of PPM	35	25	16-54	10-494
Total of citable items	30	18,956	5,686-28,011	2,857-886,919
% open access	30	72.6	7.97-96.67	0-99.4
JIF 2022	30	4	3.3-5.6	2-16.2
Number of categories	30	1	1-2	1-4
% of papers	30	85.8	76.03-92.88	54.6-100
Number of editors	30	59	43-117	25-10,726
Days until 2nd editorial decision	9	86	73-116	25-159

PPM retracted paper mill papers, IQR interquartile range, JIF Journal Impact Factor (JCR), P25 25th percentile, P75 75th percentile

Group 1: one retracted paper mill paper. Group 2: 2-9 retracted paper mill papers. Group 3: 10 or more retracted paper mill papers

We acknowledge that the results of our study may not accurately reflect the actual scope of the phenomenon, as they depend directly on how scientific journals decide and manage to retract articles that have been identified as coming from paper mills. In other words, our findings are conditioned by the current state of retractions, which represents a partial and dynamic snapshot of a broader and evolving situation. Since many journals may still have fraudulent articles published that have not yet been detected, our results should be interpreted as provisional. These results may change as new detection tools are implemented and as journals continue their internal investigations.

Nonetheless, and precisely because of the dynamic nature of the phenomenon, we believe it is important to describe what is currently known. Therefore, studying the characteristics of the journals affected with the data available at this moment is relevant and necessary. Understanding the profile of the journals that have been targeted by these fraudulent practices allows us to identify patterns of vulnerability, generate hypotheses, and guide editorial policies and preventive mechanisms—even with the understanding that this portrayal is necessarily incomplete. Having a characterization, although subject to future revision, is preferable to inaction or to waiting for a perfect picture that we may never have.

Secondly, the content obtained from the JCR database may be biased due to the deindexation of journals which do not comply with ethical standards, underrepresented journals with a greater record of retractions. Furthermore, data relating to editorial times and the total number of editors are reported by the journals themselves, which could introduce inherent biases due to the omission of data and lack of independent verification.

Thirdly, while we acknowledge that the Retraction Watch database may contain omissions or errors due to limitations inherent to its data collection

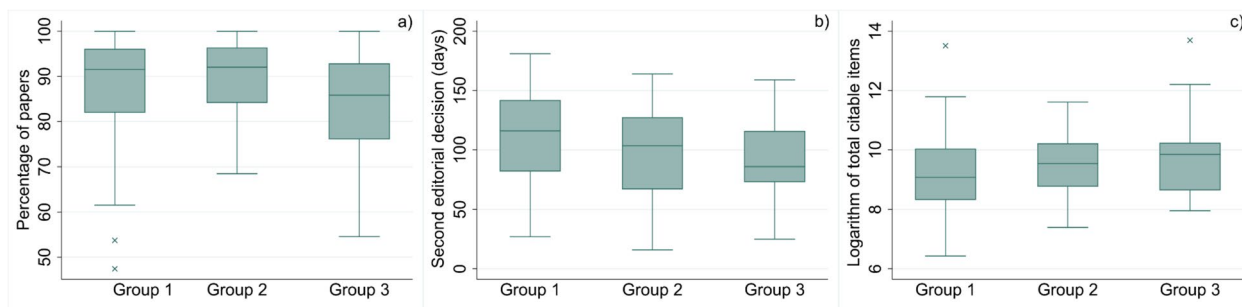


Fig. 2 Graphical representation of the median of (a) percentage of retracted paper mill papers, (b) days until second editorial decision, and (c) total citable items (expressed as logarithm), of journals categorized by group. The Boxplot depicts the interquartile range; the horizontal line in the Boxplot shows the median; the lower whisker indicates the lowest value excluding atypical values; and the upper whisker indicates the highest value. Atypical values are represented by crosses

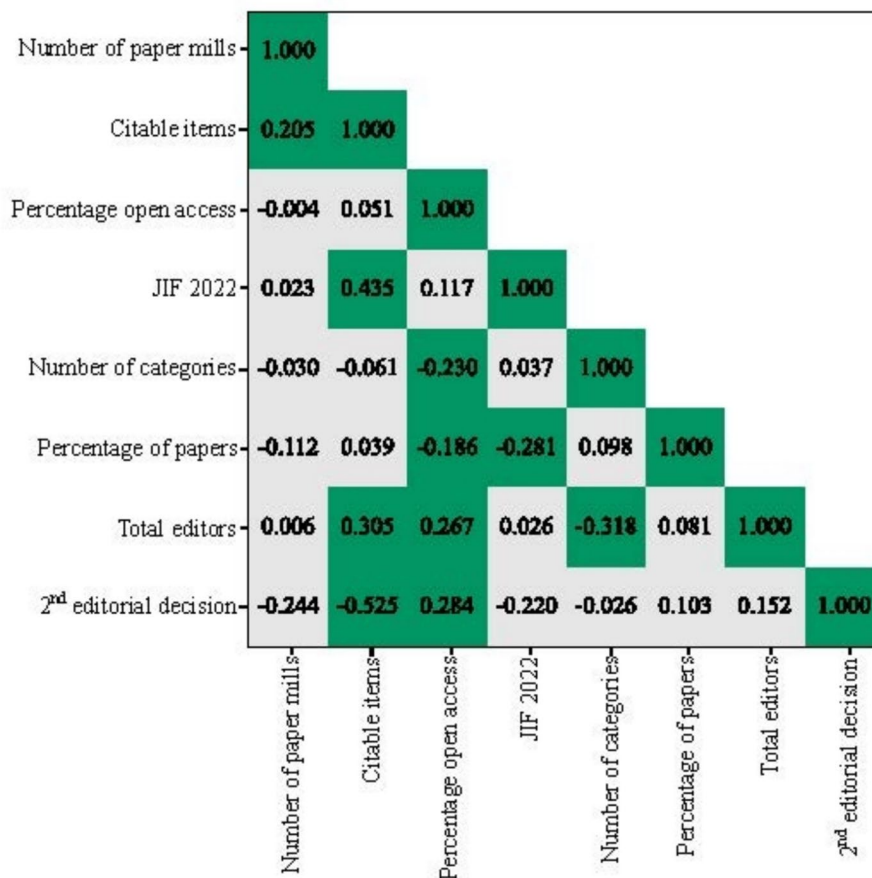


Fig. 3 Map of Spearman correlations: identification of statistically significant relationships. Green indicates statistically significant correlations; non-statistical correlations are shown in gray. Central numbers represent the value of the Spearman correlation coefficient (Rho)

methodology—for instance, its reliance on public reports and the lack of systematic practices in some editorial sources—we consider it, at present, to be the most comprehensive, accessible, and rigorous source available for systematically studying the phenomenon of retractions related to paper mills. Retraction Watch uses a variety of data sources to ascertain the reasons for retraction of the papers included in their database, including editorials, direct communication with journal editors, PubPeer commentaries, and investigations conducted by external experts. While we consider this approach to be the most robust currently available for identifying paper mill-related retractions, it remains possible that some studies were misclassified—either by being incorrectly labeled as paper mill products or by not being identified as such when they should have been. We also acknowledge that Retraction Watch might not be the best source to analyze in depth the characteristics of paper mill retractions, but we are not aware of a better source of information for this.

Furthermore, this study lacks a control group, which limits its capacity to make significant comparisons and establish solid conclusions about the distinctive characteristics. Future studies should consider the inclusion of a control group to evaluate whether characteristics differ. Additionally, the study would be enriched by a differentiated analysis of predatory journals to better understand their impact on the results. However, there is no objective and universally accepted tool to distinguish and classify journals as predatory or not. The problem is compounded by the possible lack of motivation of these journals to retract articles originating from paper mills, which means that certain journals are likely underrepresented in our dataset and may affect the results obtained. Finally, the study period of two years, with a minimum of 12 months after retraction, might be insufficient for the inclusion and follow-up of all retractions. The inclusion of a greater number of years would increase the robustness of the associations observed.

With respect to the advantages of the study, on the one hand, the use of acknowledged, comprehensive data

source, such as Retraction Watch, enables all retractions to be systematically included. Along with this, the use of JCR for gathering most of the study data ensures a solid basis for analysis. To our knowledge, previous studies have not exhaustively analyzed as many journal-related variables. Similarly, the use of a wide range of journals and their categorization into groups makes it possible to study differential characteristics by reference to paper mill penetration. To the best of our knowledge, no previous study has conducted an analysis of this nature, which suggests a novel contribution that may prove useful for future research.

Conclusions

The study reveals that retracted paper mill papers are concentrated in a small number of journals with certain common characteristics and suggests an inverse correlation between retractions and the percentage of original papers published. No evidence was found to show that the size of the editorial team might affect retractions. Editorial decision times should be investigated as a possible indicator of journals prone to retract papers. The results of this study provide a clear, detailed picture of the characteristics of journals affected by the problem of paper mills, and knowing these may be fundamental to improve the evaluation processes of journals and guide researchers in the selection of appropriate journals for submitting papers, thereby ensuring the integrity and quality of academic literature.

Abbreviations

COPE	Committee on Publication Ethics
JIF	Journal impact factor
JCR	Journal Citation Reports
PCA	Principal component analysis
Q	Quartile

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41073-025-00177-9>.

Supplementary Material 1.

Transparency declaration

The lead author states that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities
Not applicable.

Patient and public involvement

It was not appropriate or possible to involve patients or the public in the design, conduct, reporting or dissemination plans of our research.

Authors' contributions

NMF: Methodology, Formal analysis, Investigation, Data curation, Writing-Original Draft. CCP: Conceptualization, Methodology, Formal analysis, Writing-Review and Editing, Supervision. GG: Visualization, Writing-Review and Editing. JSR: Methodology, Writing-Review and Editing. ARR: Conceptualization, Methodology, Writing-Review and Editing, Supervision. LMG: Methodology, Visualization, Writing-Review and Editing.

Funding

This research received no external funding.

Data availability

Part of the data that support the findings of this study are available from Retraction Watch via CrossRef. The database can be accessed in the following link: <https://www.crossref.org/documentation/retrieve-metadata/retraction-watch/>.

Declarations

Ethics approval and consent to participate

Since this study used publicly available materials and did not involve human subjects, human subjects' ethics committee approval was not required. Patient consent was not required as no patients participated in this study.

Competing interests

Dr. Ross currently receives research support through Yale University from Johnson and Johnson to develop methods of clinical trial data sharing, from the Food and Drug Administration for the Yale-Mayo Clinic Center for Excellence in Regulatory Science and Innovation (CERSI) program (U01FD005938), from the Agency for Healthcare Research and Quality (R01HS022882), and from Arnold Ventures; formerly received research support from the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology (NEST) and from the National Heart, Lung and Blood Institute of the National Institutes of Health (NIH) (R01HS025164, R01HL144644); and in addition, Dr. Ross was an expert witness at the request of Relator's attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen Inc. that was settled September 2022. The other authors declare no conflicts of interest.

Author details

¹Department of Preventive Medicine and Public Health, Facultad de Medicina, C/San Francisco S/N Universidad de Santiago, de Compostela, 15782 Santiago de Compostela, Spain. ²Health Research Institute of Santiago de Compostela (Instituto de Investigación Sanitaria de Santiago de Compostela-IDIS), Santiago de Compostela, Galicia, Spain. ³Consortium for Biomedical Research in Epidemiology and Public Health (CIBER en Epidemiología y Salud Pública-CIBERESP), Madrid, Spain. ⁴Section of General Internal Medicine, Department of Medicine, Yale School of Medicine, Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA. ⁵Center for Outcomes Research and Evaluation, Yale-New Haven Health, New Haven, CT, USA. ⁶Cross-Disciplinary Research in Environmental Technologies (CRETUS), University of Santiago de Compostela, Santiago de Compostela, Spain.

Received: 24 January 2025 Accepted: 7 August 2025

Published online: 01 October 2025

References

- COPE & STM. Paper Mills - Research report from COPE & STM - English. 2022.
- Christopher J. The raw truth about paper mills. *Febs Lett.* 2021;595(13):1751–7.
- Byrne JA, Christopher J. Digital magic, or the dark arts of the 21(st) century-how can journals and peer reviewers detect manuscripts and publications from paper mills? *Febs Lett.* 2020;594(4):583–9.
- Perron B, Hiltz-Perron O, Bryan G. Revealed: The inner workings of a paper mill. *Retraction Watch.* 2021. Available from: <https://retractionwatch.com/2021/12/20/revealed-the-inner-workings-of-a-paper-mill/>.

5. Day A. Exploratory analysis of text duplication in peer-review reveals peer-review fraud and paper mills. *Scientometrics*. 2022;127(10):5965–87.
6. Mayta-Tristán P, Borja-García R. Malas prácticas en investigación: las fábricas de manuscritos en Perú. *Rev Peru Med Exp Salud Publica*. 2022;39:388–91.
7. Seifert R. How Naunyn-schmiedeberg's archives of pharmacology deals with fraudulent papers from paper mills. *Naunyn Schmiedebergs Arch Pharmacol*. 2021;394(3):431–6.
8. Byrne JA, Labbé C. Striking similarities between publications from China describing single gene knockdown experiments in human cancer cell lines. *Scientometrics*. 2017;110(3):1471–93.
9. COPE. COPE Forum 4 September 2020: paper mills | COPE: Committee on Publication Ethics.
10. Perez-Neri I, Pineda C, Sandoval H. Threats to scholarly research integrity arising from paper mills: a rapid scoping review. *Clin Rheumatol*. 2022;41(7):2241–8.
11. Bik EM, Casadevall A, Fang FC. The prevalence of inappropriate image duplication in biomedical research publications. *MBio*. 2016. <https://doi.org/10.1128/mBio.00809-16>.
12. Candal-Pedreira C, Ross JS, Ruano-Ravina A, Egilman DS, Fernandez E, Perez-Rios M. Retracted papers originating from paper mills: cross sectional study. *BMJ*. 2022;379:e071517.
13. Oransky I, Fremes SE, Kurlansky P, Gaudino M. Retractions in medicine: the tip of the iceberg. *Eur Heart J*. 2021;42(41):4205–6.
14. Parker L, Boughton S, Bero L, Byrne JA. Paper mill challenges: past, present, and future. *J Clin Epidemiol*. 2024;176:111549.
15. Joelving F. Retraction Watch. Paper trail. *Science*. 2024. Available from: <https://www.science.org/content/article/paper-mills-bringing-editors-scholarly-journals-science-investigation-finds>.
16. Qi X, Deng H, Guo X. Characteristics of retractions related to faked peer reviews: an overview. *Postgrad Med J*. 2017;93(1102):499–503.
17. Yang W, Sun N, Song H. Analysis of the retraction papers in oncology field from Chinese scholars from 2013 to 2022. *J Cancer Res Ther*. 2024;20(2):592–8.
18. Else H, Van Noorden R. The fight against fake-paper factories that churn out sham science. *Nature*. 2021;591(7851):516–9.
19. Kincaid E. Sage retracting three dozen articles for 'compromised' peer review. *Retraction Watch*. 2023. Available from: <https://retractionwatch.com/2023/07/21/sage-retracting-three-dozen-articles-for-compromised-peer-review/>.
20. Oransky I. Publisher retracts 350 papers at once. *Retraction Watch*. 2022. Available from: <https://retractionwatch.com/2022/02/23/publisher-retracts-350-papers-at-once/>.
21. Sebo P. Chinese authors are overrepresented in medical articles retracted for fake peer review or paper mill. *Intern Emerg Med*. 2024;19(8):2369–71.
22. Al-Ghareeb A, Hillel S, McKenna L, Cleary M, Visentin D, Jones M, et al. Retraction of publications in nursing and midwifery research: a systematic review. *Int J Nurs Stud*. 2018;81:8–13.
23. Bricker-Anthony C, Herzog RW. Distortion of journal impact factors in the era of paper mills. *Mol Ther*. 2023;31(6):1503–4.
24. Boumil MM, Salem DN. In ... and out: open access publishing in scientific journals. *Qual Manag Health Care*. 2014;23(3):133–7.
25. Dadkhah M, Rahimnia F, Darbyshire P, Borchardt G. Ten (bad) reasons researchers publish their papers in hijacked journals. *J Clin Nurs*. 2021;30(19–20):e60–3.
26. Likis FE. Predatory publishing: the threat continues. *J Midwifery Womens Health*. 2019;64(5):523–5.
27. Richtig G, Berger M, Lange-Asschenfeldt B, Aberer W, Richtig E. Problems and challenges of predatory journals. *J Eur Acad Dermatol Venereol*. 2018;32(9):1441–9.
28. Wang T, Xing QR, Wang H, Chen W. Retracted publications in the biomedical literature from open access journals. *Sci Eng Ethics*. 2019;25(3):855–68.
29. Van Noorden R. More than 10,000 research papers were retracted in 2023 - a new record. *Nature*. 2023;624(7992):479–81.
30. Dadkhah M, Bianciardi G. Ranking predatory journals: solve the problem instead of removing it! *Adv Pharm Bull*. 2016;6(1):1–4.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.