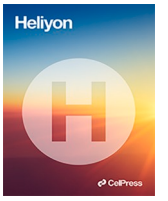
Heliyon 10 (2024) e38620

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Retracted articles in scientific literature: A bibliometric analysis from 2003 to 2022 using the Web of Science

Malcolm Koo [a,b], Shih-Chun Lin [c,d,*]

[a] Department of Nursing, Tzu Chi University, Hualien City, Hualien, 970302, Taiwan
[b] Dalla Lana School of Public Health, University of Toronto, Toronto, ON, M5T 3M7, Canada
[c] Department of Nursing, Dalin Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Dalin, Chiayi, 62247, Taiwan
[d] Graduate Institute of Nursing, National Taipei University of Nursing and Health Sciences, Taipei City, 365, Taiwan

ARTICLE INFO

Keywords:
Retractions
Retraction of publication
Scientific misconduct
Publishing
Bibliometric analysis
Bibliometrix

ABSTRACT

Retractions serve a crucial role in maintaining the integrity and accuracy of scientific literature. There has been growing interest in understanding the patterns behind retractions. This bibliometric study analyzed retracted articles published between 2003 and 2022, indexed by the Science Citation Index Expanded of the Web of Science Core Collection database. A total of 8466 retracted articles were identified, revealing an overall increase up to 2019, followed by a decline. A total of 109 countries contributed to the retracted articles, with China and the United States having the highest absolute numbers. In addition, the articles were published in 2347 different journals, with Tumor Biology recording the largest number of retracted articles. The top 10 most cited retracted articles indicated that data and image integrity issues were the main reasons for retraction. The primary reasons for retractions, identified by linking the retracted articles to the Retraction Watch Database, were data and results issues followed by plagiarism and duplication. In conclusion, the present bibliometric study offered an overview of the status of retracted articles indexed by the Web of Science Core Collection over the past two decades. These findings provide insight into areas where scientific integrity may be compromised and serve as a guide to foster a responsible research environment.

1. Introduction

Ensuring the integrity and accuracy of published findings is of paramount importance for maintaining the credibility and reliability of scientific knowledge [1]. Retractions serve as a formal mechanism within the scientific publishing process, designed to correct or remove research articles identified as containing significant flaws, inaccuracies, or instances of misconduct. Retractions uphold the integrity of the scientific literature by addressing issues such as data fabrication, plagiarism, methodological errors, duplicate publication, or ethical breaches.

A retraction may be initiated by the authors of the papers or their institution if they discover errors, inaccuracies, or misconduct in their published work. In such cases, they are expected to contact the journal editor, provide an explanation of the issue, and request a retraction or correction. In addition, when journal editors become aware of potential issues in a published article through notifications from readers, reviewers, or their own detection methods, they can conduct preliminary investigations and contact the authors to

* Corresponding author. Department of Nursing, Dalin Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Dalin, Chiayi, 62247, Taiwan.
E-mail address: df559699@tzuchi.com.tw (S.-C. Lin).

https://doi.org/10.1016/j.heliyon.2024.e38620
Received 11 September 2023; Received in revised form 25 September 2024; Accepted 26 September 2024
Available online 26 September 2024

address the concerns. If a retraction is deemed necessary, the journal will issue a retraction notice, providing specific reasons for the retraction. The journal's online and print records will be updated to reflect the retraction and ensure that future readers are aware of the article's status [2,3].

While retractions can create substantial citation penalties [4] and may have negative implications for the authors involved, such as reputational damage or loss of credibility [5], they serve to prevent the dissemination of false or misleading information. This can help to avert potential negative impacts on future research, wasted resources, or misguided clinical practices [6]. An analysis of 180 retracted English-language papers published from 2000 to 2010, along with their 851 secondary studies citing them, showed that the retracted papers were cited over 5000 times. These retracted studies put over 428,000 enrolled participants and 79,690 at-risk patients [7]. Studies examining the nature of citations of retracted articles revealed a lack of awareness among investigators, journal editors, and peer reviewers about the perpetuation of retracted work [8,9].

The number of retracted scientific articles has been increasing in recent years. Some authors suggested that this trend indicates a rise in scientific misconduct. A review of 2047 biomedical and life science research articles indexed by PubMed as retracted on May 3, 2012, revealed that 67.4 % of retractions were attributable to misconduct, including fraud or suspected fraud (43.4 %), duplicate publication (14.2 %), and plagiarism (9.8 %) [10]. A comprehensive survey of retracted articles from 42 large bibliographic databases across major scholarly fields showed that the number of articles retracted per year increased by a factor of 19 from 2001 to 2010. The factor decreased to 11.4 when excluding repeat offenders and adjusting for the growth of the published literature [11].

However, other authors offered an alternative interpretation, suggesting that the rise in retractions could also result from growing scientific integrity and increasing efficiency of the detection system rather than a surge in scientific misconduct [6,12]. A comprehensive survey spanning 42 bibliographic databases and publisher websites across various scholarly disciplines identified 4449 scholarly publications retracted between 1928 and 2011. According to the authors, the proportion of published scholarly literature affected by retraction remained relatively small, and most retracted articles did not contain flawed data [11]. Overall, while the increasing number of retracted scientific articles is a concerning trend, it is essential to consider multiple possible explanations for the phenomenon.

There has been a growing interest in understanding the patterns and reasons behind retractions in recent years. This interest is driven by a desire to improve research integrity and the overall quality of scientific literature. Investigations into retractions have been conducted across various fields, including anesthesiology [13], cancer [14], cardiovascular disease [15], dentistry [16], neurosurgery [17], nursing [18,19], obstetrics and gynecology [20], orthopedy [21], and veterinary medicine [22].

In addition to systematic analysis [23], bibliometric methodology has emerged as an important tool to understand further the patterns and trends of scientific research [24]. Bibliometrics, as a field of research, applies quantitative methods to evaluate the impact of scientific publications, such as identifying highly cited papers, influential authors, and top-ranked journals [25]. Several studies have used bibliometric analysis to examine the characteristics of retractions in specific domains, such as COVID-19 [26], oncology [27], rehabilitation [28], and schizophrenia [29]. Nevertheless, there remains a need for a comprehensive analysis of retractions across multiple fields. To address this gap, the present study used bibliometric analysis to investigate the patterns of retractions in scientific research across various disciplines. Retracted articles published between 2003 and 2022 were identified using data from the Science Citation Index Expanded (SCI-Expanded) of the Web of Science (WoS) Core Collection database. Country-level, journal-level, article-level, and keyword co-occurrence analyses were conducted. Moreover, retraction reasons were analyzed by linking retracted articles to the Retraction Watch Database using Digital Object Identifiers (DOIs).

The objectives of this study are as follows: (1) To analyze the trends and patterns of retracted articles in scientific literature from 2003 to 2022, focusing on the distribution of retractions across various disciplines and journals, citation trends, and the geographic distribution; (2) To identify the primary reasons for article retractions during this period, with a particular emphasis on issues related to data integrity, plagiarism, and ethical concerns; (3) To provide insights into the factors contributing to retractions and to offer recommendations for improving research integrity and publication practices.

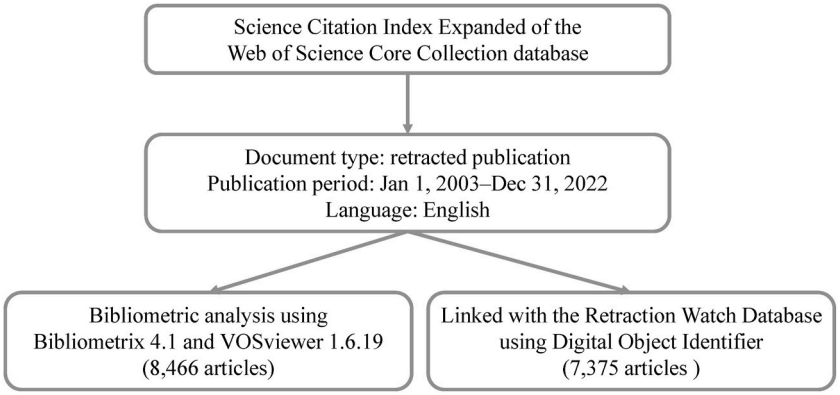


Fig. 1. Study flowchart.

2. Methods

The SCI-Expanded of the WoS Core Collection database (Clarivate Analytics, Philadelphia, PA, USA) was used to identify articles in this bibliometric analysis (Fig. 1). According to Clarivate, the SCI-Expanded indexes over 9500 of the world's most impactful journals across 178 scientific disciplines, with more than 53 million records dating from 1900 to the present [30]. While Scopus, a major commercial bibliographic database introduced by Elsevier in 2004 [31], is frequently utilized in bibliometric studies [32], our choice of WoS for the current study was driven by its unique inclusion of a dedicated document type for "retracted publication." This specific feature facilitates the straightforward identification and exclusion of retracted articles, thereby enhancing the reliability and accuracy of our search results.

The search was conducted on the same day (July 1, 2023) to avoid changes in the content due to updates in the database. For analysis, the document type "retracted publication" was used to identify retracted articles. In WoS, this document type indicates that the article in question has been formally withdrawn. Conversely, the document type "retraction" was not used because it represents merely a formal notice in WoS, which lacks data on citations and references.

The time frame of the search was limited to 2003 to 2022 to provide a two-decade snapshot of retractions in scientific literature. This allows for the identification of patterns in the retraction landscape while maintaining relevance to the current research environment. To ensure proper interpretation of the results, the publication language was restricted to English using the WoS field tag LA.

To conduct the bibliometric analysis, we utilized Bibliometrix 4.1 (Naples, Italy) [33] for country-level and journal-level analyses, and VOSviewer version 1.6.19 [34] for co-occurrence analysis of article keywords. Our keyword analysis used KeyWords Plus, which is a feature in WoS that augments author-generated keywords for a given article by adding additional index terms, thereby revealing more relevant terms potentially overlooked by the authors [35]. To maintain readability, only the top 200 terms were included to generate the co-occurrence network. In the network figure, the size of the circle reflects the number of articles in which the term occurs. The proximity between two linked terms indicates the relatedness of the terms based on their number of co-occurrences in documents.

To explore the reasons for retraction, the articles identified in the WoS were matched with the records in the Retraction Watch Database [36] by linking with DOIs. In the Retraction Watch Database, a total of 107 reasons for retraction were listed [37]. To improve interpretability, these reasons were reclassified into 12 categories: (1) Data and Results Issues, (2) Plagiarism and Duplication, (3) Investigations and Findings, (4) Authorship and Ethical Concerns, (5) Misconduct and Fraud, (6) Image Manipulation and Fabrication, (7) Peer Review and Editorial Issues, (8) Withdrawal and Retraction Notices, (9) Institutional and Policy Issues, (10) Complaints and Objections, (11) Miscellaneous, and (12) Procedural and Legal Issues (Table S1). In addition, as each retracted article can have multiple reasons for retraction, with no specific order of importance, the denominator for calculating the percentage of each retraction reason is the total number of retraction reasons rather than the number of articles.

To assess whether specific subject categories are more susceptible to certain types of retraction reasons, we focused on the three most prevalent retraction reasons—"Data and Results Issues," "Plagiarism and Duplication," and "Investigations and Findings" —along with their combinations. A contingency table was constructed to cross-tabulate the reasons for retraction with the main subject categories of the retracted articles. The subject category refers to the field of study most likely to be referenced or search for the information provided in the article. While many articles may overlap across different subject categories, only the most relevant subject category, as identified by the Retraction Watch Database, was assigned to each article in the present study. To determine the association between retraction reasons and subject categories, a Chi-square test was performed. Furthermore, post-hoc comparisons of column proportions were conducted using the Bonferroni correction to identify statistically significant differences between subject categories.

3. Results

During the 20-year period from 2003 to 2022, 8466 retracted articles were identified from the SCI-Expanded of the WoS Core Collection database. Fig. 2 shows the annual trend of retracted articles over this time. The trend in the number of retracted articles exhibited apparent fluctuations from 2003 to 2022. Initially, there was a consistent increase, rising from 137 articles in 2003 to a peak

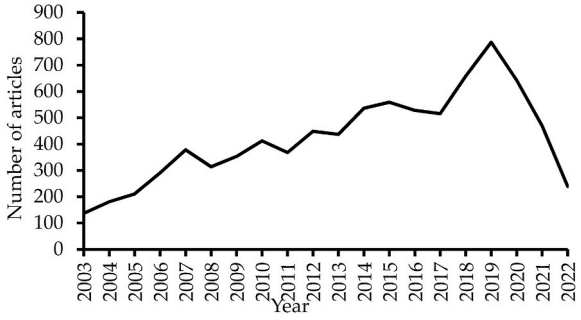


Fig. 2. Plot of annual number of retracted articles from 2003 to 2022.

of 787 in 2019. However, after reaching the peak in 2019, the trend sharply reversed, with the number of retracted articles falling to 239 by 2022.

A total of 109 countries contributed to the articles analyzed in the present study. Table 1 shows the top 10 countries, based on the affiliation of the corresponding author, of retracted articles. China (n = 3,669, 43.3 %) represented the largest share of retracted articles, followed by the United States (n = 1,246, 14.7 %), and India (n = 529, 6.2 %). The United States had the highest average citation number (51.9), followed by the United Kingdom (51.7). In addition, China showed the highest ratio of multiple-country publications to total publications (9.0 %), indicating a high inter-country collaboration. In contrast, India showed the lowest ratio at 1.2 %.

The retrieved articles were published in 2347 different journals. A total of 1137 retracted articles (13.4 %) were published in the top 10 journals. *Tumor Biology*, the journal with the largest number of articles (n = 153), had a Journal Impact Factor of 3.7 in 2016 but was no longer indexed by WoS afterward. Among the top 10 journals, the *Journal of Clinical Anesthesia* had the highest Journal Impact Factor of 6.7. Regarding WoS categories, oncology was the most prevalent, and most journals belonged to either quartile 2 or 3 (Table 2).

Fig. 3 represents the number of retracted articles for the top 10 journals over time. The lines showed a general increase in the number of articles published after 2011, with occasional fluctuations. *Tumor Biology* experienced a sharp rise in 2012–2014. The curve discontinued after 2016 as the journal was delisted by WoS in July 2017 [38]. *Journal of Cellular Biochemistry*, *European Review for Medical and Pharmacological Sciences*, *RSC Advances*, and *Cluster Computing* also exhibited sharp rises in 2017, followed by decreases. *Journal of Biological Chemistry* displayed a steady number of retracted articles, while other journals showed an increase from 2014 to 2018, then declined.

Table 3 showcases the top 10 leading retracted articles based on the total number of citations during the study period. The top article, a large multicenter trial in Spain investigating the protective effect of a Mediterranean diet supplemented with extra-virgin olive oil or nuts, was first published in 2013 in the *New England Journal of Medicine* [39] and received a total of 2363 citations. However, due to errors in methods, the article was retracted [40] and republished in 2018 [41] after excluding study sites that violated randomization principles. *BMC Evolutionary Biology* retracted the article by Jobb et al. [42] due to the decision by the corresponding author that the software could not be made available publicly, which breached the journal's editorial policy on software availability [43]. The remaining eight articles were retracted due to either data or image integrity issues.

The retracted articles belonged to 204 subject categories of WoS. Table 4 lists the 20 most frequently occurring subject categories of WoS in retracted articles from 2003 to 2022. Oncology (12.3 %), biochemistry & molecular biology (10.0 %), cell biology (8.5 %), medicine research & experimental (7.3 %), and pharmacology & pharmacy (6.7 %) represented the top five categories, accounting for a total of 3798 articles (44.9 %) of the 8466 articles.

The co-occurrence analysis conducted using VOSviewer revealed distinct thematic clusters, as depicted in Fig. 4. Four interconnected clusters were observed. The green cluster focused on expression with terms related to cancer research, such as proliferation, invasion, metastasis, migration, and progression. The red cluster is composed of terms associated with molecular studies, such as apoptosis, activation, inflammation, inhibition, and pathway. The yellow cluster focused on terms related to pathogenesis, such as disease, therapy, survival, management, risk, and epidemiology. The blue cluster presents a broader array of terms, such as model, prediction, identification, mechanism, and protein.

To examine the reasons for retractions, the 8164 articles identified using the WoS were linked to records in the Retraction Watch Database based on DOI of the articles. Of the 7375 articles (90.3 %) successfully linked, the total count of the reasons was 21,422 cases were obtained because each article can be assigned with multiple reasons according to the Retraction Watch Database (Table 5). The most prevalent issue, accounting for 28.8 % (6160 cases), is related to data and results issues. Plagiarism and duplication follow with 19.1 % (4085 cases), and investigations and findings constitute 16.8 % (3588 cases). Authorship and ethical concerns represent 9.0 % (1918 cases), while misconduct and fraud account for 6.0 % (1288 cases). Image manipulation and fabrication are noted in 5.1 % (1092 cases) of the issues. Peer review and editorial issues, withdrawal and retraction notices, and institutional and policy issues make up 4.6 % (985 cases), 3.6 % (763 cases), and 3.0 % (643 cases) respectively. Complaints and objections constitute 2.0 % (418 cases), miscellaneous issues are 1.9 % (412 cases), and procedural and legal issues are the least frequent at 0.3 % (70 cases).

Table 1
The top 10 countries, based on the affiliation of the corresponding author, of retracted articles from 2003 to 2022 (N = 8466).

Rank	Country	Number of Articles (%)	Average Article Citations	MCP/Total Ratio (%)
1	China	3669 (43.3)	16.7	9.0
2	The United States	1246 (14.7)	51.9	2.2
3	India	529 (6.2)	22.7	1.2
4	Japan	440 (5.2)	27.0	1.4
5	Iran	374 (4.4)	18.8	2.0
6	Korea	232 (2.7)	19.2	1.6
7	Italy	174 (2.1)	23.6	3.0
8	The United Kingdom	171 (2.0)	51.7	3.9
9	Germany	160 (1.9)	41.0	3.3
10	Egypt	120 (1.4)	23.8	2.6

MCP: multiple country publication.

Table 2
The top 10 journals based on the number of retracted articles from 2003 to 2022 (N = 8466).

Rank	Journal	Number of Articles (%)	Journal Impact Factor[a]	Web of Science Category [Quartile]
1	*Tumor Biology*	153 (1.8)	3.7 (last indexed in 2016)	Oncology [2]
2	*Journal of Cellular Biochemistry*	142 (1.7)	4.0	Biochemistry & Molecular Biology [2] Cell Biology [3]
3	*Journal of Biological Chemistry*	133 (1.6)	4.8	Biochemistry & Molecular Biology [2]
4	*RSC Advances*	128 (1.5)	3.9	Chemistry, multidisciplinary [2]
5	*European Review for Medical and Pharmacological Sciences*	124 (1.5)	3.3	Pharmacology & Pharmacy [3]
6	*PLoS One*	117 (1.4)	3.7	Multidisciplinary Sciences [2]
7	*Molecular Medicine Reports*	102 (1.2)	3.4	Oncology [3] Medicine, Research & Experimental [3]
8	*Oncology Reports*	83 (1.0)	4.2	Oncology [2]
9	*Cluster Computing*	80 (0.9)	4.4	Computer Science, Theory & Methods [2] Computer Science, Information Systems [1]
10	*Journal of Clinical Anesthesia*	75 (0.9)	6.7	Anesthesiology [1]

[a] Journal Impact Factors were obtained from the 2022 Journal Citation Reports, unless otherwise indicated.

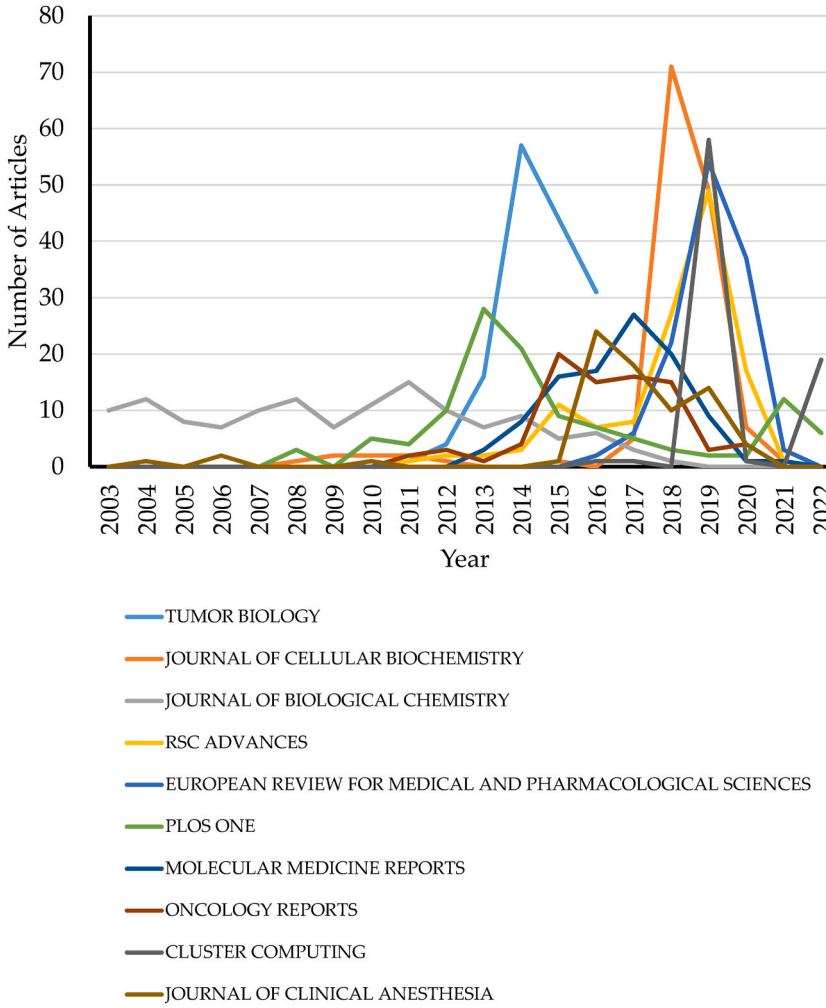


Fig. 3. Plot of annual number of retracted articles in the top 10 journals with the highest num-ber of retracted articles from 2003 to 2022 (N = 8466).

Table 3
The top 10 retracted articles based on the number of citations from 2003 to 2022 (N = 8466).

Rank	First Author (No. of Total Authors)	Title	Journal	Year of Publication	Total Citations[a]	Year of Retraction	Reason for retraction
1	Estruch R (18 & study team)	Primary prevention of cardiovascular disease with a Mediterranean diet	*New England Journal of Medicine*	2013	2362	2013	Errors in methods
2	Fukuhara A (22)	Visfatin: a protein secreted by visceral fat that mimics the effects of insulin	*Science*	2005	1490	2007	Data integrity issue, result issue
3	Voinnet O (4)	An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus	*Plant Journal*	2003	1336	2015	Image integrity issue
4	Erler JT (9)	Lysyl oxidase is essential for hypoxia-induced metastasis	*Nature*	2006	1094	2020	Image integrity issue
5	Jobb G (3)	TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics	*BMC Evolutionary Biology*	2004	1007	2015	Could not make software available
6	Bolli R (20)	Cardiac stem cells in patients with ischaemic cardiomyopathy (SCIPIO): initial results of a randomised phase 1 trial	*Lancet*	2011	1000	2019	Data integrity issue
7	Mehra MR (5)	Cardiovascular disease, drug therapy, and mortality in Covid-19	*New England Journal of Medicine*	2020	854	2020	Data integrity issue
8	Raj L (14)	Selective killing of cancer cells by a small molecule targeting the stress response to ROS	*Nature*	2011	849	2018	Image integrity issue
9	Rubio D (7)	Spontaneous human adult stem cell transformation	*Cancer Research*	2005	803	2010	Data integrity issue
10	Nakao N (6)	Combination treatment of angiotensin-II receptor blocker and angiotensin-converting-enzyme inhibitor in non-diabetic renal disease (COOPERATE): a randomised controlled trial	*Lancet*	2003	742	2009	Data integrity issue

[a] Total citation included citation before and after retraction.

Table 4
The top 20 subject categories of Web of Science with retracted articles from 2003 to 2022 (N = 8466).

Rank	Subject categories	n (%)	Cumulative n (%)
1	Oncology	1044 (12.3)	1044 (12.3)
2	Biochemistry & Molecular Biology	848 (10.0)	1892 (22.3)
3	Cell Biology	717 (8.5)	2609 (30.8)
4	Medicine, Research & Experimental	619 (7.3)	3228 (38.1)
5	Pharmacology & Pharmacy	570 (6.7)	3798 (44.9)
6	Chemistry, Multidisciplinary	427 (5.0)	4225 (49.9)
7	Materials Science, Multidisciplinary	425 (5.0)	4650 (54.9)
8	Multidisciplinary Sciences	343 (4.1)	4993 (59.0)
9	Engineering, Electrical & Electronic	337 (4.0)	5330 (63.0)
10	Neurosciences	266 (3.1)	5596 (66.1)
11	Physics, Applied	243 (2.9)	5839 (69.0)
12	Chemistry, Physical	240 (2.8)	6079 (71.8)
13	Biotechnology & Applied Microbiology	225 (2.7)	6304 (74.5)
14	Immunology	200 (2.4)	6504 (76.8)
15	Computer Science, Information Systems	187 (2.2)	6691 (79.0)
16	Nanoscience & Nanotechnology	177 (2.1)	6868 (81.1)
17	Medicine, General & Internal	168 (2.0)	7036 (83.1)
18	Surgery	163 (1.9)	7199 (85.0)
19	Telecommunications	161 (1.9)	7360 (86.9)
20	Environmental Sciences	155 (1.8)	7515 (88.8)

The cross-tabulation of the top three retraction reasons and main subject categories for articles retracted between 2003 and 2022 revealed distinct patterns in the distribution of retraction causes across different fields, with an overall Chi-square p-value of <0.001. A total of 6883 articles were analyzed, excluding those with missing data in subject categories. “Data and results issues” were significantly more prevalent in Basic Life Sciences (24.8 %) and Health Sciences (24.4 %) compared to Business and Technology (18.0 %). “Plagiarism and duplication” were significantly more common in Physical Sciences (35.9 %) and the "Other" category (34.6 %), which includes humanities, social sciences, and environmental sciences, compared to Basic Life Sciences (18.1 %) and Business and Technology (26.4 %). Retractions due to “Investigations and findings” were most prevalent in Health Sciences (22.7 %), significantly exceeding the proportions in Business and Technology (7.2 %) and Physical Sciences (7.2 %). Combinations of any of the top three

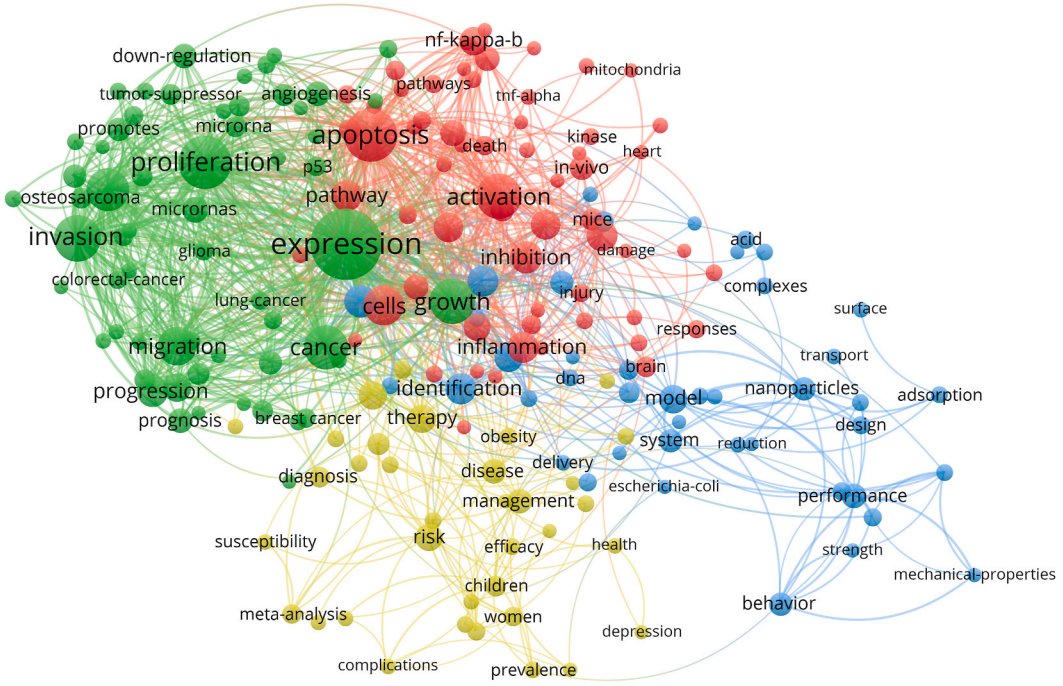


Fig. 4. A co-occurrence network of keywords in retracted articles from 2003 to 2022 (N = 8466).

Table 5
Distribution of retraction reasons of retracted articles from 2003 to 2022.

Retraction Reason	Count (%)
Data and Results Issues	6160 (28.8)
Plagiarism and Duplication	4085 (19.1)
Investigations and Findings	3588 (16.8)
Authorship and Ethical Concerns	1918 (9.0)
Misconduct and Fraud	1288 (6.0)
Image Manipulation and Fabrication	1092 (5.1)
Peer Review and Editorial Issues	985 (4.6)
Withdrawal and Retraction Notices	763 (3.6)
Institutional and Policy Issues	643 (3.0)
Complaints and Objections	418 (2.0)
Miscellaneous	412 (1.9)
Procedural and Legal Issues	70 (0.3)
Total	**21422 (100)**

Of the 8466 articles identified in the Web of Science, 7375 were successfully linked with records from the Retraction Watch Database. As each article may have multiple retraction reasons, a total count of 21,422 for the various reasons was documented.

retraction reasons were prevalent in Business and Technology (36.6 %) and Basic Life Sciences (35.6 %) compared to other categories. Lastly, other reasons and their combinations were more common in the "Other" category (18.5 %), Physical Sciences (17.6 %), and Health Sciences (14.6 %) compared to other categories (Table 6).

4. Discussion

Overall, this bibliometric study analyzed retracted articles published between 2003 and 2022 and indexed by the SCI-Expanded of WoS Core Collection database. The results revealed the patterns and implications of article retractions, contributing to the theoretical understanding of research integrity and publication ethics. The study highlighted the significance of regional and institutional factors, as well as disciplinary vulnerabilities, indicating that systemic pressures and the complexity of research methodologies in certain fields are closely linked to the incidence of retractions. Regarding practical implications, this study showed the importance of strengthening peer review processes, as illustrated by the *Tumor Biology* case, and suggests the need of new technologies to detect fraudulent reviews. Moreover, standardized practices for linking retraction notices directly to the retracted articles in databases should be implemented to

Table 6
Cross-tabulation of main retraction reasons and subject categories of retracted articles from 2003 to 2022.

Reason	Subject category, n (column %)					
	Business and Technology	Basic Life Sciences	Health Sciences	Physical Sciences	Other	Row Total
Data and Results Issues	110[a] (18.0)	1034[b] (24.8)	197[b] (24.4)	243[a,b] (20.9)	29[a,b] (22.3)	1613 (23.4)
Plagiarism and Duplication	161[a,b] (26.4)	756[c] (18.1)	186[b] (23.0)	417[d] (35.9)	45[a,d] (34.6)	1565 (22.7)
Investigations and Findings	44[a] (7.2)	502[b] (12.0)	184[c] (22.7)	84[a] (7.2)	11[a,b] (8.5)	825 (12.0)
Combinations of the above three reasons	223[a] (36.6)	1487[a] (35.6)	124[b] (15.3)	213[b] (18.3)	21[b] (16.2)	2068 (30.0)
Other reasons and combinations[a]	72[a,b] (11.8)	394[b] (9.4)	118[a,c] (14.6)	204[c] (17.6)	24[a,c] (18.5)	812 (11.8)
Column Total	610 (100)	4173 (100)	809 (100)	1161 (100)	130 (100)	6883 (100)

Each subscript letter denotes a subset of subject category whose column proportions do not differ significantly from each other at the 0.05 level, adjusted using the Bonferroni correction.

[a] Other subject categories included humanities (n = 6), social sciences (n = 14), and environmental sciences (n = 110).

mitigate the spread of invalidated research findings. In addition, the correlation between competitive pressures in specific fields and retractions signifies the need to address systemic issues within the scientific community. Targeted educational programs in research ethics for fields more prone to retractions is recommended.

During the 20-year period, a total of 8466 retracted articles were identified. There has been an increase in the number of retracted articles up to 2019. While the increase could correspond to the growth in the total number of publications during the same period, it does not explain the drop after 2019. The decline after 2019 might be attributed to the time delay required for retractions to occur, considering that retractions may take several years. Based on the top 10 cited retracted articles, the period before retraction ranged from less than a year to 14 years, with a median of 6.5 years. A study of retracted articles from 1960 to 2020 indicated that at least half of the retractions occurred no more than three years after publication [9]. Therefore, a drop in the number of retracted articles after 2019 might be attributed to the time delay required for retractions to take place. However, it is also possible that the scientific community has become more stringent in peer-review processes as well as the recent initiatives to increase transparency [44] might have contributed to this effect. Moreover, the introduction of new tools for the detection of plagiarism might have enabled earlier detection and reduced fraudulent activities leading to retractions later on.

The present study identified China and the United States as the countries with the largest number of retracted articles, accounting for 58 % of the 8466 retracted articles. Similar findings have been reported in other studies that analyzed different periods and databases. A study analyzed retracted articles indexed by WoS between 1981 and 2020 revealed that authors from China have higher retractions (25.7 %), followed by the United States (16.1 %), and India (5.3 %) [45]. Another PubMed-based study on articles published between 2008 and 2012 showed that the United States retracted the most articles, whereas China retracted the most articles for plagiarism and duplicate publication [46]. An early study based on 834 WoS retractions also found that the number of retracted papers from Chinese authors had increased between 1997 and 2016, with approximately three-quarters of which were triggered by fake peer review, plagiarism, or falsification [47]. Nevertheless, these results should be interpreted within the context of the size of the body of literature. Both China and the United States are among the largest producers of scientific research globally. Their substantial publication volume might contribute to a ‘‘publication inflation’’ that partially explains the higher number of retractions [11]. When evaluating with a different index, such as retractions per paper published, Iran takes the top spot with 14 retractions per 10,000 papers, followed by Romania with 10.4 retractions per 10,000 papers. China became the seventh spot with five retractions per 10,000 papers [48]. It should be noted that while focusing on specific countries provides significant insights, understanding retraction patterns requires a broader, global perspective.

Tumor Biology, a medical journal covering clinical and experimental oncology, had the highest number of retracted articles. Starting in 2012, the journal showed a sharp rise in the number of articles that were subsequently retracted. In 2017, its publisher Springer retracted 107 articles after discovering they were accepted based on fake peer reviews. As a result of the mass retraction, the journal was delisted by WoS starting July 2017 [38]. The incident with *Tumor Biology* underscores the importance of authentic peer review and indicates the need for rigorous measures to prevent potential lapses in the process. To help editors choose credible reviewers and close peer review loopholes, it is imperative to create an expert database with verified information on experts' credentials. In addition, new technologies should be explored to assist the detection of suspiciously positive reviews [49].

Findings from top 10 cited retracted articles indicated that most were retracted due to image and data integrity issues. The former often involved duplication of images, whereas the latter was often linked to the implausibility of data, where the results were statistically or scientifically inconsistent with previous findings. Articles were retracted because the authors failed to provide raw data when requested, which can be due to poor data management practices or intentional concealment of fraudulent or flawed data. Moreover, the high number of citations received by these articles is another important concern. Previous studies have demonstrated that articles continued to be cited years after retraction [50]. One contributing factor to this issue is that retraction notices often lack clear indication and linkage to the retracted articles [9]. Putting flawed data in the scientific literature can mislead both fellow

scientists and the broader community. As a result, the scientific community needs to enhance the clarity and transparency of retraction notices and uphold the integrity of the scientific literature.

Findings from the analysis of WoS subject categories and co-occurrence network analysis of keywords showed that articles in the field of oncology, biochemistry, molecular biology, and cell biology appeared to be particularly prone to retractions. This pattern might reflect underlying challenges and complexities within these disciplines or broader systemic issues within the scientific community. Studies in these fields are characterized by intricate experimental designs and advanced technologies, which might increase the likelihood of errors. The competitive nature and pressure to produce novel findings in these fields may also inadvertently encourage behaviors leading to retractions. A systematic survey of 571 retracted cancer publications revealed that academic misconduct in the form of plagiarism, duplicate publication, and fraud accounts for 61 % of retracted cancer publications [14]. A recent bibliometric study on retracted publications in oncology used Retraction Watch Database to explore the reasons for retractions. The top three reasons included investigation by journal/publisher (30.3 %), duplication of image (24.9 %), and fake peer review (22.6 %), accounting for a combined total of 77.8 % of the retractions analyzed [51]. In addition, the prominence of these fields may lead to higher levels of scrutiny by peers, leading to more retractions being detected and acted upon. Therefore, the findings might reflect a combination of underlying complexities and systemic issues within these disciplines.

The analysis of retraction reasons showed that the top two most prevalent reason for retractions are “data and results issues” and “plagiarism and duplication”, accounting for 48 % of the cases. Approximately 29 % of the retraction reasons are related to the category “data and results issues”. This finding indicates that important role of data integrity in scholarly research. Issues such as unreliable or error in data and results can adversely affect the credibility of research findings and are critical areas for scrutiny and improvement. The category “plagiarism and duplication” was the second most common reasons for retraction. Despite the widespread availability of plagiarism detection tools, their current application is insufficient to address the problem. Journals need to adopt a more vigilant approach to plagiarism detection at multiple stages of the publication process, including initial submission, post-peer review, and pre-publication to uphold the standards of academic publishing and ensure the integrity of scholarly research.

The cross-tabulation of subject categories and retraction reasons revealed distinct patterns that may be attributed to the nature and demands of the respective disciplines. “Data and results issues” were more prevalent in Basic Life Sciences and Health Sciences. This trend possibly arises from the complex and often intricate methodologies required in these fields, where experimental design, data collection, and analysis are highly susceptible to error, misinterpretation, or even manipulation. In contrast, “plagiarism and duplication” are more common in the Physical Sciences. This observation suggests variations in ethical standards across disciplines, where the rigor of plagiarism detection and enforcement might differ. The structure of research in Physical Sciences, which often involves collaborative and cumulative work, might also make it easier for such ethical breaches to occur unnoticed. In addition, the higher rate of retractions observed in Health Sciences may reflect the intense scrutiny and rigorous investigative processes applied to research that has direct implications for public health.

This study has some limitations worth mentioning. First, the articles analyzed in this study were retrieved from the WoS. While it is well-known that various databases have different strengths and weaknesses [52], WoS was chosen for this study because it contains citation metadata, particularly the availability of a dedicated document type for “retracted publication.” Second, an author-level analysis was not conducted due to the significant issue of author name ambiguity, particularly the Romanized names of Chinese authors [53]. Third, 9.7 % of the retracted articles identified in the WoS could not be linked to the Retraction Watch Database using DOI. The absence of these articles may potentially affect the magnitude of the counts. However, given the substantial difference in the frequency of the top three retraction reasons, it is unlikely that this limitation will alter their ranking.

5. Conclusion

Retractions play a vital role in maintaining research integrity by ensuring the accuracy of the scientific record, promoting transparency and accountability, deterring misconduct, and fostering a culture of continuous improvement within the scientific community. Our bibliometric analysis of 8466 retracted articles from 2003 to 2022, indexed by the SCI-Expanded of the WoS Core Collection, offered an overview of the current state of retractions and reasons for retractions.

The study reveals a notable increase in retractions until 2019, followed by a decline, indicating evolving awareness and response to scientific misconduct and errors in the research community. China and the United States were found to lead in the number of retractions, reflecting their large contributions to global scientific output. Specific journals and subject areas, particularly oncology, biochemistry, and cell biology, were more prone to retractions, highlighting the vulnerability of these fields.

Our study emphasizes the importance of robust peer review processes and the use of advanced tools for detecting misconduct. The significant impact of retracted articles, evidenced by their high citation counts even after retraction, suggests the need for better dissemination of retraction notices and greater transparency in the retraction process. Promoting a culture of integrity and accountability within the scientific community can reduce the adverse effects of retractions and enhance the reliability of published research.

Our findings provide a basis for developing targeted educational programs and policies to improve research ethics. As the scientific landscape evolves, maintaining vigilance and commitment to ethical standards is crucial to ensure the progression of credible scientific literature.

Funding

This work was supported by Tzu Chi University (formerly Tzu Chi University of Science and Technology, TCCT-1111A08).

Data availability statement

Data will be made available on request.

CRediT authorship contribution statement

Malcolm Koo: Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Shih-Chun Lin:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e38620.

References

[1] D.B. Resnik, A.E. Shamoo, The Singapore statement on research integrity, Account. Res. 18 (2011) 71–75.
[2] M.C. Atlas, Retraction policies of high-impact biomedical journals, J. Med. Libr. Assoc. 92 (2004) 242–250.
[3] COPE. Retraction guidelines. (Retrieved September 9, 2023) https://publicationethics.org/files/retraction-guidelines.pdf.
[4] S.F. Lu, G.Z. Jin, B. Uzzi, B. Jones, The retraction penalty: evidence from the web of science, Sci. Rep. 3 (2013) 3146.
[5] A.M. Stern, A. Casadevall, R.G. Steen, F.C. Fang, Financial costs and personal consequences of research misconduct resulting in retracted publications, Elife 3 (2014) e02956.
[6] D. Fanelli, Why growing retractions are (mostly) a good sign, PLoS Med. 10 (2013) e1001563.
[7] R.G. Steen, Retractions in the medical literature: how many patients are put at risk by flawed research? J. Med. Ethics 37 (2011) 688–692.
[8] D.G. Hamilton, Continued citation of retracted radiation oncology literature-do we have a problem? Int. J. Radiat. Oncol. Biol. Phys. 103 (2019) 1036–1042.
[9] T.K. Hsiao, J. Schneider, Continued use of retracted papers: temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine, Quant. Sci. Stud. 2 (2022) 1144–1169.
[10] F.C. Fang, R.G. Steen, A. Casadevall, Misconduct accounts for the majority of retracted scientific publications, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 17028–17033.
[11] M.L. Grieneisen, M. Zhang, A comprehensive survey of retracted articles from the scholarly literature, PLoS One 7 (2012) e44118.
[12] R.G. Steen, A. Casadevall, F.C. Fang, Why has the number of scientific retractions increased? PLoS One 8 (2013) e68397.
[13] S. Nair, C. Yean, J. Yoo, J. Leff, E. Delphin, D.C. Adams, Reasons for article retraction in anesthesiology: a comprehensive analysis, Can. J. Anaesth. 67 (2020) 57–63.
[14] A. Bozzo, K. Bali, N. Evaniew, M. Ghert, Retractions in cancer research: a systematic survey, Res. Integr. Peer Rev 2 (2017) 5.
[15] K. Audisio, N.B. Robinson, G.J. Soletti, G. Cancelli, A. Dimagli, C. Spadaccio, R.P. Olaria, D. Chadow, M. Rahouma, M. Demetres, D.Y. Tam, U. Benedetto, L. N. Girardi, P. Kurlansky, S.E. Fremes, M. Gaudino, A survey of retractions in the cardiovascular literature, Int. J. Cardiol. 349 (2022) 109–114.
[16] C.M. Faggion Jr., R.S. Ware, N. Bakas, J. Wasiak, An analysis of retractions of dental publications, J. Dent. 79 (2018) 19–23.
[17] J. Wang, J.C. Ku, N.M. Alotaibi, J.T. Rutka, Retraction of neurosurgical publications: a systematic review, World Neurosurg 103 (2017) 809–814.e1.
[18] A. Al-Ghareeb, S. Hillel, L. McKenna, M. Cleary, D. Visentin, M. Jones, D. Bressington, R. Gray, Retraction of publications in nursing and midwifery research: a systematic review, Int. J. Nurs. Stud. 81 (2018) 8–13.
[19] S. Joaquim, J. Longhini, A. Palese, What can we learn from retracted studies in the nursing field in the last 20 years? Findings from a scoping review, Acta Biomed. 93 (2022) e2022193.
[20] C. Bennett, L.M. Chambers, L. Al-Hafez, C.M. Michener, T. Falcone, M. Yao, V. Berghella, Retracted articles in the obstetrics literature: lessons from the past to change the future, Am. J. Obstet. Gynecol. MFM. 2 (2020) 100201.
[21] J. Yan, A. MacDonald, L.P. Baisi, N. Evaniew, M. Bhandari, M. Ghert, Retractions in orthopaedic research: a systematic review, Bone Joint Res 5 (2016) 263–268.
[22] M.M. Christopher, Comprehensive analysis of retracted journal articles in the field of veterinary medicine and animal health, BMC Vet. Res. 18 (2022) 73.
[23] S.Y. Hwang, D.K. Yon, S.W. Lee, M.S. Kim, J.Y. Kim, L. Smith, A. Koyanagi, M. Solmi, A.F. Carvalho, E. Kim, J.I. Shin, J.P.A. Ioannidis, Causes for retraction in the biomedical literature: a systematic review of studies of retraction notices, J. Korean Med. Sci. 38 (2023) e333.
[24] D. Donthu, S. Kumar, D. Mukherjee, N. Pandey, W.M. Lim, How to conduct a bibliometric analysis: an overview and guidelines, J. Bus. Res. 133 (2021) 285–296.
[25] A. Ninkov, J.R. Frank, L.A. Maggio, Bibliometrics: methods for studying academic publishing, Perspect. Med. Educ. 11 (2022) 173–176.
[26] H. Khan, P. Gupta, O. Zimba, L. Gupta, Bibliometric and altmetric analysis of retracted articles on COVID-19, J. Korean Med. Sci. 37 (2022) e44.
[27] P. Pantziarka, L. Meheus, Journal retractions in oncology: a bibliometric study, Future Oncol. 15 (2019) 3597–3608.
[28] M. Bordino, E. Ravizzotti, S. Vercelli, Retracted articles in rehabilitation: just the tip of the iceberg? A bibliometric analysis, Arch. Physiother. 10 (2020) 21.
[29] P. Chen, X.H. Li, Z. Su, Y.L. Tang, Y. Ma, C.H. Ng, Y.T. Xiang, Characteristics of global retractions of schizophrenia-related publications: a bibliometric analysis, Front. Psychiatry 13 (2022) 937330.
[30] Clarivate. Web of science: Science citation index expanded. (Retrieved September 9, 2023) https://clarivate.com/webofsciencegroup/solutions/webofscience-scie/.
[31] J. Baas, M. Schotten, A. Plume, G. Côté, R. Karimi, Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies, Quant. Sci. Stud. 1 (2020) 377–386.
[32] R. Pranckute, Web of Science (WoS) and Scopus: the Titans of Bibliographic Information in Today's Academic World, vol. 9, 2021, p. 12. *Publications.* 2021.
[33] M. Aria, C. Cuccurullo, Bibliometrix: an r-tool for comprehensive science mapping analysis, J. Informetr. 11 (2017) 959–975.
[34] N.J. van Eck, L. Waltman, Software survey: vosviewer, a computer program for bibliometric mapping, Scientometrics 84 (2010) 523–538.
[35] J. Zhang, Q. Yu, F. Zheng, C. Long, Z. Lu, Z. Duan, Comparing keywords plus of WOS and author keywords: a case study of patient adherence research, J. Assoc. Inf. Sci. Technol. 67 (2015) 967–972.

[36] Retraction Watch Database. (Retrieved May 20, 2024)http://retractiondatabase.org/.
[37] Retraction Watch Database. User Guide Appendix B: Reasons. (Retrieved May 20, 2024) https://retractionwatch.com/retraction-watch-database-user-guide/retraction-watch-database-user-guide-appendix-b-reasons/.
[38] Watch R. When a journal retracts 107 papers for fake reviews, it pays a price.(Retrieved May 20, 2024) https://retractionwatch.com/2017/08/16/journal-retracts-107-papers-fake-reviews-pays-price/.
[39] R. Estruch, E. Ros, J. Salas-Salvado, M.I. Covas, D. Corella, F. Aros, E. Gomez-Gracia, V. Ruiz-Gutierrez, M. Fiol, J. Lapetra, R.M. Lamuela-Raventos, L. Serra-Majem, X. Pintó, J. Basora, M.A. Muñoz, J.V. Sorlí, J.A. Martínez, M.A. Martínez-González, PREDIMED Study Investigators, Primary prevention of cardiovascular disease with a Mediterranean diet, N. Engl. J. Med. 368 (2013) 1279–1290.
[40] R. Estruch, E. Ros, J. Salas-Salvado, M.I. Covas, D. Corella, F. Aros, E. Gomez-Gracia, V. Ruiz-Gutierrez, M. Fiol, J. Lapetra, R.M. Lamuela-Raventos, L. Serra-Majem, X. Pintó, J. Basora, M.A. Muñoz, J.V. Sorlí, J.A. Martínez, M.A. Martínez-González, Retraction and republication: primary prevention of cardiovascular disease with a Mediterranean diet, N. Engl. J. Med. 378 (2018) 2441–2442.
[41] R. Estruch, E. Ros, J. Salas-Salvado, M.I. Covas, D. Corella, F. Aros, E. Gomez-Gracia, V. Ruiz-Gutierrez, M. Fiol, J. Lapetra, R.M. Lamuela-Raventos, L. Serra-Majem, X. Pintó, J. Basora, M.A. Muñoz, J.V. Sorlí, J.A. Martínez, M. Fitó, A. Gea, M.A. Hernán, M.A. Martínez-González, PREDIMED Study Investigators, Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts, N. Engl. J. Med. 378 (2018) e34.
[42] G. Jobb, A. von Haeseler, K. Strimmer, Treefinder: a powerful graphical analysis environment for molecular phylogenetics, BMC Evol. Biol. 4 (2004) 18.
[43] G. Jobb, A. von Haeseler, K. Strimmer, Retraction note: treefinder: A powerful graphical analysis environment for molecular phylogenetics, BMC Evol. Biol. 15 (2015) 243.
[44] K.C. Elliott, D.B. Resnik, Making open science work for science and society, Environ. Health Perspect. 127 (2019) 75002.
[45] K. Sharma, Team size and retracted citations reveal the patterns of retractions from 1981 to 2020, Scientometrics 126 (2021) 8363–8374.
[46] K.A. Amos, The ethics of scholarly publishing: exploring differences in plagiarism and duplicate publication across nations, J. Med. Libr. Assoc. 102 (2014) 87–91.
[47] L. Lei, Y. Zhang, Lack of improvement in scientific integrity: an analysis of wos retractions by Chinese researchers (1997-2016), Sci. Eng. Ethics 24 (2018) 1409–1420.
[48] I. Oransky, Volunteer watchdogs pushed a small country up the rankings, Science 362 (2018) 395.
[49] Y. Xiao, J. Chen, X.H. Wu, Q.M. Qiu, High retraction rate of Chinese articles: it is time to do something about academic misconduct, Postgrad. Med. J. 98 (2022) 653–654.
[50] J. Bar-Ilan, G. Halevi, Temporal characteristics of retracted articles, Scientometrics 116 (2018) 1771–1783.
[51] B. Gholampour, S. Gholampour, A. Noruzi, C. Arsenault, T. Haertlé, A.A. Saboury, Retracted articles in oncology in the last three decades: frequency, reasons, and themes, Scientometrics 127 (2022) 1841–1865.
[52] S.A.S. AlRyalat, L.W. Malkawi, S.M. Momani, Comparing bibliometric analysis using PubMed, scopus, and web of science databases, J. Vis. Exp. 152 (2019) e58494.
[53] A. Strotmann, D. Zhao, Author name disambiguation: what difference does it make in author-based citation analysis? J. Am. Soc. Inf. Sci. Technol. 63 (2012) 1820–1833.