science editing
pISSN 2288-8063
eISSN 2288-7474
Sci Ed 2020;7(1):34-44
https://doi.org/10.6087/kcse.187

Original Article

Characteristics of retracted articles based on retraction data from online sources through February 2019

Quan-Hoang Vuong[1,2], Viet-Phuong La[2], Manh-Tung Ho[2,3], Thu-Trang Vuong[4], Manh-Toan Ho[2]

[1]Scientific Council on Basic Research in the Social Sciences and Humanities, National Foundation for Science and Technology Development (NAFOSTED), Hanoi; [2]Centre for Interdisciplinary Social Research, Phenikaa University, Hanoi, Vietnam; [3]Ritsumeikan Asia Pacific University, Beppu, Japan; [4]Sciences Po Paris, Paris, France

Abstract

Purpose: Although retractions are commonly considered to be negative, the fact remains that they play a positive role in the academic community. For instance, retractions help scientific enterprise perform its self-correcting function and provide lessons for future researchers; furthermore, they represent the fulfillment of social responsibilities, and they enable scientific communities to offer better monitoring services to keep problematic studies in check. This study aims to provide a thorough overview of the practice of retraction in scientific publishing from the first incident to the present.

Methods: We built a database using SQL Server 2016 and homemade artificial intelligence tools to extract and classify data sources including RetractionWatch, official publishers' archives, and online communities into ready-to-analyze groups and to scan them for new data. After data cleaning, a dataset of 18,603 retractions from 1,753 (when the first retracted paper was published) to February 2019, covering 127 research fields, was established.

Results: Notable retraction events include the rise in retracted articles starting in 1999 and the unusual number of retractions in 2010. The Institute of Electrical and Electronics Engineers, Elsevier, and Springer account for nearly 60% of all retracted papers globally, with Institute of Electrical and Electronics Engineers contributing the most retractions, even though it is not the organization that publishes the most journals. Finally, reasons for retraction are diverse but the most common is "fake peer review".

Conclusion: This study suggests that the frequency of retraction has boomed in the past 20 years, and it underscores the importance of understanding and learning from the practice of retracting scientific articles.

Keywords

Academic publishing; Retraction; Scientific publication; Self-correcting capability

Received: August 26, 2019
Accepted: October 21, 2019

Correspondence to Manh-Toan Ho
toan.homanh@phenikaa-uni.edu.vn

ORCID
Quan-Hoang Vuong
https://orcid.org/0000-0003-0790-1576
Viet-Phuong La
https://orcid.org/0000-0002-4301-9292
Manh-Tung Ho
https://orcid.org/0000-0002-4432-9081
Thu-Trang Vuong
https://orcid.org/0000-0002-7262-9671
Manh-Toan Ho
https://orcid.org/0000-0002-8292-0120

Introduction

***Background*:** Retraction is described by the Committee on Publication Ethics as a mechanism for correcting the literature and alerting readers to publications that contain serious flaws or erroneous data to the extent that their findings and conclu-sions cannot be relied upon [1]. However, most readers and scientists regard retraction as an unfortunate negative outcome of the scientific enterprise. Retraction is seen as a source of embarrassment for all involved [2]. This is partially due to the public perceptions associated with the phenomenon: adverse consequences to the authors, wasted funds, wasted time and effort of the host institutions, and loss of the public's trust when the reputation of science is tainted by fraud [3], to name just a few.

It is thought that retraction can be an opportunity for learning and improvement [4]. Future researchers can learn from the reasons behind retraction [5]. Publicly available retraction notices represent the fulfillment of the social responsibilities of journals and publishers [6]. Open review communities, such as PubPeer, can offer better monitoring services to keep problematic studies in check.

***Specific goals*:** To better facilitate this truly powerful and posi-

Example of SQL code for data validation

```
-- Update status of duplicated Pubmed records
WITH CTE AS
(
SELECT *,ROW_NUMBER() OVER (PARTITION BY
dbo.ToPubmedUrl(OriginalPaperPubmed) ORDER BY
dbo.ToPubmedUrl(OriginalPaperPubmed)) AS RN
FROM retArticle WHERE isDupRec=0 AND
dbo.IsValidPubmed(OriginalPaperPubmed)=1
)
UPDATE CTE SET isDupRec=1 WHERE RN<>1
-- Detect the documents that have titles with a similarity (more than 90%)
select * from retArticle as aa where
exists (select 1 from retArticle as bb where
dbo.CalculateSimilarity(lower(aa.ArticleTitle),lower(bb.ArticleTitle))>0.9
and aa.Id<>bb.Id)
```

C# code for web crawling and parsing

```
string htmlText = string.Empty;
htmlText = WebDriverHelper.BrowseForTag(url, By.XPath(valid), 20);

if (!string.IsNullOrWhiteSpace(htmlText)&&
!htmlText.Contains(token["NotFoundText"].ToString()))
{
    if (maxPage == 1)
    {
        maxPage = ParseHelper.ParseMaxPage(htmlText,
(JObject)token["MaxPageCfg"]);
    }

    JObject list = (JObject)token["ListCfg"];
    string item = ParseHelper.ParseItemHtml(htmlText,
list["List"].ToString());

    if (!string.IsNullOrWhiteSpace(item))
    {
        var items = ParseHelper.ParseListHtml(item,
list["Item"].ToString(), string.Empty);

        if (items != null)
        {
            JObject itemCfg = (JObject)token["ItemCfg"];

            foreach (var ret in items)
            {

                ArticleData data = new ArticleData();
                data.ArticleTite = ParseHelper.ParseItem(ret,
itemCfg["Title"].ToString());
                data.Authors = ParseHelper.ParseItem(ret,
itemCfg["Author"].ToString());
                data.Journal = ParseHelper.ParseItem(ret,
itemCfg["Journal"].ToString());
                data.OriDOI = ParseHelper.ParseItem(ret,
itemCfg["OriDOI"].ToString());
                data.RetractDOI = ParseHelper.ParseItem(ret,
itemCfg["RetractDOI"].ToString());
                data.OriDate = (DateTime)ParseHelper.ParseDate(ret,
itemCfg["PubDate"].ToString());
                data.RetractDate = (DateTime)ParseHelper.ParseDate(ret,
itemCfg["RetractDate"].ToString());
```

Fig. 1. An example of code used.

tive function of retraction, a comprehensive database of retraction will be highly beneficial. Useful insights can be drawn from retraction data by asking questions. How old is the phenomenon? Are some specific publishers/journals more prone to retraction? Are retractions concentrated in certain fields? When did retractions begin to become more visible to the world? How long does it take for a journal to issue a retraction? Hence, by extracting insights from a homemade retraction database, of which the sources were RetractionWatch, official publishers' archives, and online communities, we aimed to answer the above questions and provide suggestions for changes in scientific publishing. By doing so, we can make use of the wisdom of the retracted papers and avoid issues associated with retraction altogether in the future.

Methods

***Ethics statement*:** No informed consent was required because this is a literature-based study.
Study design: This is a descriptive study that utilized database analysis.
Setting: A rise in scholarly publication retractions has been seen in recent years, according to sources of information such as RetractionWatch and publishers' retraction notices, which have fostered open discussions of retracted publications categorized by author, country, journal, subject, and type [7,8]. Yet, the large amounts of data stored in different systems may easily lead to omissions in results obtained by searching manually.

To bolster the value of retraction data, we embarked on a project to replicate data retrieved from online platforms, such as RetractionWatch, online journal archives, and online discussion communities. We scanned retractions that these sources may have missed, then stored the data in a database. We built this database using SQL Server 2016 (Microsoft, Seattle, WA, USA) and employed a web crawler tool to scan the data (see the file retractionCrawler (code).pdf at https://osf.io/7ahsn/ in [9] for the code for the web crawler tool).

Then, articles collected by the web crawler tool were cleaned and assessed for duplication using the DOI and PubMed databases.

Additionally, a fuzzy matching Levenshtein distance algorithm was used to find articles that had titles with a similarity of more than 90% (see file and validData (code).pdf at https://osf.io/c2zvj/ in [9] for the code for data validation). A code snippet is provided in Fig. 1.

After we eliminated 430 duplicate and incorrect records, the dataset contained 18,603 retractions, covering 127 research fields, from 1753 (when the first retracted paper was

Table 1. The list of the ten oldest retracted articles

Date of retraction	Date of publication	Bibliographic information of the retracted article
June 24, 1756	January 1, 1753	Treatise upon electricity. *Philosophical Transactions* (Royal Society Publishing)
April 1, 1927	April 1, 1926	The trend-seasonal normal in time series. *Journal of the American Statistical Association*, 21 (155), 321-329 (Taylor and Francis)
December 1, 1940	December 1, 1940	Naturwissenschaft und reale Aussenwelt. *Die Naturwissenschaften*, DOI: 10.1007/BF01488952 (Springer)
February 1, 1942	February 1, 1942	Sinn und Grenzen der exakten Wissenschaft. *Die Naturwissenschaften*, DOI: 10.1007/BF01475382 (Springer)
February 1, 1960	February 1, 1955	Change of venue and the conflict of laws. *The University of Chicago Law Review*, 22(405) (University of Chicago Law School)
October 1, 1966	October 1, 1959	On the primary site of nuclear RNA synthesis. *The Journal of Cell Biology*, DOI: 10.1083/jcb.6.2.301 (Rockefeller University Press)
August 26, 1968	September 6, 1963	Unmineralized fossil bacteria. *Science*, DOI: 10.1126/science.141.3584.919 (American Association for the Advancement of Science)
October 1, 1971	October 1, 1971	Hyperextensibility and weakness in cerebral palsy apparently opposed expression of the same muscular disorder. *Revue de Chirurgie Orthopedique et Reparatrice de L'appareil Moteur*, URL: https://www.ncbi.nlm.nih.gov/pubmed/4261570 (Elsevier)
February 24, 1977	November 1, 1975	Effects of cholinergic agents and sodium ions on the levels of guanosine and adenosine 3':5'-cyclic monophosphates in neuroblastoma and neuroblastoma X glioma hybrid cells. *FEBS Letters*, DOI: 10.1016/0014-5793(75)80344-9 (Wiley)
February 24, 1977	September 1, 1976	The effects of noradrenaline, acetylcholine, cyclic AMP, cyclic GMP, and other agents on the concentration of unesterified fatty acids in synaptosomes and synaptic membranes. *FEBS Letters*, DOI: 10.1016/0014-5793(76)80541-8 (Wiley)

published) to February 2019. Raw data for the dataset of 18,603 retractions covering 127 research fields from 1753 until February 2019 are available in both .csv and .xlsx format in the files named retraction_18603.csv (https://osf.io/2kymw/) [10] and retraction_18603.xlsx (https://osf.io/a2w8h/), respectively [11]. The dataset, code examples, and all figures are stored and publicly available in the OSF system [9].

Statistical methods: Having organized the dataset, we then

Table 2. The 10 retracted articles with the longest interval between publication and retraction

Date of retraction	Date of publication	Duration (yr)	Bibliographic info of the retracted article
December 23, 2003	April 22, 1923	80	Een geval van uroptoë [A case of uropters]. *Nederlands Tijdschrift voor Geneeskunde*, 67, 1855-1857 (Bohn Stafleu van Loghum)
November 1, 2007	January 1, 1955	52	Information, reproduction, and the origin of life. *American Scientist*, URL: http://www.jstor.org/stable/27826595 (Sigma Xi)
February 4, 2016	July 1, 1978	38	Hidrotische ektodermale dysplasie. *Journal of Orofacial Orthopedics / Fortschritte der Kieferorthopädie*, DOI: 10.1007/BF02225787 (Springer)
October 1, 2017	May 1, 1980	37	Jealousy, attention, and loss. A. O. Rorty (ed.), *Explaining Emotions*. University of California Press, 465-488.
February 6, 2009	May 11, 1974	35	Cello scrotum. *BMJ: British Medical Journal*, DOI: 10.1136/bmj.2.5914.335-a (BMJ Publishing)
February 26, 2018	October 1, 1985	33	A mos oncogene-containing retrovirus, myeloproliferative sarcoma virus, transforms rat thyroid epithelial cells and irreversibly blocks their differentiation pattern. *Journal of Virology,* DOI: 10.1128/JVI.56.1.284-292.1985 (American Society for Microbiology)
March 1, 2018	September 1, 1987	31	One- and two-step transformations of rat thyroid epithelial cells by retroviral oncogenes. *Molecular and Cellular Biology*, DOI: 10.1128/MCB.7.9.3365 (American Society for Microbiology)
February 8, 2016	October 1, 1986	30	Diffusion and solubility of N-alkanes in polyolefines. *Journal of Applied Polymer Science*, DOI: 10.1002/app.1986.070320501 (Wiley)
July 10, 2015	June 01, 1986	29	Volume replacement with a new hydroxyethyl starch preparation (3 percent HES 200/0.5) in heart surgery. *Transfusion Medicine and Hemotherapy*, https://www.ncbi.nlm.nih.gov/pubmed/2427448 (Karger)
October 01, 2017	September 01, 1988	29	Freud on unconscious affects, mourning, and the erotic mind. B. P. McLaughlin & A. O. Rorty (eds.), *Perspectives on Self-Deception*. University of California Press, 46-263

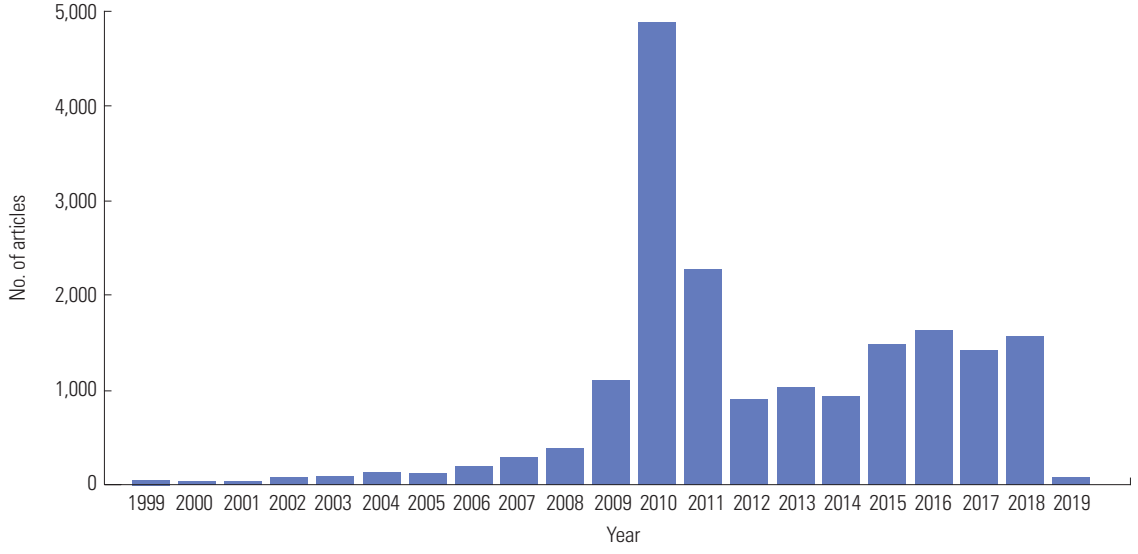

Fig. 2. Number of retracted articles per year since 1999.

Table 3. Number of publishers, journals, countries, and fields in which retraction decisions were made by year

Year	Total no.	No. of publishers	No. of journals	No. of countries	No. of fields
1756	1	1	1	1	1
1927	1	1	1	1	1
1940	1	1	1	1	2
1942	1	1	1	1	1
1960	1	1	1	1	2
1966	1	1	1	1	4
1968	1	1	1	1	2
1971	1	1	1	1	1
1977	9	5	4	2	8
1980	12	7	7	3	12
1981	3	3	3	1	7
1982	5	5	5	3	11
1983	12	7	7	1	16
1984	8	7	7	4	13
1985	13	7	7	3	13
1986	12	6	6	2	14
1987	13	9	8	1	20
1988	4	3	3	3	6
1989	13	10	10	3	21
1990	19	12	18	9	27
1991	21	13	17	5	19
1992	25	17	19	9	26
1993	23	11	18	7	22
1994	24	13	21	9	21
1995	25	11	21	9	31
1996	14	10	12	3	17
1997	40	22	30	10	39
1998	59	23	35	16	35
1999	49	22	31	14	37
2000	38	16	27	16	38
2001	33	19	29	16	36
2002	68	27	49	20	57
2003	83	30	60	29	48
2004	122	44	88	31	69
2005	117	41	91	27	64
2006	185	69	143	34	79
2007	290	74	186	43	100
2008	370	89	233	47	95
2009	1,096	93	355	55	110
2010	4,867	92	401	70	118
2011	2,240	91	452	60	112
2012	873	137	567	68	112
2013	997	142	626	71	109
2014	891	142	609	82	106
2015	1,414	201	797	70	116
2016	1,547	210	913	91	115
2017	1,362	186	818	74	113
2018	1,515	195	822	80	110
2019	67	33	50	19	56
Total	18,586	2,162	7,613	1,128	2,192

Table 4. Publishers with the largest number of retracted papers

Publisher	Total no. of retracted papers	No. of journals with retractions	No. of fields with retractions	Start year	Recent year
Institute of Electrical and Electronics Engineers	6,763	174	109	2002	2018
Elsevier	2,438	877	114	1971	2019
Springer	1,368	609	110	1940	2019
Wiley	987	399	104	1977	2019
Taylor and Francis	486	283	111	1927	2019
American Society for Biochemistry and Molecular Biology	305	4	42	1992	2019
SAGE Publications	297	113	93	1999	2019
Wolters Kluwer	252	105	70	1986	2019
BioMed Central	211	80	64	2001	2017
American Chemical Society	179	35	47	1995	2019
Hindawi	175	74	75	2007	2019
PLoS	173	6	69	2007	2019
Nature Publishing Group	166	39	55	1977	2018
Oxford Academic	131	73	74	1990	2019
International Union of Crystallography	130	1	7	2009	2012
National Academy of Sciences	121	1	41	1982	2018
American Association for Cancer Research	121	7	27	1992	2019
Lippincott Williams & Wilkins	121	25	39	1987	2018
American Society for Microbiology	119	13	26	1989	2019
Oxford University Press	109	46	61	1995	2017
American Association for the Advancement of Science	106	4	42	1968	2018
Royal Society of Chemistry	94	27	35	2003	2019
De Gruyter	90	48	71	1991	2018
BMJ Publishing	83	26	53	1989	2018
Cell Press	78	16	30	1984	2018
Mary Ann Liebert	76	32	52	1989	2018
American Heart Association	73	5	23	1982	2018
American Association of Immunologists	58	2	21	1987	2017
e-Century Publishing Corporation	57	6	34	2013	2019
Cambridge University Press	53	31	60	1982	2018
American Physiological Society	53	9	27	1983	2018
Spandidos	52	8	33	2010	2018
American Society of Hematology	49	1	17	1996	2018
IOP Publishing	49	22	26	2008	2019
AIP Publishing	45	8	18	2003	2018
Frontiers	45	18	36	2012	2019
The Company of Biologists	44	5	19	2005	2018
Dove Press	43	18	42	2012	2018
MDPI	42	25	51	2010	2019
Karger	39	21	36	2006	2019
American Society for Clinical Investigation	39	1	24	1980	2019
American Physical Society	39	6	14	2002	2018
American Diabetes Association	37	2	22	1980	2019
American Medical Association	35	11	36	1989	2018
Bentham Open	34	13	42	2015	2018
EMBO Press	34	3	15	1990	2018
Total	18,601	4,507	5,817		

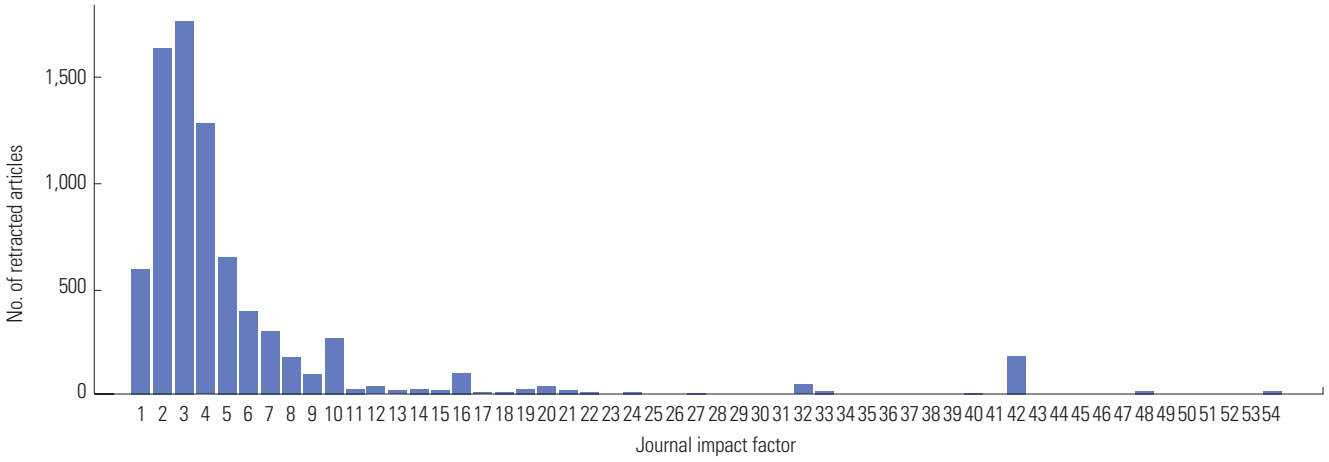

Fig. 3. Distribution of the number of retracted articles by journal impact factor.

calculated descriptive statistics to present a clear overview of the practice of retraction in scientific publishing.

Results

Retractions were found in 4,289 journals belonging to 753 publishers (or publishing organizations). From the analysis of data through February 2019, 18,603 retractions were found. In the past, this phenomenon was rare. Table 1 presents information regarding the 10 oldest retracted articles, with the oldest dating back to 1756. The next recorded retraction occurred in 1927; following that, retractions were typically recorded as taking place every several years. The first five articles on this list are not accessible because no digitized document is available.

The increasing number of retractions in recent years [12] may also reflect trends in time to retraction (the time from the publication of the article to the publication of the retraction note) [4]. We measured the time to retraction for the 10 articles with the longest time to retraction, and the longest interval before retraction was 80 years (Table 2). Four of the 10 articles listed below are not available online.

Although the first retraction was issued in 1756, retraction only began to become more common in 1999, as shown in Fig. 2, with 2010 being an anomaly.

The number of retractions and the numbers of publishers, journals, countries, and fields in which retraction decisions were made per year are reported in Table 3. Despite the increase in journals issuing retractions in recent years, the number of retractions per retracting journal has not increased. As shown in Table 3, in 2010, 4,867 papers were retracted by 401 journals or publications associated with 92 publishers. The authors of articles retracted that year came from 70 different

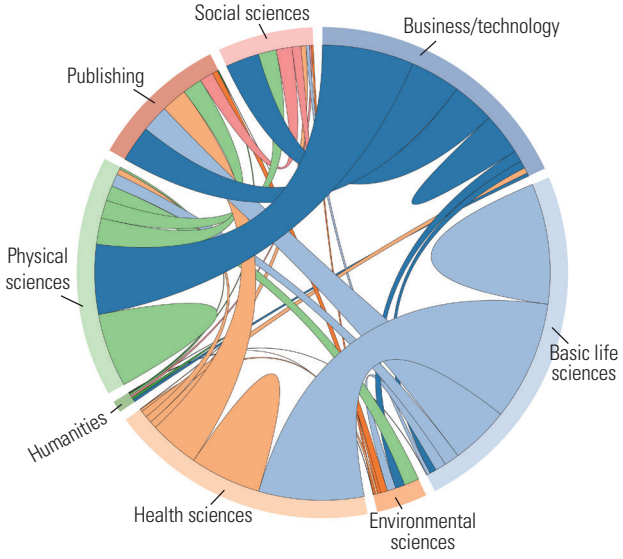

Fig. 4. Chord diagram for retractions of papers in different fields.

countries, and their papers covered 118 research fields.

Among the 753 publishers with retracted papers, the highest number of papers belonged to the Institute of Electrical and Electronics Engineers (IEEE), with 6,763 retracted articles. Elsevier had the most journals that have had papers retracted: 877 journals covering 114 research fields. The IEEE, Elsevier, and Springer accounted for 56.81% (10,569 of 18,603) of all retracted papers globally. Basic data concerning the publishers with the most retractions are given in Table 4.

Fig. 3 illustrates the distribution of the number of retracted papers by 2017 journal impact factor (JIF). It indicates that 7,836 out of 18,603 papers were published in journals with a JIF, of which more than three-fourths were published in jour-

Table 5. Top 15 countries by number of retracted articles

Country	Total no. of retracted articles	No. of publishers	First year articles were counted	Most recent year articles were counted
China	8,612	165	1995	2019
United States	3,179	237	1927	2019
India	934	130	1992	2019
Japan	874	126	1986	2019
Germany	623	94	1940	2019
United Kingdom	593	96	1756	2019
Iran	582	69	2005	2019
South Korea	520	112	1999	2019
Italy	434	92	1994	2019
Canada	307	63	1982	2018
Taiwan	302	45	1993	2019
France	301	64	1985	2019
Spain	258	70	1999	2019
Australia	254	66	1982	2018
The Netherlands	222	51	1990	2018

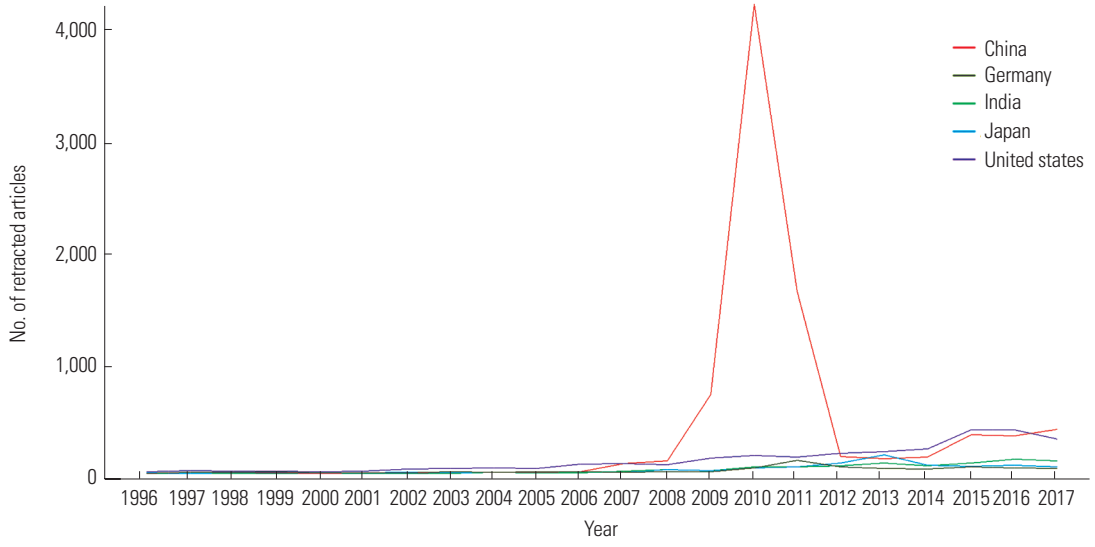

Fig. 5. Top five countries according to the number of retracted articles by year.

nals with a JIF of 5 or lower.

Data regarding retractions of papers in various fields are shown in Fig. 4.

China ranked first in the top 15 countries by number of retracted articles, as presented in Table 5.

A closer look at the top five countries showed a spurt in the retractions of articles by Chinese authors around 2010, as depicted in Figs. 5 and 6.

Discussion

Key results: RetractionWatch is among the few databases tracking retractions exclusively on the global scale; hence, making the best use of this resource can greatly benefit the scientific community. Recognizing this fact, we have collected a comprehensive database on scientific retractions from 1753 to February 2019 using SQL Server 2016 and homemade arti-

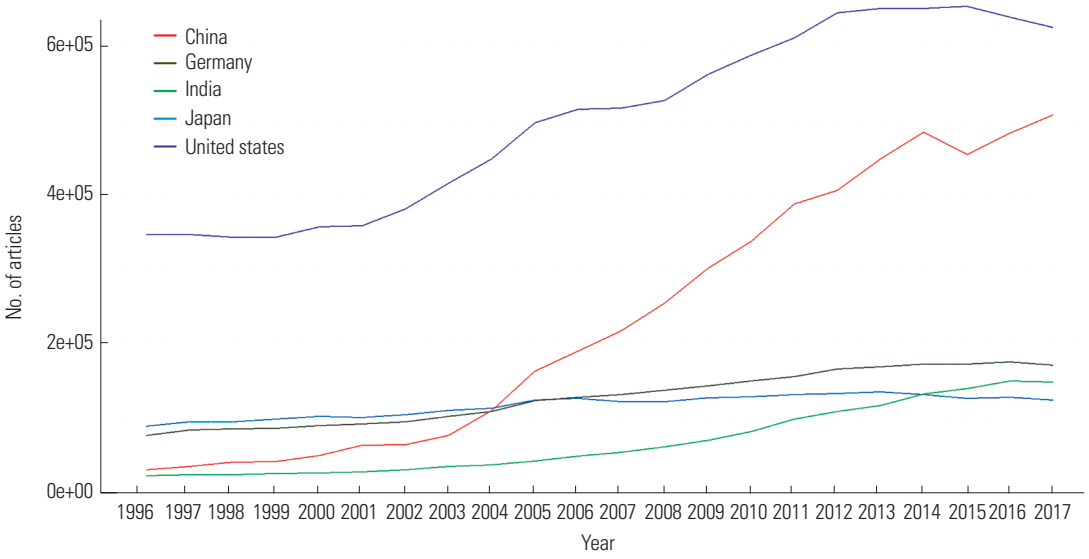

Fig. 6. Number of articles published each year of the top five countries with regard to the number of retracted articles.

ficial intelligence tools. This database enabled us to answer the questions posed in this paper. We found that although retraction is an old phenomenon, with the first retraction of a paper dating back to 1756 (Table 1); it became a common practice in 1999, and the most retractions were issued in 2010.

Moreover, the longest duration that a retracted paper stayed in the literature was 80 years (Table 2). Most notably, the IEEE, Elsevier, and Springer together accounted for nearly 60% of all retracted papers, with the IEEE accounting for the most. Of the reasons for retraction, "fake peer review" was the most common. Additionally, our database noted a sharp rise in the number of retracted papers from China (Table 5). These insights suggest that future studies can continue to explore various aspects of retractions.

Interpretation: This rise of retraction that began in 1999 (as shown in Fig. 2) is nearly consistent with the findings of Brembs et al., which concluded that the retraction rate of articles had remained stable since the 1970s and began to increase rapidly in the early 2000s. They also saw the creation and popularization of a website dedicated to monitoring retractions in 2010 [13]. However, this increase may be a sign that journal editors are becoming more skillful at identifying and removing flawed publications [14].

Diverging from previous results that held that journals with higher impact factors have a higher rate of retractions [15], our finding showed an non-significant correlation between JIF and the probability of article retraction (Fig. 3). This result is consistent with Singh et al. [3], who found a statistically non-significant relationship between the impact factor and the number of articles retracted. Different fields also had different numbers of retracted papers (Fig. 4). The majority of retractions were associated with business and technology, physical sciences, basic life science, and the health sciences. Meanwhile, the social sciences, humanities, environmental science, and publishing accounted for a small portion of all retractions. The relationships among retractions in different fields is also presented in Fig. 4. For instance, basic life sciences and health sciences had a significant number of shared retracted articles. In fields with few retractions, most of the retracted articles were shared with fields with high numbers of retractions.

The reasons for retraction can be diverse, and one paper is usually retracted for multiple reasons [4,7]. Since 2012, "fake peer review" has become a major reason, with 676 retractions for that reason during the last 7 years. About 30% (5,602) of retracted papers had undergone some investigation (Office of Research Integrity official investigation, investigation by a third party, investigation by a company/institution, or investigation by a journal/publisher) before being retracted. The findings of Qi et al. [8] also indicate that the number of retractions due to fake peer review differs among journals and countries; a majority (74.8%) of retracted papers were determined to be written by Chinese researchers.

This result may be due to China's current national situation (Table 5 and Figs. 5, 6). Greater amounts of funding and awards for conducting scientific research make researchers more eager to publish; however, measures to enforce publishing ethics may not have caught up [8]. However, it is important to note that when considering the number of retractions per publication and the amount of research funding, respectively, Iran and Romania are the top countries [16].

Limitation: This article is not exempt from limitations. First,

this study mainly employed descriptive statistics, which serve only to provide an overview and do not dive into any specific issue. Thus, future studies should make use of the resources provided by this report and focus on tackling specific problems, such as reasons for retraction or case studies of publishers or countries. Different statistical approaches, such as frequentist statistics [17] or Bayesian statistics [18], should be used. Analyses of these specific topics using different statistical methods will yield a more in-depth understanding of the practice of retraction. Second, due to paywalls, our artificial intelligence tools were unable to scan beyond basic information unless the retracted articles were open-access and available in HTML format. Similarly, this study used the 2017 JIF, also because of an accessibility issue. In the future, new technology and open-access policies of publishers may enable us to access more information.

With regard to lessons that can be learned from the above findings, what we present is only a macro-level view of the entire practice of retraction. The data, when organized and analyzed properly, will be much more useful for various stakeholders. As an example, the story of China and the drastic 2010 peak in retracted articles suggest that countries that are newcomers to the academic world should take care to avoid getting too caught up in productivity boosts, particularly in developing countries, where policy failure can be extremely consequential [19]. The provision of science financing and grants is, of course, a welcome action on the part of the government [20]; however, science policies ought not to incentivize researchers to sacrifice quality for quantity. In the face of the increase in the frequency of retractions across all fields in global academia, nurturing a culture of honesty and humility is just as important as output. Editors and publishers, as well as researchers and policy-makers, have something to learn from the story of retraction. Publishers can hold the key to mitigating the fierce competition on a playing field often leveled against emerging countries, thus supporting more sustainable practices in scientific publishing [21].

***Conclusion*:** In essence, science is a continuous process of trial and error, and only by accepting the possibility of failure can a scientist make progress [22]. Thus, this study offers an overview of retraction offered from various perspectives, in which the data was examined with regard to articles, publishers, fields, and countries. This overview suggests that retraction has boomed in the past 20 years, and that the lessons that can be learned from retractions must be taken more seriously.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Data Availability

Raw data for the dataset of 18,603 retractions covering 127 research fields from 1753 until February 2019 are available in both .csv and .xlsx format under the files named retraction_18603.csv (https://osf.io/2kymw/) and retraction_18603.xlsx (https://osf.io/a2w8h/), respectively. The dataset, code examples, all figures, and other files are deposited and publicly available in OSF (https://osf.io/pbwv3/).

Acknowledgments

This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under the National Research Grant no. 502.01-2018.19. We would also like to thank RetractionWatch for their contributions to science.

References

1. Katavic V. Retractions of scientific publications: responsibility and accountability. Biochem Med 2014;24:217-22. https://doi.org/10.11613/BM.2014.024
2. Byrne J. We need to talk about systematic fraud. Nature 2019;566:9. https://doi.org/10.1038/d41586-019-00439-9
3. Singh HP, Mahendra A, Yadav B, Singh H, Arora N, Arora M. A comprehensive analysis of articles retracted between 2004 and 2013 from biomedical literature: a call for reforms. J Tradit Complement Med 2014;4:136-9. https://doi.org/10.4103/2225-4110.136264
4. Steen RG, Casadevall A, Fang FC. Why has the number of scientific retractions increased?. PLoS One 2013;8:e68397. https://doi.org/10.1371/journal.pone.0068397
5. Wager E, Williams P. Why and how do journals retract articles? An analysis of Medline retractions 1988-2008. J Med Ethics 2011;37:567-70. https://doi.org/10.1136/jme.2010.040964
6. Bar-Ilan J, Halevi G. Post retraction citations in context: a case study. Scientometrics 2017;113:547-65. https://doi.org/10.1007/s11192-017-2242-0
7. Ribeiro MD, Vasconcelos SM. Retractions covered by Retraction Watch in the 2013-2015 period: prevalence for the most productive countries. Scientometrics 2018;114:719-34. https://doi.org/10.1007/s11192-017-2621-6
8. Qi X, Deng H, Guo X. Characteristics of retractions related to faked peer reviews: an overview. Postgrad Med J 2017;93:499-503. https://dx.doi.org/10.1136/postgradmedj-2016-133969
9. Vuong QH, La VP. Retractions data mining #1 [Internet]. Charlottesville, VA: OSF; 2019 [cited 2019 Jul 24]. Avail-

able from: https://doi.org/10.17605/OSF.IO/PBWV3
10. Vuong QH, La VP. Retractions data mining #1: retraction_18603.csv (version: 1) [dataset]. 2019 Feb 26 [cited 2019 Oct 12]. OSF. Available from: https://osf.io/2kymw/
11. Vuong QH, La VP. Retractions data mining #1: retraction_18603.xlsx (version: 1) [dataset]. 2019 Feb 26 [cited 2019 Oct 12]. OSF. Available from: https://osf.io/a2w8h/
12. Redman BK, Yarandi HN, Merz JF. Empirical developments in retraction. J Med Ethics 2008;34:807-9. https://doi.org/10.1136/jme.2007.023069
13. Brembs B, Botton K, Munafo M. Deep impact: unintended consequences of journal rank. Front Hum Neurosci 2013;7:291. https://doi.org/10.3389/fnhum.2013.00291
14. Fanelli D. Why growing retractions are (mostly) a good sign. PLoS Med 2013;10:e1001563. https://doi.org/10.1371/journal.pmed.1001563
15. Fang FC, Casadevall A. Retracted science and the retraction index. Infect Immun 2011;79:3855-9. https://doi.org/10.1128/IAI.05661-11
16. Oransky I. Volunteer watchdogs pushed a small country up the rankings. Science 2018;362:395. https://doi.org/10.1126/science.362.6413.395
17. Agresti A, Kateri M. Categorical data analysis. Berlin: Springer; 2011.
18. Vuong QH, Bui QK, La VP, et al. Cultural evolution in Vietnam's early 20th century: a Bayesian networks analysis of Hanoi Franco-Chinese house designs. Soc Sci Humanit Open 2019;1:100001. https://doi.org/10.1016/j.ssaho.2019.100001
19. Vuong QH. The (ir)rational consideration of the cost of science in transition economies. Nat Hum Behav 2018; 2:5. https://doi.org/10.1038/s41562-017-0281-4
20. Vuong QH. Breaking barriers in publishing demands a proactive attitude. Nat Hum Behav 2019;3:1034. https://doi.org/10.1038/s41562-019-0667-6
21. Vuong QH. The harsh world of publishing in emerging regions and implications for editors and publishers: the case of Vietnam. Learn Publ 2019;32:314-24. https://doi.org/10.1002/leap.1255
22. Toan HM, Kong NT, Trang VT, Hoang NM, Tung HM. To walk on the Penrose stairs of science [Internet]. London: A Community from Nature Research; 2019 Oct 10 [cited 2019 Oct 12]. Available from: https://socialsciences.nature.com/users/301097-ho-manh-toan/posts/54541-to-walk-on-the-penrose-stairs-of-science